Systematics of *Buddleja* (Scrophulariaceae): phylogenetic relationships, historical biogeography, and phylogenomics

John H. Chau

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Richard G. Olmstead, Chair

Verónica S. Di Stilio

Adam D. Leaché

Program Authorized to Offer Degree:

Department of Biology

University of Washington

Abstract

Systematics of *Buddleja* (Scrophulariaceae):
phylogenetic relationships, historical biogeography, and phylogenomics

John H. Chau

Chair of the Supervisory Committee:
Professor Richard G. Olmstead
Department of Biology

Plants display incredible diversity, in morphology and spatial distribution, which can best be understood in an evolutionary context. The reconstruction of how this diversity has evolved can illuminate patterns and trends in the evolution of functionally and ecologically important traits and on how modern plant communities have formed around the globe. Case studies of individual taxa that encompass such diversity allow for thorough taxonomic sampling and detailed analysis of traits and distribution. The tribe Buddlejeae in Scrophulariaceae comprises 108 species of trees and shrubs in five genera: *Buddleja*, *Chilianthus*, *Emorya*, *Gomphostigma*, and *Nicodemia*. They are variable in flower color and shape, inflorescence architecture, fruit type, leaf shape and texture, and habitat preference, among other traits. They also have a wide distribution in tropical montane and subtropical regions of Africa, Madagascar, Asia, North America, and South America. Prior phylogenetic studies including the group have had limited taxonomic sampling, and evolutionary relationships between species and genera remained unknown. In Chapter 1, I infer a phylogeny for tribe Buddlejeae with extensive taxonomic sampling from all five genera and all major areas of distribution, using multiple nuclear and

plastid markers. *Buddleja* and *Chilianthus* were resolved to be non-monophyletic, with *Buddleja* paraphyletic with respect to the other four genera. A new classification is proposed in which the other four genera are combined with *Buddleja* and seven sections in *Buddleja* are erected. Ancestral character state reconstructions show that some traits, including stamen exsertion, corolla shape, and inflorescence type, converged on similar states multiple times. The plesiomorphic trait states in Buddlejeae include capsular fruits, included stamens, white and tube-shaped corollas, and paniculate inflorescences. In Chapter 2, I infer a time-calibrated phylogeny for *Buddleja*, reconstruct ancestral distributions, and test for shifts in diversification rate dynamics. We found that from an ancestral distribution on continental Africa, *Buddleja* expanded its range to the New World, Asia, and Madagascar, one time each, in the mid to late Miocene. Long-distance dispersal or migration through northern high-latitude corridors may have allowed for these range expansions. An increase in speciation rate early in the diversification of the New World clade suggests conditions conducive to speciation in the American cordilleras. In Chapter 3, I use phylogenomic methods to infer a better-supported phylogeny for *Buddleja*, with particular focus on the Asian clade. Four locus sets were identified as targets for sequence capture and high-throughput sequencing. A "taxon-specific" locus set was developed using genomic and transcriptome data from two species of *Buddleja*. Three "general" locus sets were chosen from previous studies that used genomic data from several distantly related angiosperms. A greater number of loci were developed for the "taxon-specific" set. All sets had a very high proportion of target loci with assembled sequences for *Buddleja* species, but "general" sets had greater assembly for outgroup taxa. The "taxon-specific" and PPR loci had the highest average percentage of variable sites. A fully resolved and highly supported phylogeny for the Asian *Buddleja* clade can serve as a framework for future evolutionary studies.

ACKNOWLEDGEMENTS

I am grateful to my committee for all their support and advice over my time at UW. I especially thank my advisor, Dick Olmstead, for his encouragement of my work, for the independence he has given me, and for being a role model to me as a scholar and person. I also thank Adam Leaché and Verónica Di Stilio for their helpful suggestions and critiques of my work and for their generosity in sharing their lab resources. I am thankful to Sarah Reichard for reminding me to think more broadly about *Buddleja*'s place in the natural world, and to Eric Turnblom for his openness and interest as I was finishing my dissertation.

I thank Eliane Norman for all her work with *Buddleja* and insights shared through our conversations. My research rests on the foundations she and other plant systematists laid. I thank Nataly O'Leary for her invaluable contributions to this research in procuring specimens and generating plastid sequence data and for her patience. I thank Sun Wei-Bang for facilitating collaborations with his lab in Kunming, which allowed me to have nearly complete sampling of Asian taxa and to experience these plants in the field.

I thank the graduate students in the Olmstead lab, past and present, for our conversations about science, for passing on to me ancient lab knowledge, and for being a constant source of humor and support: Valerie Soza, Yuan Yaowu, Ryan Miller, Pat Lu-Irving, Laura Frost, Audrey Ragsac, and Ana Maria Bedoya. I also thank the many undergrads I have had the opportunity to work with, who have helped me generate data for this project: Bailey Craig, Anna O'Brien, Shulamit Harkavy, Ben Meersman, Calvin Lee, and Georgia Seyfried. I give special thanks to Wolf Rahfeldt for his programming expertise and excellent companionship.

I thank the herbarium (WTU), especially David Giblin, for facilitating my work with specimens and for their financial and personal support. I also thank the Botany Greenhouse, especially Jeanette Milne and Doug Ewing, and the Medicinal Herb Garden and Keith Possee for allowing me to grow *Buddleja* in Seattle and see some in flower for the first time, for making teaching about plants immensely more interesting, and for many enjoyable and enlightening hours spent with plants.

The staff of the UW Biology department has helped me tremendously in various ways through my years here. I especially thank Karen Bergeron for her guidance when I submitted my proposal for the NSF DDIG and for her affability.

I thank the many institutions that have provided sources of plant specimens for my research: CAS, DAV, GH, GRA, K, KUN, LPB, MO, NBG, NY, SI, TEX, UC, WTU, Arnold Arboretum, Hoyt Arboretum, Kunming Botanical Garden, Kirstenbosch Botanical Garden, Los Angeles County Arboretum, Rancho Santa Ana Botanic Garden, San Francisco Botanical Garden, and University of California-Davis Arboretum. I especially thank the University of Washington Botanic Gardens for accepting into their collections the *Buddleja* specimens I have started and grown. Systematics and biodiversity research would be vastly more difficult without these important collections and records of life on Earth.

I thank the many people who have made my fieldwork immeasurably easier and more enjoyable: Segundo Leiva and Mario Zapata Cruz in Peru; Dave Tank and Diego Morales-Briones in Bolivia; the Organization for Tropical Studies and Ricardo Kriebel in Costa Rica; Alistair Barker, Tony Dold, and Nigel Barker in South Africa; Sun Wei-Bang, Xin Zhang, and the other members of Prof. Sun's lab in China; and especially Simon Uribe-Converse in Peru and Bolivia, who I explored beautiful Peru with on my first field trip and who inspired me to be a better field botanist and scientist.

I thank the sources of financial support that have made this research possible: the Graduate Research Fellowship and Doctoral Dissertation Improvement Grant from the National Science Foundation; the Sargent Award, Giles Botanical Field Research Award, Washington Research Foundation and Benjamin Hall Fellowship, and Melinda Denton Writing Fellowship from the UW Department of Biology; and student research grants from the Botanical Society of America, American Society of Plant Taxonomists, and Society of Systematic Biologists.

I thank my friends, in Seattle and beyond, who have provided camaraderie and welcome distractions from graduate school.

Finally, I am grateful for my education in the public school system of the United States, in the Los Angeles Unified School District, Cornell University, University of California-Davis, and University of Washington, for the opportunities it has given me to engage in the pursuit and sharing of knowledge.

CONTENTS

CHAPTER 1:

Phylogenetic relationships in tribe Buddlejeae (Scrophulariaceae) based on multiple nuclear and plastid markers

Citation:
Chau JH, O'Leary N, Sun W-B, & Olmstead RG. 2017. Phylogenetic relationships in tribe Buddlejeae (Scrophulariaceae) based on multiple nuclear and plastid markers. *Botanical Journal of the Linnean Society* 184: 137-166.
doi: https://doi.org/10.1093/botlinnean/box018

Botanical Journal of the Linnean Society, 2017, **184**, 137–166. With 4 figures.

Phylogenetic relationships in tribe Buddlejeae (Scrophulariaceae) based on multiple nuclear and plastid markers

JOHN H. CHAU[1]*, NATALY O'LEARY[2], WEI-BANG SUN[3] and RICHARD G. OLMSTEAD[1]

[1]*Department of Biology and Burke Museum, University of Washington, Box 351800, Seattle, WA 98195, USA*
[2]*Instituto de Botánica Darwinion, Labardén 200, San Isidro, Argentina*
[3]*Kunming Botanical Garden, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China*

Received 26 June 2016; revised 14 November 2016; accepted for publication 14 March 2017

Buddlejeae comprise *c*. 108 species in five commonly accepted genera: *Buddleja*, *Chilianthus*, *Emorya*, *Gomphostigma* and *Nicodemia*. Conflicting generic and infrageneric level classifications based on morphology attest to a need to evaluate relationships and trait evolution in a molecular phylogenetic framework. We use multiple independent loci from the nuclear and plastid genomes and representative taxonomic sampling to infer phylogenetic relationships using maximum likelihood and Bayesian analyses with single-locus and concatenated data and Bayesian multispecies coalescent analyses. *Nicodemia* and *Gomphostigma* are resolved as monophyletic. *Chilianthus* is not monophyletic, with three species in one clade and *Buddleja glomerata* (=*Chilianthus lobulatus*) possibly separate. *Buddleja* is paraphyletic with respect to *Chilianthus*, *Emorya*, *Nicodemia* and, probably, *Gomphostigma*. We propose a new classification to reflect phylogenetic relationships in Buddlejeae. Only *Buddleja* is retained at the generic level. *Chilianthus*, *Nicodemia*, *Gomphostigma* and *Emorya* are combined with *Buddleja*, with a new name and new combination erected for the two *Emorya* spp., *Buddleja normaniae* and *B. rinconensis*. Sectional classification of *Buddleja* is revised, with two new monotypic sections being proposed, *Salviifoliae* and *Pulchellae*, and *Gomphostigma* being lowered to sectional rank. Reproductive morphological traits traditionally used to define genera, including stamen exsertion, corolla shape and inflorescence type, were reconstructed on the phylogenetic tree and are inferred to have converged on similar states multiple times. Plesiomorphic trait states in Buddlejeae include capsular fruits, included stamens, white and tube-shaped corollas and paniculate inflorescences.

ADDITIONAL KEYWORDS: *Buddleja* – *Chilianthus* – chloroplast DNA – classification – *Emorya* – *Gomphostigma* – morphology – *Nicodemia* – PPR loci.

INTRODUCTION

Scrophulariaceae *s.s.* were first recognized as a distinct clade in the more broadly circumscribed and polyphyletic Scrophulariaceae *s.l.* by Olmstead & Reeves (1995) and were subsequently upheld in additional phylogenetic analyses of DNA markers (Oxelman, Backlund & Bremer, 1999; Kornhall, Heidari & Bremer, 2001; Olmstead *et al.*, 2001; Oxelman *et al.*, 2005; Rahmanzadeh *et al.*, 2005). Scrophulariaceae *s.l.* were predominantly bilateral in corolla symmetry and cosmopolitan in distribution, including many charismatic taxa of the northern temperate flora (e.g. *Antirrhinum* L., *Castilleja* Mutis ex L.f., *Digitalis* L., *Mimulus* L., *Penstemon* Schmidel, *Scrophularia* L., *Verbascum* L., *Veronica* L.), whereas Scophulariaceae *s.s.* as currently circumscribed (Olmstead *et al.*, 2001; APG II, 2003; Tank *et al.*, 2006; APG IV, 2016) are composed mostly of taxa with radial or sub-radial corolla symmetry and distribution in the Southern Hemisphere. Phylogenetic studies of Scrophulariaceae *s.s.* identified eight tribes (Kornhall *et al.*, 2001; Kornhall & Bremer, 2004; Oxelman *et al.*, 2005), including Buddlejeae, which comprise *c.* 108 species and are one of only two tribes that have major radiations in the Northern and Southern Hemispheres (Tank *et al.*, 2006). Buddlejeae are typically shrubs or

*Corresponding author. E-mail: jhchau@uw.edu

trees with opposite leaves and interpetiolar stipules, stellate, glandular trichomes and tetramerous, radially symmetrical flowers arranged in cymes. Buddlejeae have a broad distribution, encompassing tropical, subtropical and warm-temperate areas of Africa, Asia and North and South America, and display wide morphological diversity, especially in the flower and inflorescence (Norman, 2000; Oxelman, Kornhall & Norman, 2004). Several species are known for their horticultural value [e.g. *Buddleja davidii* Franch. (butterfly bush), *B. alternifolia* Maxim., *B. globosa* Hope], invasiveness following introductions outside their native range [e.g. *B. davidii* and *B. madagascariensis* Lam. (=*Nicodemia madagascariensis* (Lam.) R.Parker)] and use in traditional medicine [e.g. *B. officinalis* Maxim. (Chinese: mi meng hua), *B. coriacea* J.Rémy and *B. incana* Ruiz & Pav. (Quechua: kiswar)] (Li & Leeuwenberg, 1996; Norman, 2000; Tallent-Halsell & Watt, 2009).

Tribe Buddlejeae as now recognized have had a complicated taxonomic history (see Norman, 2000, for a detailed review). They have been considered at various ranks as part of Scrophulariaceae (Bentham, 1835, 1846) or Loganiaceae (Bentham, 1857; Bentham & Hooker, 1876; Solereder, 1895; Leeuwenberg & Leenhouts, 1980) or separated as the family Buddlejaceae (Wilhelm, 1910; Wagenitz, 1964; Hutchinson, 1973; Takhtajan, 1980; Cronquist, 1981; Dahlgren, 1983; Thorne, 1983, 1992; Norman, 2000; Oxelman *et al.*, 2004). Molecular data from the plastid genome resolved the position of the group in Scrophulariaceae *s.s.* (Olmstead & Reeves, 1995; Oxelman *et al.*, 1999; Olmstead *et al.*, 2001), which supported earlier evidence of affinity from embryology (Wagenitz, 1964; Hakki, 1980), palynology (Punt & Leenhouts, 1967) and phytochemistry (Jensen, Nielsen & Dahlgren, 1975). Molecular phylogenetic studies also clarified the positions of several taxa that were once thought to be closely related and included in the group. *Androya* H.Perrier was transferred to tribe Myoporeae in Scrophulariaceae, *Nuxia* Lam. to Stilbaceae, *Polypremum* L. to Tetrachondraceae and *Peltanthera* Benth. and *Sanango* G.S.Bunting & J.A.Duke to or near Gesneriaceae (Oxelman *et al.*, 1999; Refulio-Rodriguez & Olmstead, 2014), leaving five genera, *Buddleja* L., *Chilianthus* Burch., *Nicodemia* Ten., *Gomphostigma* Trucz. and *Emorya* Torr., in Buddlejeae (Oxelman *et al.*, 2004).

The majority of the species diversity and distributional area of the tribe is encompassed by *Buddleja*, which includes > 90 species distributed in Africa, Asia, North America and South America. Reproductive morphology in the genus is variable especially in corolla shape (short and cup-shaped to long and tubular), corolla colour (various shades of white, yellow, orange or purple) and architecture of the inflorescence in which cymes are arranged (paniculate, thyrsoid, spiciform or capitate) (Leeuwenberg, 1979; Norman, 2000; Oxelman *et al.*, 2004). *Buddleja* was last comprehensively treated by Bentham (1846), who divided the genus based on differences in floral and inflorescence morphology. The Asian species were reclassified by Marquand (1930) and Li (1982), who erected infrageneric taxa based on phyllotaxy and floral traits. Leeuwenberg (1979) conducted a study of the African and Asian species and proposed a global classification based on reproductive morphology, in which most species were placed in a single section. Norman (2000) completed a monograph of the New World species and proposed 12 series based on morphology and ecogeography. A summary of generic and infrageneric classifications is presented in Table 1.

Four species in Buddlejeae from southern Africa have been treated as members of *Buddleja* (Leeuwenberg, 1979) or the segregate genus *Chilianthus* (Bentham, 1846; Norman, 2000; Oxelman *et al.*, 2004). This group of species has been recognized because their floral morphology is distinguished by short, cup-shaped corollas, stamens with relatively long filaments that are partly or fully exserted and cymes in highly branched paniculate inflorescences. Some studies have suggested, however, that these morphological characters are neither constant in, nor exclusive, to these four species (Phillips, 1946; Leeuwenberg, 1979). Leeuwenberg (1979), who completed the most recent taxonomic study of African members of Buddlejeae, recognized the group at the section level in *Buddleja*. Additionally, he removed one species, *B. loricata* Leeuwenberg, from this group because it has anthers with shorter filaments that are barely exserted from the corolla. Earlier studies suggested an affinity between *Chilianthus* and *Nuxia* due to similarities in floral and pollen morphology (Leeuwenberg, 1979; Punt, 1980). However, phylogenetic analyses of plastid DNA sequences showed that *Nuxia* is outside Scrophulariaceae (Oxelman *et al.*, 1999).

Eight species from Madagascar are distinct in having fleshy, indehiscent berry-like fruits instead of dry, dehiscent capsules as in all other members of Buddlejeae. Although originally described in *Buddleja* and sometimes treated at an infrageneric rank there (Bentham, 1846; Leeuwenberg, 1979; Li, 1982; Norman, 2000), these species have also been segregated into the genus *Nicodemia* (Marquand, 1930; Oxelman *et al.*, 2004). A subset of these species was placed in another segregate genus *Adenoplea* Radlk. because they have four-celled rather than two-celled ovaries as found in the rest of Buddlejeae. Another genus *Adenoplusia* Radlk. was erected because its members, which have all been combined with the species *Buddleja axillaris* Willd., have drupe-like fruits with a chartaceous endocarp (Bruce & Lewis, 1960; Leenhouts, 1962; Leeuwenberg, 1979).

Table 1. Selected generic and infrageneric classifications for Buddlejeae

Bentham (1846)	Marquand (1930)	Leeuwenberg (1977, 1979)	Li (1982)	Norman (2000)	Oxelman *et al.* (2004)	Chau *et al.* (this study)
Genus *Buddleja* Section *Lozada* Subsection *Paniculatae* (NW) Subsection *Globosae* (NW) Subsection *Verticillatae* (NW) Section *Neemda* Subsection *Glomeratae* (Af, As, M, NW) Subsection *Thyrsoideae* (Af, NW) Subsection *Stachyoideae* (NW) Subsection *Macrothyrsae* (Af, As, M)	Genus *Buddleja* Series *Gynandrae* (As) Series *Alternifoliae* (As) Series *Curviflorae* (As) Series *Rectiflorae* (Af, As)	Genus *Buddleja* Section *Buddleja* (NW) Section *Neemda* (Af, As, NW)	Genus *Buddleja* Subgenus *Buddleja* Section *Alternifoliae* (As) Section *Neemda* Series *Curviflorae* (As) Series *Rectiflorae* (As)	Genus *Buddleja* Section *Buddleja* Series *Thyrsoides* (NW) Series *Oblongae* (NW) Series *Stachyoides* (NW) Series *Globosae* (NW) Series *Anchoenses* (NW) Series *Glomeratae* (NW) Series *Brachiatae* (NW) Series *Lanatae* (NW) Series *Scordioides* (NW) Series *Buddleja* (NW) Series *Verticillatae* (NW) Series *Cordatae* (NW)	Genus *Buddleja*	Genus *Buddleja* Section *Salviifoliae* (Af) Section *Pulchellae* (Af) Section *Alternifoliae* (As) Section *Buddleja* (NW)

Table 1. *Continued*

Bentham (1846)	Marquand (1930)	Leeuwenberg (1977, 1979)	Li (1982)	Norman (2000)	Oxelman *et al.* (2004)	Chau *et al.* (this study)
Subsection *Axilliflorae* (M)	Genus *Nicodemia*	Section *Nicodemia* (M)	Subgenus *Nicodemia* (M)	Section *Nicodemia* (M)	Genus *Nicodemia*	Section *Nicodemia* (Af, M)
Genus *Chilianthus*	–	Section *Chilianthus* (Af)	–	Genus *Chilianthus*	Genus *Chilianthus*	Section *Chilianthus* (Af)
Genus *Gomphostigma*	–	Genus *Gomphostigma*	–	–	Genus *Gomphostigma*	Section *Gomphostigma* (Af)
–	–	–	–	Genus *Emorya*	Genus *Emorya*	(part of Section *Buddleja*)

Groups corresponding to five previously accepted genera are shown on same row. Distribution is indicated for infrageneric groups (Af = continental Africa; As = Asia; M = Madagascar; NW = New World). For additional classification schemes, see Leeuwenberg & Leenhouts (1980).

Gomphostigma includes two species from southern Africa. They were first described as members of *Buddleja*, but were later segregated on the basis of their distinctive inflorescences, which are racemose rather than cymose, and flowers with corollas that are short and cup-shaped rather than tubular. Recent taxonomic treatments have kept this group distinct from *Buddleja* (Leeuwenberg, 1977; Oxelman *et al.*, 2004).

Emorya, with two species occurring in northern Mexico and the adjoining south-western United States, is distinct in its floral morphology from other Buddlejeae in North America. Their flowers have exserted stamens with long filaments and an exserted style and the corollas are tubular and much longer than those in all North American *Buddleja*. The corolla morphology in *Emorya* is similar to that in South American members of *Buddleja* series *Stachyoides* (Benth.) E.M.Norman. However, South American *Buddleja* spp. have stamens and styles that are included (Norman & Moore, 1968; Norman, 2000). Taxonomic treatments have always treated *Emorya* as distinct from *Buddleja* (Norman, 2000; Oxelman *et al.*, 2004).

Molecular phylogenetic studies including members of these five genera have shown that they form a well-supported clade in Scrophulariaceae (Oxelman *et al.*, 1999; Kornhall *et al.*, 2001; Kornhall & Bremer, 2004; Oxelman *et al.*, 2005). These studies have focused on higher-level relationships or other groups in the family and included at most one or two exemplars from each genus of Buddlejeae. Additionally, New World *Buddleja*, which is the most species-rich group in the tribe, has been represented by only a single species in one study (Kornhall & Bremer, 2004). It remains uncertain whether each of the five genera is monophyletic and what the pattern of relationships is among and within them. Moreover, all prior molecular data have come from the non-recombining plastid genome. Single gene trees may not accurately reflect species evolutionary history due to confounding factors, including incomplete lineage sorting, hybridization and introgression (Maddison, 1997). Single- and low-copy loci from the nuclear genome provide a source of independent data and are also often more quickly evolving (Sang, 2002; Small *et al.*, 2004), which may be more appropriate for studies at the level of species in Buddlejeae.

We present here the first molecular phylogenetic analysis of tribe Buddlejeae with broadly representative taxonomic sampling, including members of all recognized genera and extensive sampling of species in the large genus *Buddleja* from all parts of its range. We use sequence data from the nuclear ribosomal locus external transcribed spacer (ETS), three low-copy nuclear genes from the pentatricopeptide repeat (PPR) gene family (At1G31430/PPR24, At4G30825/PPR97, At5G39980/PPR123) and three plastid regions (*rpo*A,

trnD-trnT, *trnS-trnfM*). Our goals are to assess monophyly of the genera in Buddlejeae, evaluate relationships of major clades against current classifications, investigate the evolutionary history of morphological traits traditionally important in delimiting genera and establish a revised classification that reflects the phylogenetic trees.

MATERIAL AND METHODS

Taxon sampling

Representative species from all genera and major areas of distribution were selected for this study. We follow the species names used in the most recent monographic works for the Old World and New World taxa (Leeuwenberg, 1977, 1979; Norman, 2000) and subsequent reports of newly described and resurrected species (Liu & Peng, 2004, 2006; Morales & González, 2007; Zhang *et al.*, 2014). The species that have been segregated into *Chilianthus* and *Nicodemia* are referred to by their name in *Buddleja*, as in the monograph by Leeuwenberg (1979), but their phylogenetic coherence and position will be a focus of our analyses.

Seventy-three out of 104 *Buddleja* spp. were sampled, including all four species sometimes treated as *Chilianthus* and six of eight species sometimes treated as *Nicodemia*. In *Buddleja*, we included all four species from Africa without synonyms in *Chilianthus* or *Nicodemia*, 20 of 24 species from Asia, 12 of 19 species from North America and 28 of 46 species from South America, including two subspecies of *B. elegans* Cham. & Schltdl. All series and sections of Marquand (1930), Li (1982) and Leeuwenberg (1979) for Old World species and 11 of the 12 series proposed by Norman (2000) for New World species are represented. Both *Gomphostigma* spp. and one of two *Emorya* spp. were sampled. Six species were included as outgroups based on prior studies (Oxelman et al., 1999, 2005; Kornhall *et al.*, 2001), including two taxa from the sister clade to Buddlejeae [*Oftia africana* (L.) Bocq. and *Phygelius capensis* E.Mey. ex Benth.], two more distant taxa in the Scrophulariaceae (*Scrophularia nodosa* L. and *Nemesia fruticans* Benth.) and two additional taxa in Lamiales (*Nuxia floribunda* Benth. in Stilbaceae and *Lantana depressa* Small in Verbenaceae). Voucher information and collection localities for all specimens are presented in Table A1.

Molecular methods

Leaf tissue was sampled from specimens either as silica gel-preserved material from plants collected in the field or as fragments from herbarium specimens. Total DNA was extracted from leaf tissue using a modified CTAB procedure (Doyle & Doyle, 1987) and purified by isopropanol precipitation. For some specimens from herbarium material, DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). For four specimens, DNA was directly obtained from the DNA banks at the Missouri Botanical Garden or the Royal Botanic Gardens, Kew (Table A1).

PCR amplification reactions for nuclear markers were performed in 25 µL volumes with 1 µL genomic DNA, 0.125 µL *Taq* DNA polymerase and final concentrations of 1× PCR buffer, 3 mM $MgCl_2$, 1 µg/µL bovine serum albumin, 0.25 mM dNTP mix and 0.25 µM each of the forward and reverse primers. Where amplification proved difficult, 1× TBT-PAR was included in the reaction mix (Samarakoon, Wang & Alford, 2013). Reactions were run in a MJ Research (Bio-Rad, Hercules, CA, USA) thermocycler with the following conditions: initial denaturation at 94 °C for 2 min; followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 50 °C for 30 s, and extension at 72 °C for 1.5–2.5 min; and a final extension at 72 °C for 10 min. Four nuclear loci were targeted: the ETS region of ribosomal DNA and three PPR loci. The universal 18S-IGS (Baldwin & Markos, 1998) and Lamiales-specific ETS-B (Beardsley, Yen & Olmstead, 2003) primers were used to amplify ETS. Lamiales-specific primers were designed to amplify and sequence two PPR loci (Table A2). For locus At1G31430, hereafter called PPR24 according to its position in table 1 of Yuan *et al.* (2009), primers PPR24-140F and PPR24-1354R were used. For locus At4G30825 (PPR97), primers PPR97-781F and PPR97-1585R were used. For a third PPR locus, At5G39980 (PPR123), the Lamiidae-specific primers 550F and 1890R (Yuan *et al.*, 2010) were used. Three regions in the plastid genome were also targeted. Amplification reactions for plastid markers followed the protocols used in Yuan & Olmstead (2008). The *trnD-trnT* region was amplified with primers $trnD^{GUC}$F and $trnT^{GGU}$, the *trnS-trnfM* region with primers $trnS^{UGA}$ and $trnfM^{CAU}$ (Demesure, Sodzi & Petit, 1995; Shaw *et al.*, 2005) and the *rpo*A region with primers RPOA2 and RPOA5 (Petersen & Seberg, 1997). Amplification products were cleaned using polyethylene glycol precipitation.

Sanger cycle sequencing was performed using the standard Applied Biosystems protocol with BigDye v3.1 and PCR or internal primers (Table A2). Sequencing reaction products for nuclear loci were purified by filtering through Sephadex G-50 columns or precipitation with sodium acetate and ethanol and then read on an Applied Biosystems 3130XL or 3730 Genetic Analyzer (Thermo Fisher Scientific, Grand Island, NY, USA). Plastid loci and some ETS sequencing reactions were performed by Macrogen Inc. using Applied Biosystems PRISM BigDye Terminator Cycle Sequencing Kits with AmpliTaq DNA polymerase (Applied Biosystems, Seoul, South Korea). For most

of the length of each locus, at least two overlapping sequence fragments were generated to check for random sequencing errors. Sequence fragment data were edited and assembled into full sequences using Sequencher 4.7 (Gene Codes Corp.). Sites with multiple peaks were coded as ambiguities. All sequences have been deposited in GenBank (Table A1).

Phylogenetic analyses

For each locus, sequences were aligned with MAFFT v7 (Katoh & Standley, 2013) using the default strategy and parameters (scoring matrix = 200PAM/K = 2; gap opening penalty = 1.53). Alignments were checked by eye and minor adjustments performed manually using Se-Al v2.0a11. A few plastid sequences (*B. blattaria* J.F.Macbr.: *trnS-trnfM*, *B. incana*: *trnS-trnfM*, *B. lanata* Benth.: *trnD-trnT*, *B. rufescens* Willd. ex Schultes & Schultes: *trnS-trnfM*) had regions that were difficult to align and these were deleted from the sequence.

Statistical analyses were used to reconstruct phylogenetic trees for each of the four nuclear loci, a concatenated plastid three-locus dataset and a concatenated nuclear and plastid seven-locus dataset. In the concatenated datasets, sequences from multiple accessions of the same species were combined in order to maximize the number of loci with sequence data for each species. Phylogenetic analyses with ETS sequences from all accessions were performed (Supporting Information, Fig. S2) and sequences were combined for a species only if there was no support for non-monophyly among accessions of that species. Although it has been suggested that composite taxa may give misleading results in phylogenetic analyses (Malia, Lipscomb & Allard, 2003), it has been demonstrated that their use can perform as well as or better than data matrices with more missing data, especially when there is evidence that combined taxa are monophyletic (Campbell & Lapointe, 2009).

The substitution model for each locus was chosen according to the Akaike information criterion (AIC) as calculated using jModeltest 2.1.4 (Guindon & Gascuel, 2003; Darriba *et al.*, 2012) with three substitution schemes. To reduce the problem of large sampling error, models that account for among-site rate variation using both a gamma distribution and proportion of invariable sites were excluded in favour of those that use only a gamma distribution (Sullivan, Swofford & Naylor, 1997). Concatenated datasets were partitioned by locus for analyses such that all evolutionary model parameters were unlinked.

Maximum likelihood analyses were performed in GARLI 2.0 (Zwickl, 2006; http://garli.googlecode.com). For the full search analyses, the generation termination condition was set at 20 000 and the score improvement threshold was set at 0.001. All other settings were left at the default. Search runs were repeated until at least two replicates resulted in best-scoring trees with the same topology or 100 replicates were performed. For bootstrapping, 1000 replicates were performed with the generation termination condition decreased to 10 000 and the number of search replicates per bootstrap replicate set at 1.

Bayesian analyses were performed using MrBayes 3.2.1 or 3.2.3 (Ronquist *et al.*, 2012) on CIPRES Science Gateway (http://www.phylo.org/index.php). For each analysis, two runs with four chains each were performed. Analyses were run for 10 000 000 generations with a sampling frequency of 1000 for single-locus and concatenated plastid datasets and 30 000 000 generations with a sampling frequency of 3000 for the concatenated seven-locus dataset. Convergence was assessed by checking that the average standard deviation of split frequencies was < 0.05, the estimated sample size of parameters was > 200 as calculated in Tracer v1.5 (Rambaut & Drummond, 2009) and the plot of split frequencies showed high correlation as generated in AWTY (Wilgenbusch, Warren & Swofford, 2004). The initial 25% of trees sampled were discarded as burn-in. To evaluate the appropriateness of concatenating data from separate loci, the topologies of individual gene trees were visually examined for incongruences that are well supported [bootstrap percentage (BP) > 70% and posterior probability (PP) > 0.90].

Species tree estimation under the multispecies coalescent model was performed using *BEAST in BEAST v1.8.1 or v1.8.0 (Drummond *et al.*, 2012) on CIPRES, with data from all seven loci. Each of the four nuclear loci and the combined plastid dataset were treated as independent and set to have unlinked trees and clock models. In addition, all individual loci, including each of the three plastid loci, were set to have unlinked substitution models. The clock model for each locus was set as an uncorrelated lognormal relaxed clock with a mean having an exponential distribution with a mean of 10. The birth-death process was used as the species tree prior. Two runs were performed, each for 700 000 000 generations with a sampling frequency of 40 000. Convergence was assessed by evaluating the estimated sample size of parameters and checking for stationarity in the plot of log-likelihoods using Tracer v1.5. The initial 25% of trees was removed as burn-in and trees from both runs were combined before generating the maximum clade credibility tree with median node heights in TreeAnnotator v1.8.1.

Topology testing

Topology tests were used to assess the monophyly of proposed genera as previously circumscribed. The maximum likelihood tree was inferred using GARLI 2.0 for the full concatenated dataset, with topological

constraints such that species traditionally placed in genera formed a clade. Six different constraints were tested: (1) *Chilianthus s.l.*, including *B. loricata* as monophyletic; (2) *Chilianthus s.s.*, excluding *B. loricata* as monophyletic; (3) *Buddleja s.l.*, including members of *Chilianthus s.l.* and *Nicodemia* as monophyletic; (4) *Buddleja s.s.*, excluding members of *Chilianthus s.l.* and *Nicodemia* as monophyletic; (5) *Buddleja* excluding only members of *Chilianthus s.l.* as monophyletic; and (6) *Buddleja* excluding only members of *Nicodemia* as monophyletic. All constrained maximum likelihood trees were compared with the unconstrained maximum likelihood tree by performing the Shimodaira–Hasegawa (SH) test in PAUP* using the RELL method and 1000 bootstraps. Because the SH test is relatively conservative, the approximately unbiased (AU) test (Shimodaira, 2002) was also performed. TREE-PUZZLE (Schmidt *et al.*, 2002) was used to compute site-log-likelihood values under the HKY + G model, which were then used to perform the AU test in CONSEL (Shimodaira & Hasegawa, 2001).

Morphological character state reconstruction

We investigated the evolution of reproductive characters that have been important in generic delimitation. For each species in Buddlejeae in our phylogenetic tree, traits were classified into categories based on species descriptions in taxonomic treatments (Leeuwenberg, 1977, 1979; Norman, 2000). For fruit type, fleshy fruits were coded as 'berry' and dry fruits were coded as 'capsule'. For stamens, those that extend outside the corolla tube were coded as 'exserted' and those that are hidden inside the corolla tube were coded as 'included'. Corolla shape could not be easily divided into categories because of continuous variation in this trait. The ratio of corolla tube length to corolla lobe length was compared to verbal descriptions from published treatments and a ratio of 1.8 was chosen as the dividing point between 'cup-shaped' (< 1.8) and 'tube-shaped' (> 1.8) corollas. Most corollas with a ratio < 1.8 are described as cup-shaped or funnelform in species descriptions and most with a ratio > 1.8 are described as tubular, cylindrical or salverform. For corolla colour, the colour of the majority of the corolla, generally including the lobes and outer tube, was classified as yellow, orange, purple or white. In many species, the corolla throat, or inner tube, has a different colour, which was not considered. For inflorescences, those with sessile flowers and peduncled cymes on a primary branch were considered 'capitate', those with sessile cymes and sessile flowers were considered 'spiciform', those with peduncled cymes and pedicellate flowers were considered 'thyrsoid', those with greater than one order of branching were considered 'paniculate' regardless of presence or absence of peduncles and pedicels and those with single-flowered cymes in a raceme were considered 'racemose' (Table A3).

All taxa were coded as having a single state for each trait, although in rare cases another state occurs at low frequency. States of taxa outside Buddlejeae were coded as missing because outgroup taxa represent large clades that typically include large variation in trait states and sampling was not sufficient to be representative. Maximum likelihood analyses were conducted under the one-rate Mk1 model in Mesquite v.2.75 (Maddison & Maddison, 2015) using the majority-rule consensus tree from Bayesian analyses of the concatenated seven-locus dataset. Bayesian analyses were conducted in BayesTraits v2.0 (Pagel & Meade, 2014) using a restricted one-rate model and the posterior distribution of trees from Bayesian analyses of the concatenated seven-locus dataset, excluding 25% burn-in. The prior for the rate was set as an exponential distribution with a mean of 10 and analyses were run for 1 000 000 generations with sampling every 1000 generations. The probabilities of trait states were averaged over generations after a burn-in of 10%.

RESULTS

Dataset characteristics

The total concatenated aligned dataset consisted of 6235 bp for each of 83 taxa, including 77 taxa in Buddlejeae. Among the characters, 2289 were variable, of which 1144 were potentially parsimony-informative. Seventy-seven taxa had sequence data for at least four of the seven loci, with 60 of these having data for at least six loci. The remaining six species had data for one or three loci and were included to increase taxonomic breadth and comprehensiveness. Characteristics of individual loci are shown in Table 2. The substitution model chosen using the AIC criterion was GTR + G for all loci.

Phylogenetic reconstructions

Topologies from maximum likelihood and Bayesian reconstructions for a dataset were generally consistent, with differences only at poorly supported nodes. We considered nodes to be strongly supported if they received support values of BP ≥ 90% and PP ≥ 0.95 and moderately supported if they received support values of 70% ≤ BP < 90% or 0.90 ≤ PP < 0.95.

Individual gene trees are shown in Supporting Information, Figure S1A–E. All individual gene trees confirm Buddlejeae to be monophyletic with strong or moderate support. The two *Gomphostigma* spp. formed a clade with strong support in all gene trees. Species that have been placed in *Nicodemia* formed a

Table 2. Characteristics of individual locus datasets

Locus	Genome	Sequenced length range (bp)	Aligned length (bp)	Variable characters (% of aligned length)	Potentially parsimony-informative characters (% of aligned length)	Taxa with sequence data [% of all taxa (n = 83)]
ETS	Nuclear	321–449	468	271 (57.9%)	170 (36.3%)	83 (100%)
PPR24	Nuclear	959–1192	1192	583 (48.9%)	327 (27.4%)	70 (84.3%)
PPR97	Nuclear	556–778	778	334 (42.9%)	164 (21.1%)	64 (77.1%)
PPR123	Nuclear	535–1276	1279	494 (38.6%)	272 (21.3%)	72 (86.7%)
trnD-trnT	Plastid	590–856	897	154 (17.2%)	61 (6.8%)	76 (91.6%)
trnS-trnfM	Plastid	522–829	889	218 (24.5%)	58 (6.5%)	48 (57.8%)
rpoA	Plastid	673–697	732	235 (32.1%)	92 (12.6%)	78 (94%)

clade with strong support in the ETS and PPR24 trees. A clade with all the Asian *Buddleja* spp. was inferred in three of the five gene trees and had strong support in the plastid tree. Species in *Buddleja s.s.* or *Chilianthus* did not form monophyletic groups in any of the five gene trees. Topologies among gene trees were not completely congruent, but no strongly supported differences occurred at deeper nodes in Buddlejeae.

The seven-locus concatenated dataset yielded congruent trees from maximum likelihood and Bayesian analyses (Fig. 1). Buddlejeae received strong support as monophyletic, as did *Gomphostigma*. Members of *Nicodemia* also formed a clade, with strong support in the Bayesian analysis. *Chilianthus* spp. did not form a monophyletic group. *Buddleja saligna* Willd. (=*Chilianthus oleaceus* Burch.) and *B. loricata* (=*C. corrugatus* Benth.) had strong support as sister species and they together with *B. dysophylla* (Benth.) Radlk. (=*C. dysophyllus* Benth.) and *B. auriculata* Benth. formed a clade, but with low support. These four species were found in a larger clade with *Gomphostigma*, which excluded *B. glomerata* H.Wendl. (=*C. lobulatus* Benth.). In *Buddleja*, there are two large well-supported clades, one comprising species from the New World and one comprising species from Asia. *Buddleja* was inferred to be paraphyletic. *Buddleja salviifolia* (L.) Lam. was sister to the rest of Buddlejeae. *Emorya* was sister to the clade of New World *Buddleja*. The Asian *Buddleja* clade was part of a well-supported clade with *B. polystachya* Fresen. and *Nicodemia*. The backbone representing relationships among these major groups generally had low support, particularly in the maximum likelihood analysis.

The species tree from the coalescent-based *BEAST analyses (Fig. 2) had a topology similar to the phylogenetic trees from the concatenated seven-locus dataset. Strongly supported relationships inferred in all analyses include monophyletic Buddlejeae, *Gomphostigma* and Asian *Buddleja*. In the species tree analysis, a clade comprising all New World *Buddleja* and *Emorya* received strong support, as did a clade comprising *B. polystachya* and *Nicodemia*; these clades also received strong support in the Bayesian analyses of concatenated data. Topological differences were at weakly supported nodes. In the species tree analyses, *Gomphostigma* was sister to the rest of Buddlejeae, but with weak support.

Topology tests

SH tests were not significant when *Chilianthus* was constrained to be monophyletic in either its narrow (P = 0.14) or broad (P = 0.29) circumscriptions. AU tests rejected the monophyly of *Chilianthus s.s.* (P < 0.01), but when *B. loricata* is included, the group could marginally not be rejected (P = 0.06). SH tests were significant when *Buddleja s.s.* (P = 0.00) or *Buddleja* without *Nicodemia* (0.03) were constrained, but not when *Buddleja s.l.* (0.43) or *Buddleja* without *Chilianthus* (P = 0.07) were constrained. In AU tests, monophyly of *Buddleja* in all of its narrower circumscriptions was rejected (P < 0.05), but the monophyly of *Buddleja s.l.* could not be rejected (P = 0.15).

Morphological character state reconstruction

Maximum likelihood and Bayesian analyses generally agreed on the highest-probability states for the nodes representing the most recent common ancestors of major clades (Table 3, Fig. 3). The most recent common ancestor of Buddlejeae was inferred to have capsular fruits, included stamens, tube-shaped, white corollas and paniculate inflorescences. The most recent common ancestor of *Nicodemia* had berries and represented the only transition to fleshy fruits. Exserted stamens and cup-shaped corollas evolved multiple times, possibly twice in African taxa with one reversal and at least once in the New World clade. Corolla colour transitioned many times: to yellow in the most recent common ancestor of the New World species; to purple in the most recent common ancestor of the

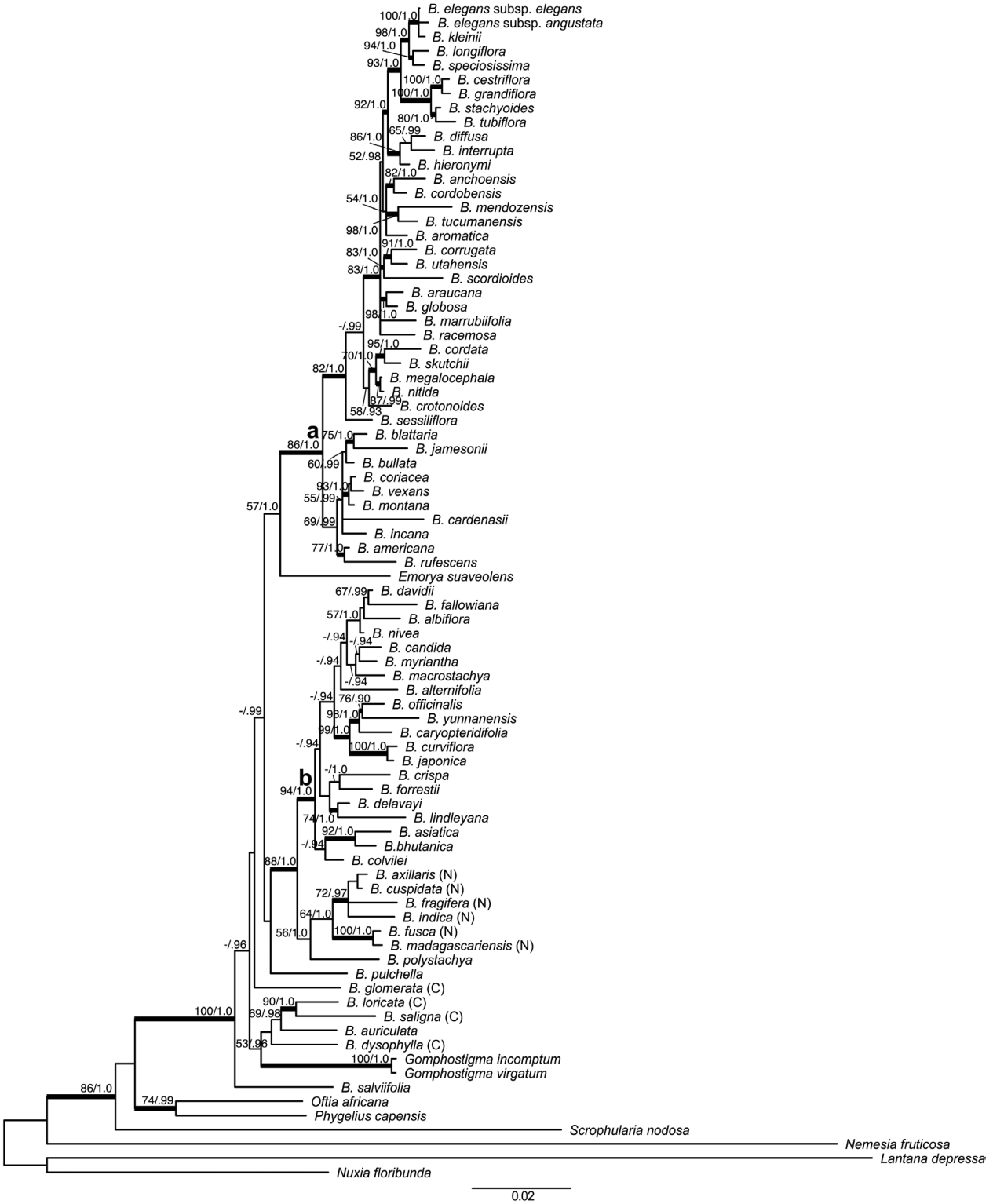


Figure 1. Majority-rule consensus phylogram from Bayesian analysis of concatenated seven-locus dataset. Values at nodes indicate support: maximum likelihood bootstrap percentage (BP)/Bayesian posterior probability (PP), if > 50% BP or 0.5 PP. Nodes with > 70% BP and 0.9 PP support are highlighted with thicker branches. Letter after species name indicates species that has also been considered a member of *Chilianthus* (C) or *Nicodemia* (N). Two nodes are marked: (a) clade of New World *Buddleja* spp. and (b) clade of Asian *Buddleja* spp.

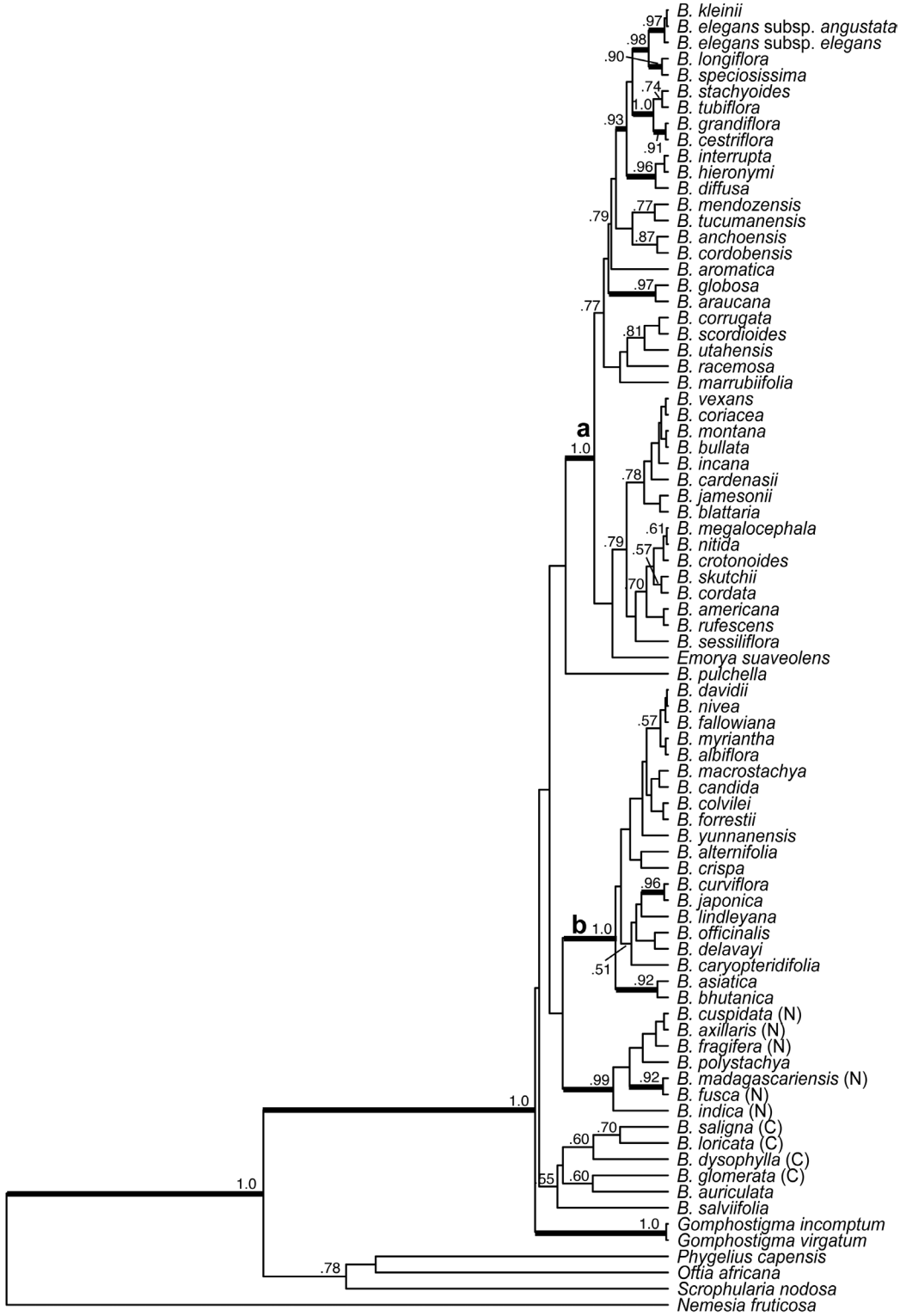


Figure 2. Maximum clade credibility tree from Bayesian multispecies coalescent analyses (*BEAST). Values at nodes indicate posterior probability (PP) support, if > 0.5. Nodes with > 0.9 PP support are highlighted with thicker branches. Letter after species name indicates species that has also been considered a member of *Chilianthus* (C) or *Nicodemia* (N). Two nodes are marked: (a) clade of New World *Buddleja* spp. and *Emorya suaveolens* and (b) clade of Asian *Buddleja* spp. Outgroups outside Scrophulariaceae are not shown.

Table 3. Probabilities of trait states at nodes corresponding to numbers in Figure 3

Node	Fruit type (berry/ capsule)	Stamen exsertion (exserted/ included)	Corolla shape (cup-shaped/ tube-shaped)	Corolla colour (yellow/ orange/purple/white)	Inflorescence (capitate/ spiciform/thyrsoid/paniculate/ racemose)
1 – Buddlejeae	0/**1** (0/**1**)	0.44/**0.56** (0.31/**0.69**)	0.32/**0.68** (0.30/**0.70**)	0.17/0.07/0.08/**0.68** (0.15/0.06/0.06/**0.72**)	0.02/0.02/0.03/**0.92**/0.02 (0.02/0.02/0.03/**0.91**/0.02)
2	0/**1** (0/**1**)	**0.51**/0.49 (**0.69**/0.31)	0.36/**0.64** (0.54/**0.46**)	0.18/0.04/0.05/**0.72** (0.27/0.06/0.07/**0.60**)	0/0/0.01/**0.98**/0 (0.01/0.01/0.04/**0.93**/0.01)
3	0/**1** (0/**1**)	0.49/**0.51** (0.41/**0.59**)	0.32/**0.68** (0.29/**0.71**)	0.25/0.05/0.07/**0.63** (**0.50**/0.10/0.12/0.28)	0/0/0.02/**0.97**/0 (0.02/0.03/0.16/**0.77**/0.02)
4	0/**1** (0/**1**)	0.14/**0.86** (0.01/**0.99**)	0.08/**0.92** (0.01/**0.99**)	0.29/0.07/0.10/**0.53** (**0.43**/0.18/0.22/0.17)	0/0/0.05/**0.94**/0 (0.03/0.05/**0.46**/0.42/0.03)
5 – section *Buddleja*	0/**1** (0/**1**)	0.14/**0.86** (0.28/**0.72**)	0.06/**0.94** (0.11/**0.89**)	**0.48**/0.10/0.09/0.33 (**0.75**/0.10/0.08/0.07)	0.01/0.01/0.06/**0.90**/0.01 (0.09/0.10/0.20/**0.53**/0.08)
6	0/**1** (0/**1**)	**0.82**/0.18 (**0.98**/02)	**0.58**/ 42 (**0.92**/0.08)	0.09/0.03/0.03/**0.85** (0.04/0.03/0.03/**0.89**)	0/0/00.01/**0.98**/0.01 (0.03/0.03/0.03/**0.83**/0.06)
7 – section *Chilianthus*	0/**1** (0/**1**)	**0.87**/0.13 (**0.97**/0.03)	**0.67**/0.33 (**0.90**/0.10)	0.04/0.01/0.02/**0.93** (0.02/0.02//02/**0.93**)	0/0/0/**0.99**/0 (0.01/0.01/0.01/**0.95**/0.01)
8	0/**1** (0/**1**)	**0.88**/0.12 (**0.74**/0.26)	**0.68**/0.32 (**0.68**/0.32)	0.02/0.01/0.01/**0.96** (0.03/0.03/0.03/**0.91**)	0/0/0/**1**/0 (0.02/0.02/0.02/**0.93**/0.02)
9	0/**1** (0/**1**)	**0.97**/0.03 (**0.99**/0.01)	**0.85**/0.15 (**0.96**/0.04)	0.02/0.01/0.01/**0.96** (0.05/0.05/0.05/**0.85**)	0/0/0/**0.99**/0 (0.04/0.04/0.04/**0.86**/0.04)
10 – section *Gomphostigma*	0/**1** (0/**1**)	**1**/0 (**1**/0)	**1**/0 (**1**/0)	0/0/0/**1.0** (0/0/0/**1.0**)	0/0/0/0/**1** (0/0/0/0/**1**)
11	0/**1** (0/**1**)	0.04/**0.96** (0/**1**)	0.05/**0.95** (0.02/**0.98**)	0.26/0.09/0.14/**0.51** (0.20/0.23/**0.33**/0.25)	0.01/0.01/0.10/**0.88**/0.01 (0.05/0.07/**0.55**/0.28/0.05)
12	0/**1** (0.03/**0.97**)	0/**1** (0/**1**)	0/**1** (0.01/**0.99**)	0.19/0.19/**0.43**/0.19 (0.08/0.32/**0.54**/0.05)	0.01/0.03/**0.82**/0.13/0.01 (0.01/0.05/**0.91**/0.02/0.01)
13 – section *Alternifoliae*	0/**1** (0/**1**)	0/**1** (0/**1**)	0/**1** (0/**1**)	0.01/0.01/**0.95**/0.03 (0/0/**0.98**/0.01)	0/0.01/**0.92**/0.06/0 (0/0.03/**0.91**/0.05/0)
14 – section *Nicodemia*	0.01/**0.99** (**0.74**/0.26)	0/**1** (0/**1**)	0/**1** (0.02/**0.98**)	0.24/**0.37**/0.25/0.14 (0.16/**0.73**/0.05/0.07)	0.01/0.07/**0.84**/0.08/0.01 (0.06/0.28/**0.56**/0.05/0.05)
15	**1**/0 (**1**/0)	0/**1** (0/**1**)	0/**1** (0.01/**0.99**)	0.38/**0.43**/0.09/0.09 (0.34/**0.53**/0.04/0.08)	0.01/0.03/**0.94**/0.01/0 (0.03/0.04/**0.89**/0.01/0.01)

Nodes representing most recent common ancestors of major clades in revised classification are indicated. The first set of numbers are from maximum likelihood analyses under an equal rates model. The second set of numbers, in parentheses, are averaged posterior probabilities from Bayesian analyses. Highest probabilities are highlighted in bold.

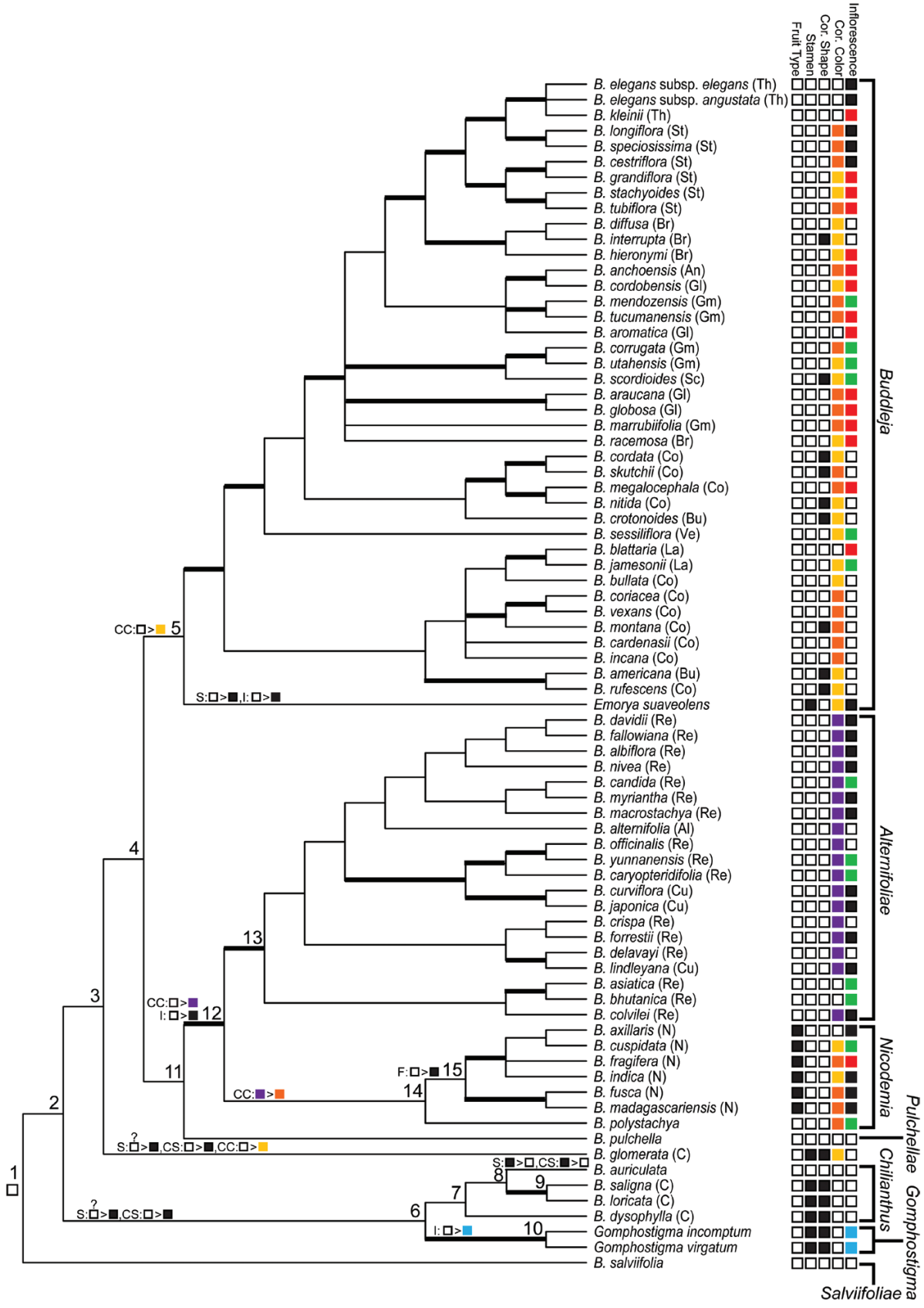
Asian *Buddleja* clade, *B. polystachya* and *Nicodemia*; and to orange in the most recent common ancestor of *B. polystachya* and *Nicodemia*. Inflorescence type has also been evolutionarily labile. The most recent common ancestor of *Gomphostigma* evolved racemose inflorescences and the most recent common ancestor of the Asian *Buddleja* clade, *B. polystachya* and *Nicodemia* probably evolved thyrsoid inflorescences.

DISCUSSION

We have inferred the first molecular phylogenetic hypotheses of species relationships in tribe Buddlejeae with extensive sampling encompassing > 70% of the species diversity in the tribe. Members of Buddlejeae form a strongly supported clade in Scrophulariaceae in all analyses of nuclear and plastid sequence data, corroborating results from previous studies of the tribe and family using plastid sequences (Olmstead & Reeves, 1995; Oxelman et al., 1999, 2005; Kornhall *et al.*, 2001). Our data from the nuclear genome also reject a close relationship between *Nuxia* and *Chilianthus* in Buddlejeae, consistent with previous analyses of plastid data (Oxelman et al., 1999, 2005).

Generic circumscription and relationships

Buddleja, in any of its previous circumscriptions, is paraphyletic. There is strong evidence from both concatenated and species tree analyses that *Emorya* and


Inflorescence
Cor. Color
Cor. Shape
Stamen
Fruit Type
B. elegans subsp. elegans (Th)
B. elegans subsp. angustata (Th)
B. kleinii (Th)
B. longiflora (St)
B. speciosissima (St)
B. cestriflora (St)
B. grandiflora (St)
B. stachyoides (St)
B. tubiflora (St)
B. diffusa (Br)
B. interrupta (Br)
B. hieronymi (Br)
B. anchoensis (An)
B. cordobensis (Gl)
B. mendozensis (Gm)
B. tucumanensis (Gm)
B. aromatica (Gl)
B. corrugata (Gm)
B. utahensis (Gm)
B. scordioides (Sc)
B. araucana (Gl)
B. globosa (Gl)
B. marrubiifolia (Gm)
B. racemosa (Br)
B. cordata (Co)
B. skutchii (Co)
B. megalocephala (Co)
B. nitida (Co)
B. crotonoides (Bu)
B. sessiliflora (Ve)
B. blattaria (La)
B. jamesonii (La)
B. bullata (Co)
B. coriacea (Co)
B. vexans (Co)
B. montana (Co)
B. cardenasii (Co)
B. incana (Co)
B. americana (Bu)
B. rufescens (Co)
Emorya suaveolens
B. davidii (Re)
B. fallowiana (Re)
B. albiflora (Re)
B. nivea (Re)
B. candida (Re)
B. myriantha (Re)
B. macrostachya (Re)
B. alternifolia (Al)
B. officinalis (Re)
B. yunnanensis (Re)
B. caryopteridifolia (Re)
B. curviflora (Cu)
B. japonica (Cu)
B. crispa (Re)
B. forrestii (Re)
B. delavayi (Re)
B. lindleyana (Cu)
B. asiatica (Re)
B. bhutanica (Re)
B. colvilei (Re)
B. axillaris (N)
B. cuspidata (N)
B. fragifera (N)
B. indica (N)
B. fusca (N)
B. madagascariensis (N)
B. polystachya
B. pulchella
B. glomerata (C)
B. auriculata
B. saligna (C)
B. loricata (C)
B. dysophylla (C)
Gomphostigma incomptum
Gomphostigma virgatum
B. salviifolia
Buddleja
Alternifoliae
Nicodemia
Pulchellae
Chilianthus
Gomphostigma
Salviifoliae
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
CC:□>■
S:□>■,I:□>■
CC:□>■
I:□>■
CC:■>
F:□>■
S:?□>■,CS:□>■,CC:□>■
S:■>□,CS:■>□
S:?□>■,CS:□>■
I:□>■

Nicodemia are derived from within *Buddleja*. New World *Buddleja* spp. are more closely related to *Emorya* than they are to the Asian or African species. *Buddleja polystachya* is more closely related to *Nicodemia* than to other *Buddleja* spp. Reconstructions from both analyses suggest that *Chilianthus* is also derived from within *Buddleja*, though with lower support; and in the concatenated analyses, *Buddleja* is also paraphyletic with respect to *Gomphostigma*. *Buddleja auriculata* is inferred to be more closely related to *Chilianthus* and, in the concatenated analyses also to *Gomphostigma*, than to other *Buddleja* spp. Topology tests rejected the monophyly of *Buddleja* when it excluded the species in either or both *Chilianthus* and *Nicodemia*, but could not reject a more inclusive clade comprising members of all three genera, leaving only *Emorya* and *Gomphostigma* outside the group. However, with strong support across phylogenetic analyses for the close relationship between New World *Buddleja* and *Emorya* and some support for a close relationship among *B. auriculata*, *Chilianthus* and *Gomphostigma*, we believe there is sufficient evidence to assert that *Buddleja* is not monophyletic even in this broadest circumscription.

The monophyly of *Chilianthus* was not supported in our analyses. Three species in the group, *B. dysophylla*, *B. loricata* and *B. saligna*, were found in a clade (hereafter, called core *Chilianthus*), which also included *B. auriculata* in analyses with concatenated data. A sister species relationship between *B. loricata* and *B. saligna* was recovered in both concatenated and species tree analyses, with strong support in the concatenated analyses. The fourth member of the *Chilianthus* group, *B. glomerata*, was consistently outside of this clade, but its precise phylogenetic position is equivocal. In the species tree analysis, *B. glomerata* together with *B. auriculata* is sister to core *Chilianthus*, whereas in the concatenated data analyses, it is more distantly related. Topology tests indicated that a monophyletic *Chilianthus* including all four species could not be rejected, but only marginally. Leeuwenberg (1979) removed *B. loricata* from the *Chilianthus* group because its stamens have shorter filaments and are barely exserted. A clade comprising the remaining three species in *Chilianthus*, *B. dysophylla*, *B. glomerata* and *B. saligna* was never recovered and topology tests rejected the monophyly of this group. The three species in core *Chilianthus* share several traits besides the typical *Chilianthus* floral morphology of short, cup-shaped corollas, long, exserted stamens and paniculate inflorescences. They also have white or cream corollas with an orange or maroon throat, pubescence on the inside of the corolla and a reticulate seed coat. *Buddleja glomerata* has the typical *Chilianthus* floral morphology, but has yellow corollas, is glabrous inside the corolla tube and has seeds with a smooth coat. *Buddleja auriculata*, which was found to be closely related to core *Chilianthus* in both analyses, is similar morphologically in having white corollas with an orange throat, pubescence inside the corolla tube and a reticulate seed coat, but the corolla shape is long and tubular and the stamens are included (Leeuwenberg, 1979). *Gomphostigma* was resolved to be closely related to core *Chilianthus* in the concatenated analyses, but without strong support. Both *Gomphostigma* and *Chilianthus* have short corolla tubes and exserted stamens, but *Gomphostigma* is distinct in having racemose inflorescences and corollas that are pure white (Leeuwenberg, 1977; Oxelman *et al.*, 2004). Relationships among core *Chilianthus*, *B. glomerata*, *B. auriculata* and *Gomphostigma* were poorly supported and inconsistent in our analyses and additional data will be required to fully resolve their history.

Nicodemia spp. formed a clade in the phylogenetic trees inferred from concatenated data, with strong support in the Bayesian analyses. They are unique in

Figure 3. Majority-rule consensus cladogram from Bayesian analysis of concatenated seven-locus dataset. Nodes with > 70% bootstrap percentage and 0.9 posterior probability support are highlighted with thicker branches. Single letter after species name indicates species that has also been considered a member of *Chilianthus* (C) or *Nicodemia* (N). Two-letter codes after species name indicate infrageneric classification in *Buddleja*. For New World species, this follows Norman (2000): series *Anchoenses* (An), *Brachiatae* (Br), *Buddleja* (Bu), *Cordatae* (Co), *Globosae* (Gl), *Glomeratae* (Gm), *Lanatae* (La), *Scordioides* (Sc), *Stachyoides* (St), *Thyrsoides* (Th), *Verticillatae* (Ve). For Asian species, this follows Marquand (1930): series *Alternifoliae* (Al), *Curviflorae* (Cu), *Rectiflorae* (Re). Circumscription of sections in revised classification of *Buddleja* shown at far right. Coloured boxes indicate trait states of taxa. First column from left – fruit type: capsule (white) or berry (black); second column – stamen exsertion: included (white) or exserted (black); third column – corolla shape: tube-shaped (white) or cup-shaped (black); fourth column – corolla colour corresponds to box colour: white, yellow, orange or purple; fifth column – inflorescence type: paniculate (white), thyrsoid (black), spiciform (green), capitate (red) or racemose (blue). Inferred ancestral states of Buddlejeae indicated by white box at root: capsule, included stamens, tube-shaped, white corolla and paniculate inflorescence. Major transitions between states are indicated above branches where inferred (F = fruit type, S = stamen exsertion, CS = corolla shape, CC = corolla colour, I = inflorescence). Question mark (?) above transition indicates equivocal reconstruction. Transitions within sections *Buddleja*, *Alternifoliae* and *Nicodemia* are generally not indicated. Numbers at nodes correspond to those in Table 3. Outgroups outside Buddlejeae are not shown.

Buddlejeae in having indehiscent fleshy fruits, usually considered berries, and they share a main distribution in Madagascar with some species also found in surrounding islands and eastern Africa. *Buddleja polystachya*, a species from eastern Africa and the Arabian peninsula not previously assigned to *Nicodemia*, was resolved to be closely related to *Nicodemia* in all analyses. It is sister to *Nicodemia* in the concatenated analyses and is nested in *Nicodemia* in the species tree analyses. *Buddleja polystachya* shares a yellow to orange corolla with many *Nicodemia* spp. and they all have thyrsoid inflorescences, which differ from the paniculate inflorescences found in the basal grade of African Buddlejeae. The fruits of *B. polystachya* may represent an intermediate condition between the dry, septicidally dehiscent capsules of most Buddlejeae and the fleshy, indehiscent berries in *Nicodemia*; its dry fruits are partially indehiscent, with valves described as 'not torn' (Leeuwenberg, 1979). Some members of *Nicodemia* have at times been placed in other segregate genera. *Adenoplea* is differentiated by its four-celled ovaries, as opposed to the two-celled ovaries found in the rest of Buddlejeae and most of Scrophulariaceae (Leenhouts, 1962; Leeuwenberg, 1979). The two species we sampled with four-celled ovaries, *B. fusca* Baker and *B. madagascariensis*, consistently formed a well-supported clade. The two other species with this trait, *B. acuminata* Poir. and *B. sphaerocalyx* Baker, need to be sampled to determine their phylogenetic position. *Adenoplusia* is distinct in having drupe-like fruits with a chartaceous endocarp (Bruce & Lewis, 1960). All of its species have been combined with *B. axillaris* Willd., which is in the *Nicodemia* clade.

The two *Gomphostigma* spp. received strong support as sister taxa in all analyses. Both species are from southern Africa and share a distinct suite of morphological traits, including racemose inflorescences, cup-shaped corollas and exserted stamens (Oxelman *et al.*, 2004). *Gomphostigma* is part of a basal grade of African members of Buddlejeae, although its exact position is not well supported. In the concatenated analyses, it is sister to a clade consisting of core *Chilianthus* and *B. auriculata*, whereas in the species tree analysis, it forms the sister group to the rest of the tribe.

Only one of two *Emorya* spp. was sampled in this study, so the monophyly of this group could not be assessed. Both species are distributed in north-central Mexico, but the unsampled species *E. rinconensis* Mayfield is known from only a single locality in Coahuila state. The two species share several traits, including long-tubular corollas, exserted styles and exserted stamens with long filaments, that suggest a close relationship, but there are also notable differences. Inflorescences are thyrsoid in *E. suaveolens* Torr. but racemose in *E. rinconensis* and pollen is tetracolporate in *E. suaveolens* but tricolporate in *E. rinconensis* (Mayfield, 1999). *Emorya suaveolens* forms a clade with New World *Buddleja* spp. with strong support. In the concatenated analyses, *Emorya* is sister to all New World *Buddleja*, whereas in the species tree analysis, it is sister to one of two main New World clades. Despite noted similarities in floral morphology, including a long corolla tube, between *Emorya* and members of the South American *Buddleja* series *Stachyoides* (Norman, 2000), a close relationship between these two groups was not found.

Infrageneric relationships in *Buddleja*

Relationships among *Buddleja* spp. show strong geographical signal, particularly at the continental level. Based on our results, infrageneric classification schemes in *Buddleja* (Table 1) that ignore geographical distribution and group species from separate continents in the same taxon do not reflect evolutionary relationships. Most systematic studies in *Buddleja* have been regionally focused and the composition of their proposed infrageneric taxa has been limited to species from a single region. However, the classifications of Bentham (1846), Marquand (1930) and Leeuwenberg (1979) included several infrageneric groups with distributions spanning multiple continents, which are not supported by our results (e.g. section *Neemda* Benth., subsection *Glomeratae* Benth., subsection *Thyrsoideae* Benth., subsection *Macrothrysae* Benth., series *Rectiflorae* Marquand and section *Neemda sensu* Leeuwenberg).

Southern African members of Buddlejeae, including *B. salviifolia*, *B. auriculata*, *Chilianthus* and *Gomphostigma*, make up a basal grade. *Buddleja salviifolia* is resolved as sister to all other species of Buddlejeae in the concatenated analyses, whereas it is in a clade with *B. auriculata* and *Chilianthus* in the species tree analyses. The remaining species in Buddlejeae are found in two major clades. One of them comprises the rest of the Old World species and forms two groups: a clade with all the Asian *Buddleja* and a clade with *Nicodemia* and *B. polystachya,* species from Madagascar and eastern Africa. The other major clade consists of all the New World species. The position of *B. pulchella* N.E.Br. from southern and eastern Africa is not well supported, but it may be sister to one of these two major clades. The relationships of *B. auriculata* with *Chilianthus* and *Gomphostigma* and of *B. polystachya* with *Nicodemia* are discussed in the preceding section on generic relationships.

Bentham (1846), Marquand (1930) and Leeuwenberg (1979) placed Asian *Buddleja* spp. in groups with

species from Africa and sometimes Madagascar and the New World, for example subsection *Glomeratae* Benth., subsection *Macrothrysae* Benth., series *Rectiflorae* Marquand and section *Neemda sensu* Leeuwenberg, none of which was supported by our analyses. Marquand (1930) focused mostly on Asian *Buddleja* and he proposed an infrageneric classification, which was generally followed by Li (1982), based on differences in leaf arrangement and floral morphology. *Buddleja alternifolia*, which is unique in the genus in having alternate leaves, is the only currently accepted species in section *Alternifoliae* Kränzl. The remaining species are considered synonyms of *B. alternifolia* or *B. asiatica* Lour., both of which are in the Asian *Buddleja* clade. (Although the type of section *Alternifoliae* is *B. amentacea* Kränzl., synonymized with *B. asiatica*, this has not always been recognized by previous taxonomists when circumscribing the group). Series *Curviflorae* Marquand, which is distinguished by curved corolla tubes, was partly supported by our phylogenetic reconstructions. *Buddleja curviflora* Hook. & Arn. and *B. japonica* Hemsl. are strongly supported as sister species in concatenated and species tree analyses, but the position of the third species, *B. lindleyana* Fortune, is uncertain. In the species tree analyses, it forms a clade with the other two species, but in the concatenated analyses, they are not closely related. The varying position of *B. lindleyana* in the different gene trees suggests that introgression or retention of ancestral polymorphism may be a factor (Maddison, 1997). Series *Rectiflorae*, which includes the majority of the Asian species and is characterized by straight corolla tubes, is paraphyletic with respect to *Alternifoliae* and *Curviflorae* (Fig. 3).

For New World *Buddleja*, the classification of Bentham (1846) included several groups, for example, subsection *Paniculatae* Benth., subsection *Globosae* Benth., subsection *Verticillatae* Benth., subsection *Stachyoides* Benth., none of which was supported as monophyletic in our phylogenetic analyses. The most recent and comprehensive study of New World *Buddleja* by Norman (2000) included 12 series based on morphology and ecogeography (Fig. 3). Our study included representatives from all series, except the monotypic *Oblongae* E.M.Norman. We sampled multiple species for each included series, except *Scordioides* E.M.Norman, *Verticillatae* (Benth.) E.M.Norman and the monotypic *Anchoenses* E.M.Norman, which enabled us to begin investigating the monophyly and relationships among these infrageneric groups.

The monophyly of series *Thyrsoides* (Benth.) E.M.Norman, including *B. elegans* and *B. kleinii* E.M.Norman & L.B.Sm., was supported by the phylogenetic analyses. Most of the species in series *Stachyoides* (Benth.) E.M.Norman also formed a well-supported clade. However, *B. longiflora* Brade and *B. speciosissima* Taub from *Stachyoides* are more closely related to series *Thyrsoides*, to which they form the sister group. These two species differ from the rest of series *Stachyoides*, but is similar to series *Thyrsoides*, in having subcoriaceous rather than membranaceous leaves and pedicellate rather than sessile flowers (Norman, 2000). Series *Thyrsoides* and *Stachyoides* are most closely related to each other and together they are sister to series *Brachiatae* E.M.Norman. These three series are mainly South American in distribution, occurring in south-eastern Brazil and the Andes (Norman, 2000). Species in *Brachiatae* form a strongly supported clade when *B. racemosa* Torr., the only North American species in the group, is excluded.

Series *Glomeratae* (Benth.) E.M.Norman is inferred to be polyphyletic. *Buddleja mendozensis* Gillies ex Benth. and *B. tucumanensis* Griseb., from Argentina and Bolivia, are sister species, but they are not closely related to the other members of series *Glomeratae*, which are mostly North American. *Buddleja mendozensis* and *B. tucumanensis* are more closely related to *B. anchoensis* Kuntze from series *Anchoenses* and *Buddleja aromatica* J.Rémy and *B. cordobensis* Griseb. from series *Globosae* (Benth.) E.M.Norman. These five South American species share similar seed morphology and sessile flowers (Norman, 2000). The other two species in series *Globosae*, *B. araucana* Phil. and *B. globosa* from Chile and Argentina, form a clade with strong support. North American members of series *Glomeratae*, *B. corrugata* M.E.Jones and *B. utahensis* Coville, are more closely related to the North American species *B. scordioides* Kunth in series *Scordioides*.

Series *Cordatae* E.M.Norman is paraphyletic with respect to series *Buddleja*, *Lanatae* E.M.Norman and *Verticillatae*. Members of these four series fall into two clades according to geographical distribution. A North American clade includes *B. cordata* Kunth, *B. megalocephala* Donn.Sm., *B. nitida* Benth. and *B. skutchii* C.V.Morton from *Cordatae*, *B. crotonoides* A.Gray from *Buddleja* and, in the species tree analyses, *B. sessiliflora* Kunth from *Verticillatae*. The other clade is South American and contains *B. cardenasii* Standl. ex E.M.Norman, *B. coriacea*, *B. incana*, *B. montana* Britton and *B. vexans* Kraenzl. & Loes. ex E.M.Norman from *Cordatae* and *B. blattaria* and *B. jamesonii* Benth. from series *Lanatae*. *Buddleja americana* L., the range of which spans North and South America, and *B. rufescens* from Peru are sister species, but their phylogenetic position is equivocal. They fall with the South American clade in the concatenated analysis and with the North American clade in the species tree analysis. The distant relationship between *B. americana* and *B. crotonoides* indicates that series *Buddleja* is polyphyletic. The sampled species in series *Lanatae* form a strongly supported clade. All species from series *Buddleja*, *Cordatae* and *Verticillatae* for which ploidy

has been determined are polyploid (Norman, 2000). There are no published chromosome counts for any species in series *Lanatae*, but we predict based on these relationships that they are also polyploid.

Trait evolution

The evolution of morphological traits traditionally used to characterize genera in Buddlejeae was investigated (Fig. 3). For *Nicodemia*, fleshy berries remain a synapomorphy and useful distinguishing character. Fruit type evolved once from dry capsules to fleshy berries in the most recent common ancestor of this group.

The other reproductive traits traditionally used to delimit genera have been evolutionary labile and evolved independently to similar states multiple times. Although traditionally used to distinguish *Buddleja* from other genera, included stamens and tube-shaped corollas are inferred to be symplesiomorphic in Buddlejeae. Exserted stamens and cup-shaped corollas evolved at least once in the African species. They may have evolved independently in the ancestor of *B. glomerata* and in the ancestor of a clade comprising core *Chilianthus*, *Gomphostigma* and *B. auriculata*, with reversals occurring in *B. auriculata*. Alternatively, exserted stamens and cup-shaped corollas may have evolved in an earlier ancestor of core *Chilianthus*, *Gomphostigma*, *B. auriculata* and *B. glomerata*, with reversals occurring in *B. auriculata* and in the ancestor of *Nicodemia* and the Asian and New World *Buddleja*. Support is low for some relationships among the African species, including on the backbone of the tree, and trait states at several of these nodes are equivocal. Phylogenetic analyses with more data to increase resolution and support for the relationships among these groups are needed to fully understand the evolution of these traits. Additionally, stamen exsertion and corolla shape underwent independent transitions in the New World clade. Exserted stamens evolved once in *Emorya* and cup-shaped corollas evolved multiple times in New World *Buddleja*.

Ancestral white corolla colour was retained in the basal African grade, including *B. salviifolia*, *B. auriculata*, core *Chilianthus* and *Gomphostigma*. Corolla colour evolved from white to purple in the most recent common ancestor of *Nicodemia*, *B. polystachya* and Asian *Buddleja* and then transitioned to orange in the most recent common ancestor of *Nicodemia* and *B. polystachya*. In *Nicodemia*, transitions to white and yellow also occurred and in Asian *Buddleja*, the purple corolla colour was mostly retained, with a single reversal back to white. In the ancestor of *Emorya* and New World *Buddleja*, corolla colour evolved from white to yellow. Yellow corollas were retained in many New World *Buddleja*, but there have also been multiple transitions to white and orange. Yellow corollas evolved independently in *B. glomerata*.

In Buddlejeae, the ancestral inflorescence form was the highly branched paniculate type and reductions in branching occurred multiple times. In *Gomphostigma*, the inflorescence was reduced to a racemose form; that inflorescence type, cup-shaped corollas and exserted stamens remain a useful suite of characters for recognizing the clade. Paniculate inflorescences were retained in the rest of the basal African grade and in the ancestor of *Emorya* and New World *Buddleja*. In the New World clade, multiple independent reductions in branching and loss of peduncles and/or pedicels produced a range of thyrsoid, capitate, spicate and racemose inflorescences. In the ancestor of *Nicodemia*, *B. polystachya* and Asian *Buddleja*, inflorescence form was reduced to the thyrsoid type. Further reductions of pedicels and/or peduncles resulted in spicate or capitate inflorescences in some species and reversals to paniculate inflorescences also occurred in Asian *Buddleja*.

Classification

We present here a revised classification for Buddlejeae reflecting our phylogenetic results. Our general philosophy is to name supported monophyletic groups in order to facilitate communication and understanding of relationships. Although relationships among named clades are not all strongly supported, they are exclusive of other clades and represent distinct lineages.

Only the genus *Buddleja* is maintained and its circumscription is expanded. Evidence shows that *Buddleja* as previously circumscribed is paraphyletic. Despite rendering *Buddleja* redundant with Buddlejeae, we take this conservative approach to the taxonomy because Buddlejeae is clearly monophyletic and uncertainty in some relationships between *Buddleja* and other small lineages (e.g. *Gomphostigma*) precludes accepting previously recognized segregate genera. All species in *Chilianthus*, *Gomphostigma*, *Nicodemia* and *Emorya* are combined with *Buddleja*. Species in *Chilianthus*, *Gomphostigma* and *Nicodemia* already have synonyms in *Buddleja*, but two new names are proposed for the species in *Emorya*.

Seven groups of species consistently obtained in analyses are recognized at the sectional rank in *Buddleja* (Fig. 3). Two new monotypic sections are recognized for *B. salviifolia* and *B. pulchella*. *Gomphostigma* is lowered from the genus to sectional rank. Section *Chilianthus* comprises *B. dysophylla*, *B. loricata*, *B. saligna* and *B. auriculata*. The position of *B. glomerata* remains equivocal and may be included in this section if additional evidence supports this relationship. Section *Nicodemia* is expanded to include *B. polystachya* in addition to the eight species traditionally

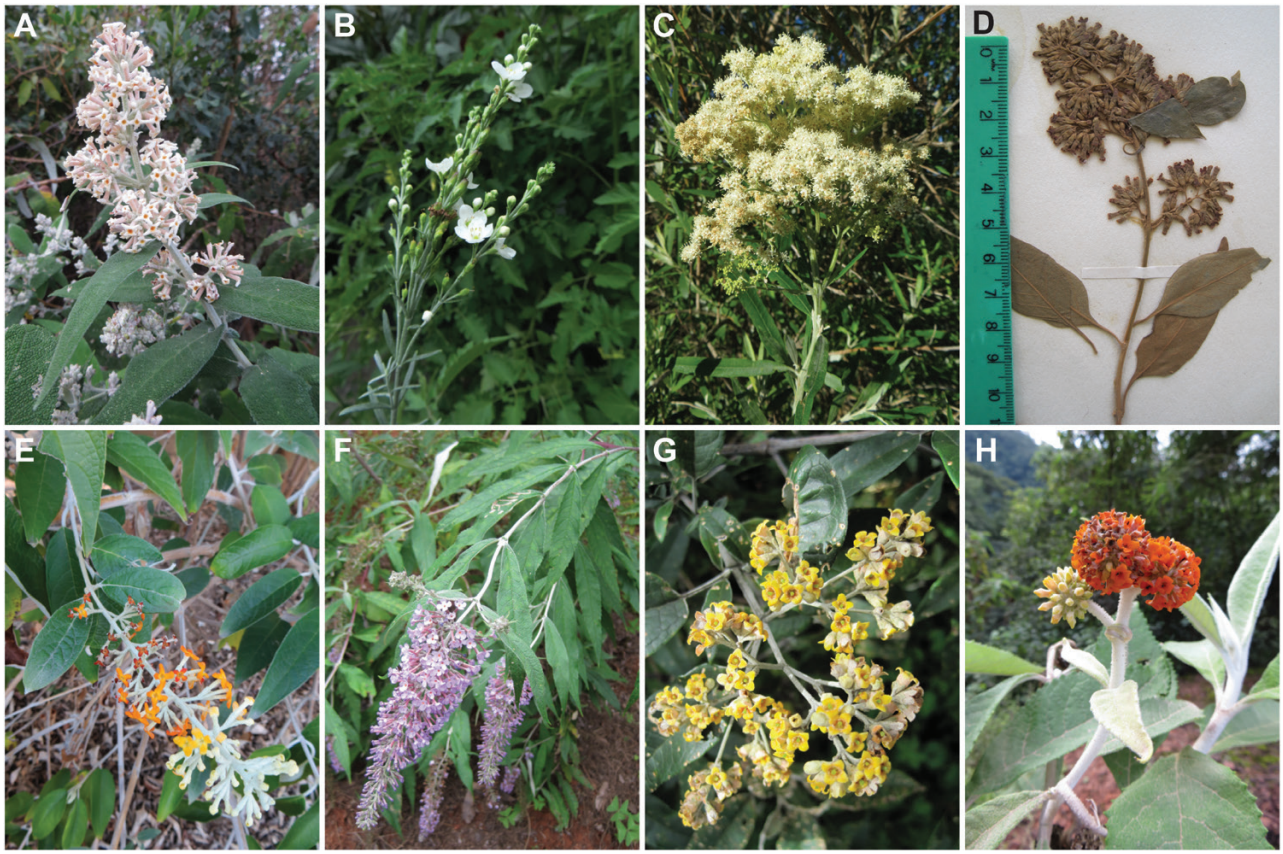


Figure 4. Representatives of seven sections of *Buddleja* in revised classification of Buddlejeae. (A) *Buddleja salviifolia*, section *Salviifoliae*, (B) *Buddleja virgata*, section *Gomphostigma*, (C) *Buddleja saligna*, section *Chilianthus*, (D) *Buddleja pulchella*, section *Pulchellae*, (E) *Buddleja madagascariensis*, section *Nicodemia*, (F) *Buddleja davidii*, section *Alternifoliae*, (G) *Buddleja nitida*, section *Buddleja*, (H) *Buddleja anchoensis*, section *Buddleja*. All photographs by J.H. Chau.

in the group. Section *Alternifoliae* is expanded to include all species of *Buddleja* distributed in Asia. All species found in North and South America are placed in section *Buddleja*, including members of *Emorya*. Circumscriptions and species names in revised classification are listed in Table A4 and select representatives of sections are shown in Figure 4.

I. *Buddleja* L., Sp. Pl. 1: 112. 1753. Type: *Buddleja americana* L.
Inflorescence paniculate, thyrsoid, capitate, spicate or racemose. Flowers with tube- or cup-shaped corollas and included or exserted stamens. Ovary two- or four-celled. Fruit a capsule or berry. Distribution: Africa, Madagascar, Asia, North America, South America.

1. Section ***Salviifoliae*** J.H.Chau, **sect. nov.** Type: *Buddleja salviifolia* (L.) Lam.
Inflorescence paniculate. Corolla white or lilac to purple, with deep orange throat; tube-shaped. Stamens included. Ovary two-celled. Fruit a capsule. Distribution: southern and eastern Africa.
2. Section ***Gomphostigma*** (Turcz.) J.H.Chau, **stat. nov.** Basionym: *Gomphostigma* Turcz., Bull. Soc. Nat. Mosc. 16: 53. 1843. Type: *Gomphostigma scoparioides* Turcz. = *Buddleja virgata* L.f.
Inflorescence racemose. Corolla white, cup-shaped. Stamens exserted. Ovary two-celled. Fruit a capsule. Distribution: southern Africa.
3. Section *Chilianthus* (Burch.) Leeuwenberg, Meded. Landbouwhogeschool Wageningen 79 (6): 7. 1979. Type: *Chilianthus oleaceus* Burch. = *Buddleja saligna* Willd.

   Inflorescence paniculate. Corolla white or cream, with orange or mauve throat; cup- or tube-shaped. Stamens exserted or included. Ovary two-celled. Fruit a capsule. Distribution: southern Africa.
4. Section ***Pulchellae*** J.H.Chau, **sect. nov.** Type: *Buddleja pulchella* N.E.Br.
Inflorescence paniculate. Corolla white, yellow or pale orange, with yellow or orange throat; tube-shaped. Stamens included. Ovary two-celled. Fruit a capsule. Distribution: southern and eastern Africa.

5. Section *Nicodemia* (Tenore) Leeuwenberg, Meded. Landbouwhogeschool Wageningen 79 (6): 9. 1979. Type: *Nicodemia diversifolia* (Vahl) Tenore = *Buddleja indica* Lam.
   Inflorescence thyrsoid, capitate or spicate. Corolla white, yellow or orange; tube-shaped. Stamens included. Ovary two- or four-celled. Fruit a berry or capsule. Distribution: Madagascar, eastern Africa, Arabian Peninsula.
6. Section *Alternifoliae* Kränzl., Bull. Jard. Imp. Bot. Petersb. 8 (4): 89. 1913. Type: *Buddleja amentacea* Kränzl. = *Buddleja asiatica* Lour.
   Inflorescence thyrsoid, spicate or paniculate. Corolla purple or white, often with orange throat; tube-shaped. Stamens included. Ovary two-celled. Fruit a capsule. Distribution: Asia.
7. Section *Buddleja*. Type: *Buddleja americana* L.
   Inflorescence paniculate, thyrsoid, capitate, spicate or racemose. Corolla white, yellow or orange; tube- or cup-shaped. Stamens included or exserted. Ovary two-celled. Fruit a capsule. Distribution: North and South America.
   i. ***Buddleja normaniae*** J.H.Chau, **nom. nov.**
      Basionym: *Emorya suaveolens* Torr., Rep. U.S. Mex. bound. 2(1): 121 t. 36. 1859.
      The epithet recognizes the work of Eliane Norman in the study of Buddlejeae, especially its New World members.
   ii. ***Buddleja rinconensis*** (Mayfield) J.H.Chau, **comb. nov.**
      Basionym: *Emorya rinconensis* Mayfield, Sida 18: 693–699. 1999.

CONCLUSIONS

Buddlejeae are among the larger and most broadly distributed tribes in Scrophulariaceae. We present the first phylogenetic reconstruction of relationships in the tribe based on multiple independent genetic markers and with extensive and representative taxonomic sampling. We show that *Buddleja* is paraphyletic with respect to *Chilianthus*, *Nicodemia*, *Emorya* and probably *Gomphostigma* and the traits used to distinguish *Buddleja*, namely flowers with included stamens and capsular fruits, are plesiomorphic. Additional data and analyses will be required to definitively resolve some relationships that remain poorly supported and their implications for patterns in trait evolution, including among some of the African taxa and in the Asian and New World clades. Extensive polyploidy in the Asian and New World clades complicates analyses due to uncertainties in orthology assessment and separation of copy sequences through cloning or next-generation sequencing methods will be necessary. Our revised classification clarifies evolutionary relationships in Buddlejeae and can serve as a framework for future investigations on evolution in this diverse group.

ACKNOWLEDGEMENTS

We thank E. Norman for insight on the study group and generosity in sharing specimens; A. Barker, N. Barker, A.L. Cazé, A. Fregonezi, J. Fregonezi, S. Leiva, D. Morales-Briones, F. Salimena, D. Tank, S. Uribe-Convers, M. Zapata Cruz and X. Zhang for assistance in the field; B. Craig, S. Harkavy, P. Lu-Irving and H. Marx for providing sequence data; A. O'Brien and B. Meersman for assistance with primer design for PPR loci and M. Moore for sharing specimens. Plant material or DNA from their collections was provided by the following botanical gardens and herbaria: Arnold Arboretum, Harry P. Leu Gardens, Los Angeles County Arboretum, Kunming Botanical Garden, Rancho Santa Ana Botanic Garden, San Francisco Botanical Garden, University of California Botanical Garden, UC-Davis Arboretum, University of Washington Botanic Gardens, CAS, DAV, K, KUN, LPB, MO, NBG, SI, TEX and UC/JEPS. This study was supported by National Science Foundation (NSF) grants DEB-0542493, DEB-1020369 to RGO and DEB-1311111 to RGO and JHC; a NSF Graduate Research Fellowship to JHC; graduate student grants from the Botanical Society of America, American Society of Plant Taxonomists and the Society of Systematic Biologists to JHC; the Giles Botanical Field Research Award and Sargent Award from the University of Washington Department of Biology to JHC, a Agencia Nacional de Promoción Científica y Tecnológica grant PICT 2010-0648 to NOL and National Natural Science Foundation of China (NSFC) grant 30970192 to W-BS.

REFERENCES

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

APG IV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181:** 1–20.

Baldwin BG, Markos S. 1998. Phylogenetic utility of the external transcribed spacer (ETS) of 18S–26S rDNA: congruence of ETS and ITS trees of *Calycadenia* (Compositae). *Molecular Phylogenetics and Evolution* **10:** 449–463.

Beardsley PM, Yen A, Olmstead RG. 2003. AFLP phylogeny of *Mimulus* section *Erythranthe* and the evolution of hummingbird pollination. *Evolution* **57:** 1397–1410.

Bentham G. 1835. *Scrophularinae indicae*. London: James Ridgeway.

Bentham G. 1846. Scrophulariaceae. In: De Candolle A, ed. *Prodromus systematis naturalis regni vegetabilis pars X*. Paris: Victor Masson, 432–447.

Bentham G. 1857. Notes on Loganiaceae. *Journal of the Linnean Society, Botany* **1:** 52–144.

Bentham G, Hooker JD. 1876. *Genera plantarum, Vol. 2*. London: Reeve & Co.

Bruce EA, Lewis J. 1960. Loganiaceae. In: Hubbard CE, Milne-Redhead K, eds. *Flora of tropical East Africa*. London: Crown Agents for Oversea Governments and Administrations, 1–45.

Campbell V, Lapointe FJ. 2009. The use and validity of composite taxa in phylogenetic analysis. *Systematic Biology* **58:** 560–572.

Cronquist A. 1981. *An integrated system of classification of flowering plants*. New York: Columbia University Press.

Dahlgren R. 1983. General aspects of angiosperm evolution and macrosystematics. *Nordic Journal of Botany* **3:** 119–149.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9:** 772.

Demesure B, Sodzi N, Petit RJ. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* **4:** 129–131.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin, Botanical Society of America* **19:** 11–15.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29:** 1969–1973.

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

Hakki MI. 1980. Embryology. In: Hiepko P, Melchior H, eds. *Die natürlichen Pflanzenfamilien, Band 28bI, Angiospermae: Ordnung Gentianales, Fam. Loganiaceae, ed. Leeuwenberg AJM*. Berlin: Duncker & Humblot, 192–201.

Hutchinson J. 1973. *The families of flowering plants, 3rd edn*. London: Oxford University Press.

Jensen SR, Nielsen BJ, Dahlgren R. 1975. Iridoid compounds, their occurrence and systematic importance. *Botaniska Notiser* **128:** 148–180.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30:** 772–780.

Kornhall P, Bremer B. 2004. New circumscription of the tribe Limoselleae (Scrophulariaceae) that includes the taxa of the tribe Manuleeae. *Botanical Journal of the Linnean Society* **146:** 453–467.

Kornhall P, Heidari N, Bremer B. 2001. Selagineae and Manuleeae, two tribes or one? Phylogenetic studies in the Scrophulariaceae. *Plant Systematics and Evolution* **228:** 199–218.

Leenhouts PW. 1962. Loganiaceae. In: van Steenis CGGJ, ed. *Flora Malesiana, Vol. 6*. Groningen: Wolters-Noordhoff, 293–387.

Leeuwenberg AJM. 1977. The Loganiaceae of Africa XVI. *Gomphostigma* Turcz. *Mededelingen Landbouwhogeschool Wageningen* **77:** 15–30.

Leeuwenberg AJM. 1979. The Loganiaceae of Africa XVIII. *Buddleja* II. Revision of the African and Asiatic species. *Mededelingen Landbouwhogeschool Wageningen* **79:** 1–163.

Leeuwenberg AJM, Leenhouts PW. 1980. Taxonomy. In: Hiepko P, Melchior H, eds. *Die natürlichen Pflanzenfamilien, Band 28bI, Angiospermae: Ordnung Gentianales, Fam. Loganiaceae, ed. Leeuwenberg AJM*. Berlin: Duncker & Humblot, 8–96.

Li P. 1982. A study on the genus *Buddleja* L. of China. *Acta Botanica Yunnanica* **4:** 241–247.

Li P, Leeuwenberg AJM. 1996. *Buddleja*. In: Wu CY, Raven PH, eds. *Flora of China, Vol. 15*. Beijing: Science Press, 329–337.

Liu E-D, Peng H. 2004. *Buddleja subcapitata* (Buddlejaceae), a new species from SW Sichuan, China. *Annales Botanici Fennici* **41:** 467–469.

Liu E-D, Peng H. 2006. *Buddleja microstachya* (Buddlejaceae), a new species from SW Yunnan, China. *Annales Botanici Fennici* **43:** 463–465.

Maddison W. 1997. Gene trees in species trees. *Systematic Biology* **46:** 523–536.

Maddison WP, Maddison DR. 2015. *Mesquite: a modular system for evolutionary analysis, Version 2.75*. Available at: http://mesquiteproject.org

Malia MJ Jr, Lipscomb DL, Allard MW. 2003. The misleading effects of composite taxa in supermatrices. *Molecular Phylogenetics and Evolution* **27:** 522–527.

Marquand CVB. 1930. Revision of the Old World species of *Buddleja*. *Kew Bulletin of Miscellaneous Information* **5:** 177–208.

Mayfield MH. 1999. A new species of *Emorya* (Buddlejaceae). *Sida* **18:** 693–699.

Morales JF, González J. 2007. *Buddleja filibracteolata* (Buddlejaceae), una nueva especie para Costa Rica. *Anales del Jardín Botánico de Madrid* **64:** 161–163.

Norman E. 2000. Buddlejaceae. *Flora Neotropica Monograph* **81:** 1–225.

Norman E, Moore RJ. 1968. Notes on *Emorya* (Loganiaceae). *Southwestern Naturalist* **13:** 137–142.

Olmstead RG, Reeves PA. 1995. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbcL* and *ndhF* sequences. *Annals of the Missouri Botanical Garden* **82:** 176–193.

Olmstead RG, Depamphilis CW, Wolfe AD, Young ND, Elisons WJ, Reeves PA. 2001. Disintegration of the Scrophulariaceae. *American Journal of Botany* **88:** 348–361.

Oxelman B, Backlund M, Bremer B. 1999. Relationships of the Buddlejaceae *s. l.* investigated using parsimony jackknife and branch support analysis of chloroplast *ndhF* and *rbcL* sequence data. *Systematic Botany* **24:** 164–182.

Oxelman B, Kornhall P, Norman EM. 2004. Buddlejaceae. In: Kubitzki K, ed. *The families and genera of vascular plants, VII: flowering plants, dicotyledons, Lamiales (except*

Acanthaceae including Avicenniaceae). In: Kadereit JW, ed. Berlin: Springer-Verlag, 39–44.

Oxelman B, Kornhall P, Olmstead RG, Bremer B. 2005. Further disintegration of Scrophulariaceae. *Taxon* **54:** 411–425.

Pagel M, Meade A. 2014. *BayesTraits, Version 2.0*. Available at: http://www.evolution.rdg.ac.uk/BayesTraits.html

Petersen G, Seberg O. 1997. Phylogenetic analysis of the Triticeae (Poaceae) based on *rpoA* sequence data. *Molecular Phylogenetics and Evolution* **7:** 217–230.

Phillips EP. 1946. The genera *Chilianthus* Burch. and *Buddleja* Linn. *Journal of South African Botany* **12:** 113–114.

Punt W. 1980. Pollen morphology. In: Hiepko P, Melchior H, ed. *Die natürlichen Pflanzenfamilien, Band 28bI, Angiospermae: Ordnung Gentianales, Fam. Loganiaceae, ed. Leeuwenberg AJM*. Berlin: Duncker & Humblot, 162–191.

Punt W, Leenhouts PW. 1967. Pollen morphology and taxonomy in the Loganiaceae. *Grana Palynologica* **7:** 461–516.

Rahmanzadeh R, Müller K, Fischer E, Bartels D, Borsch T. 2005. The Linderniaceae and Gratiolaceae are further lineages distinct from the Scrophulariaceae (Lamiales). *Plant Biology (Stuttgart, Germany)* **7:** 67–78.

Rambaut A, Drummond AJ. 2009. *Tracer, version 1.5*. Available at: http://tree.bio.ed.ac.uk/software/tracer

Refulio-Rodriguez NF, Olmstead RG. 2014. Phylogeny of Lamiidae. *American Journal of Botany* **101:** 287–299.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61:** 539–542.

Samarakoon T, Wang SY, Alford MH. 2013. Enhancing PCR amplification of DNA from recalcitrant plant specimens using a trehalose-based additive. *Applications in Plant Sciences* **1:** 1200236.

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

Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics (Oxford, England)* **18:** 502–504.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92:** 142–166.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* **51:** 492–508.

Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics (Oxford, England)* **17:** 1246–1247.

Small RL, Cronn RC, Wendel JF. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* **17:** 145–170.

Solereder H. 1895. Loganiaceae. In: Engler A, Prantl K, eds. *Die natürlichen Pflanzenfamilien IV, 2*. Leipzig: Engelmann, 19–50.

Sullivan J, Swofford DL, Naylor GJP. 1997. The effect of taxon sampling on estimating rate heterogeneity parameters of maximum-likelihood models. *Molecular Biology and Evolution* **16:** 1347–1356.

Takhtajan AL. 1980. Outline of the classification of flowering plants (Magnoliophyta). *Botanical Review* **46:** 225–359.

Tallent-Halsell NG, Watt MS. 2009. The invasive *Buddleja davidii* (butterfly bush). *Botanical Review* **75:** 292–325.

Tank D, Beardsley PM, Kelchner SA, Olmstead RG. 2006. Review of the systematics of Scophulariaceae *s.l.* and their current disposition. *Australian Systematic Botany* **19:** 289–307.

Thorne RF. 1983. Proposed new realignment in the angiosperms. *Nordic Journal of Botany* **3:** 85–117.

Thorne RF. 1992. Classification and geography of the flowering plants. *Botanical Review* **58:** 225–348.

Wagenitz G. 1964. Gentianales. In: Melchior H, ed. *Englers Syllabus der Pflanzenfamilien, ed. 12, vol. 2*. Berlin: Gebrüder Borntraeger.

Wilgenbusch JC, Warren DL, Swofford DL. 2004. *AWTY: a system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference*. Available at: http://ceb.csit.fsu.edu/awty

Wilhelm KA. 1910. *Die Samenpflanzen*. Vienna: Franz Deuticke.

Yuan YW, Olmstead RG. 2008. A species-level phylogenetic study of the *Verbena* complex (Verbenaceae) indicates two independent intergeneric chloroplast transfers. *Molecular Phylogenetics and Evolution* **48:** 23–33.

Yuan YW, Liu C, Marx HE, Olmstead RG. 2009. The pentatricopeptide repeat (PPR) gene family, a tremendous resource for plant phylogenetic studies. *The New Phytologist* **182:** 272–283.

Yuan YW, Liu C, Marx HE, Olmstead RG. 2010. An empirical demonstration of using pentatricopeptide repeat (PPR) genes as plant phylogenetic tools: phylogeny of Verbenaceae and the *Verbena* complex. *Molecular Phylogenetics and Evolution* **54:** 23–35.

Zhang X, Chen G, Gong W, Sun W. 2014. *Buddleja caryopteridifolia* (Scrophulariaceae), a species to be recognized based on morphology, floral scent, and AFLP data. *Phytotaxa* **161:** 181–193.

Zwickl DJ. 2006. *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. PhD dissertation, University of Texas at Austin.

APPENDIX

Table A1. Specimens included in study, with collection locality, voucher information and GenBank accession numbers for sequences

Taxon	Collection locality	Voucher	ETS	PPR24	PPR97	PPR123	*rpoA*	*trnD-trnT*	*trnS-trnfM*
Buddleja albiflora Hemsl. (A)	Cultivated, USA: Arnold Arboretum (acc. #: 13-92-A)	*J. Chau 260* (WTU, A)	KX827818	KX827926	KX827996	KX833264	–	–	–
Buddleja albiflora Hemsl. (B)*	China: Hubei	*1980 Sino-Amer. Exped. 257* (UC)	KX827819	–	–	–	KX856095	–	–
Buddleja alternifolia Maxim. (A)	Cultivated, USA: University of Washington Botanic Gardens	*R. Olmstead 2010-50* (WTU)	KX827820	KX827927	KX827997	KX833265	–	–	–
Buddleja alternifolia Maxim. (B)	China: Tibet	*G. Chen 070802* |(KUN)	KX827821	–	–	–	KX856096	KX828060	KX855287
Buddleja americana L.	Peru: La Libertad	*J. Chau 97* (HAO)	KX827822	KX827928	KX827998	KX833266	KX856097	KX828061	KX855288
Buddleja anchoensis Kuntze (A)	Bolivia: Tarija	*J. Chau 224* (WTU, LPB)	KX827823	KX827929	KX827999	KX833267	–	–	KX855289
Buddleja anchoensis Kuntze (B)*	Bolivia: Santa Cruz	*M. Nee 53158* (LPB)	KX827824	–	–	–	KX856098	KX828062	–
Buddleja araucana Phil. (A)	Argentina: Neuquén	*R. Olmstead 2007-94* (WTU)	KX827825	KX827930	–	KX833268	–	–	–
Buddleja araucana Phil. (B)*	Argentina: Rio Negro	*C. Calvino 743* (SI)	KX827826	–	–	–	KX856099	KX828063	KX855290
Buddleja aromatica J.Rémy (A)	Bolivia: La Paz	*J. Chau 206* (WTU, LPB)	KX827827	KX827931	KX828000	KX833269	–	–	KX855291
Buddleja aromatica J.Rémy (B)*	Bolivia: La Paz	*J. Solomon 13053* (CAS)	KX827828	–	–	–	KX856100	KX828064	–
Buddleja asiatica Lour. (A)	China: Yunnan	*J. Chau 157* (WTU)	KX827829	KX827932	KX828001	KX833270	–	–	–
Buddleja asiatica Lour. (B)	China: Yunnan	*G. Chen 015* (KUN)	KX827830	–	–	–	KX856101	KX828065	KX855292
Buddleja auriculata Benth.	South Africa: Eastern Cape	*J. Chau 246* (WTU, GRA)	KX827831	KX827933	KX828002	KX833271	KX856102	KX828066	KX855293
Buddleja axillaris Willd. [*Nicodemia*]*	Madagascar: Atsinanana	*B. Lewis & S. Razafim andimbison 687* (MO)	KX827832	–	–	–	–	–	–
Buddleja bhutanica Yamazaki*	Bhutan	*B. Barthol omew 3904* (CAS)	KX827833	–	–	–	KX856103	KX828067	–
Buddleja blattaria J.F.Macbr.	Peru: Cajamarca	*J. Chau 101* (HAO)	KX827834	KX827934	KX828003	KX833272	KX856104	KX828068	KX855294
Buddleja bullata Kunth	Peru: Cajamarca	*J. Chau 98* (HAO)	KX827835	KX827935	KX828004	KX833273	KX856105	KX828069	–
Buddleja candida Dunn	China: Tibet	*G. Chen 070817* (KUN)	KX827836	–	–	–	KX856106	KX828070	KX855295
Buddleja cardenasii Standl. ex E.M.Norman (A)	Bolivia: La Paz	*J. Chau 196* (WTU. LPB)	KX827837	KX827936	KX828005	KX833274	–	–	KX855296
Buddleja cardenasii Standl. ex E.M.Norman (B)*	Bolivia: Cochabamba	*S. Beck 14418* (LPB)	KX827838	–	–	–	KX856107	KX828071	–

Table A1. *Continued*

Taxon	Collection locality	Voucher	ETS	PPR24	PPR97	PPR123	*rpoA*	*trnD-trnT*	*trnS-trnfM*
Buddleja caryopteridifolia W.W.Sm. (A)	China: Yunnan	*J. Chau 171* (WTU)	KX827839	KX827941	KX828009	KX833279	–	–	KX855297
Buddleja caryopteridifolia W.W.Sm. (B)*	China: Sichuan	*D. Boufford et al. 29045* (CAS)	KX827840	–	–	–	KX856108	KX828072	–
Buddleja cestriflora Cham.	Brazil: Santa Catarina	*R. Olmstead 2010-213* (WTU, ICN)	KX827841	KX827937	–	KX833275	KX856109	KX828073	–
Buddleja colvilei Hook.f.	Cultivated, USA: San Francisco Botanical Garden (acc. #: XY-1801)	*J. Chau 42* (WTU)	KX827842	KX827938	KX828006	KX833276	KX856110	KX828074	–
Buddleja cordata Kunth	Cultivated, USA: Leu Gardens	*E. Norman s.n.* (FTU)	KX827843	–	–	–	KX856111	KX828075	KX855298
Buddleja cordobensis Griseb.*	Argentina: Córdoba	*F. Zuloaga 11302* (SI)	KX827844	–	–	–	KX856112	KX828076	KX855299
Buddleja coriacea J.Rémy (A)	Peru: Cajamarca	*J. Chau 110* (HAO)	KX827845	KX827939	KX828007	KX833277	KX856113	KX828077	–
Buddleja coriacea J.Rémy (B)*	Bolivia: La Paz	*E. Urtubey 498* (SI)	KX827846	–	–	–	–	–	KX855300
Buddleja corrugata M.E.Jones*	Mexico: Baja California Sur	*A. Carter & R. Moran 5330* (UC)	KX827847	–	–	–	KX856114	KX828078	–
Buddleja crispa Benth. (A)	China: Yunnan	*J. Chau 170* (WTU)	KX827848	KX827940	KX828008	KX833278	–	–	–
Buddleja crispa Benth. (B)	China: Yunnan	*G. Chen 070818* (KUN)	KX827849	–	–	–	KX856115	KX828079	KX855301
Buddleja crotonoides A.Gray+	Nicaragua: Madriz	*W. Stevens et al. 29357* (MO)	KX827850	KX827942	KX828010	KX833280	KX856116	KX828080	–
Buddleja curviflora Hook. & Arn.	Cultivated, USA: University of Washington Botanic Gardens (acc. #: 38-94)	*R. Olmstead 2010-49* (WTU)	KX827851	KX827943	–	KX833281	KX856117	KX828081	–
Buddleja cuspidata Baker [*Nicodemia*]*	Madagascar: Sava	*C. Rakotovao et al. 3263* (MO)	KX827852	–	–	–	–	–	–
Buddleja davidii Franch. (A)	Cultivated, China: Kunming Botanical Garden	*J. Chau 177* (WTU)	KX827853	KX827944	KX828011	–	–	–	–
Buddleja davidii Franch. (B)	China: Yunnan	*W. Sun 019* (KUN)	KX827854	–	–	–	KX856118	KX828082	KX855302
Buddleja davidii Franch. (C)	Cultivated, USA: University of Colorado	*R. Olmstead 92-192* (WTU)	KX827855	–	–	KX833282	–	–	–
Buddleja delavayi L.F.Gagnep.	China: Yunnan	*J. Chau 165* (WTU)	KX827856	KX827945	KX828012	KX833283	KX856119	KX828083	KX855303
Buddleja diffusa Ruiz & Pav.*	Bolivia: La Paz	*R. Seidel et al. 1314* (LPB)	KX827857	–	–	–	KX856120	KX828084	–
Buddleja dysophylla (Benth.) Radlk. [*Chilianthus dysophyllus* Benth.]	South Africa: Eastern Cape	*J. Chau 233* (WTU, LPB)	KX827858	KX827946	KX828013	KX833284	KX856121	KX828085	KX855304

Table A1. *Continued*

Taxon	Collection locality	Voucher	ETS	PPR24	PPR97	PPR123	*rpoA*	*trnD-trnT*	*trnS-trnfM*
Buddleja elegans Cham. & Schltdl. subsp. *elegans* (A)	Brazil: Rio Grande do Sul	*R. Olmstead 2010-214* (ICN)	KX827860	KX827947	–	KX833285	–	–	–
Buddleja elegans Cham. & Schltdl. subsp. *elegans* (B)	Brazil: Rio Grande do Sul	*R. Olmstead 2010-210* (WTU, ICN)	KX827861	–	–	–	KX856122	KX828086	–
Buddleja elegans Cham. & Schltdl. subsp. *angustata* (Benth.) E.M.Norman	Brazil: Rio Grande do Sul	*V. Thode et al. 399* (ICN)	KX827859	KX827982	KX828047	KX833320	KX856159	KX828122	KX855323
Buddleja fallowiana Balf.f. & W.W.Smith (A)	China: Yunnan	*J. Chau 166* (WTU)	KX827862	KX827948	KX828014	KX833286	–	–	–
Buddleja fallowiana Balf.f. & W.W.Smith (B)	China: Yunnan	*G. Chen 059* (KUN)	KX827863	–	–	–	KX856123	KX828087	KX855305
Buddleja forrestii Diels (A)	China: Yunnan	*J. Chau 161* (WTU)	KX827864	KX827949	–	–	KX856124	KX828088	KX855306
Buddleja forrestii Diels (B)	Cultivated, USA: University of California Botanical Garden (acc. #: 91.0429)	*R. Welch s.n.* (UC)	KX827865	–	KX828015	KX833287	–	–	–
Buddleja fragifera Leeuwenb. [*Nicodemia*]*	Madagascar: Atsimo-Andrefana	*P. Phillipson 3007* (MO)	KX827866	–	–	KX827817	KX856125	KX828089	–
Buddleja fusca Baker [*Nicodemia*]*	Madagascar: Vakinan karatra	*P. Phillipson et al. 5634* (MO)	KX827867	KX827950	–	KX833288	KX856126	KX828090	–
Buddleja globosa Hope	Cultivated, USA: University of Washington Botanic Gardens	*R. Olmstead 2010-46* (WTU)	KX827868	KX827951	KX828016	KX833289	KX856127	KX828091	KX855307
Buddleja glomerata H.Wendl. [*Chilianthus lobulatus* Benth.]	South Africa: Eastern Cape	*J. Chau 254* (WTU, GRA)	KX827869	KX827952	KX828017	KX833290	KX856128	KX828092	KX855308
Buddleja grandiflora Cham. & Schltdl.	Brazil: Rio Grande do Sul	*R. Olmstead 2010-207* (WTU, ICN)	KX827870	KX827953	–	KX833291	KX856129	KX828093	–
Buddleja hieronymi R.E.Fr. (A)	Bolivia: Tarija	*J. Chau 225* (WTU, LPB)	KX827871	KX827954	KX828018	–	KX856130	KX828094	KX855309
Buddleja hieronymi R.E.Fr. (B)	Argentina: Jujuy	*R. Olmstead 2007-59* (WTU)	KX827872	–	–	KX833292	–	–	–
Buddleja incana Ruiz & Pav.	Peru: Cajamarca	*J. Chau 111* (HAO)	KX827873	KX827955	KX828019	KX833293	KX856131	KX828095	KX855310
Buddleja indica Lam. [*Nicodemia diversifolia* (Vahl) Ten.]+	Madagascar: Atsinanana	*J. Rabenan toandro 1234* (MO)	KX827874	KX827956	KX828020	KX833294	KX856132	KX828096	–

Table A1. *Continued*

Taxon	Collection locality	Voucher	ETS	PPR24	PPR97	PPR123	*rpoA*	*trnD-trnT*	*trnS-trnfM*
Buddleja interrupta Kunth	Peru: Cajamarca	*J. Chau 117* (HAO)	KX827875	KX827957	KX828021	KX833295	KX856133	KX828097	KX855311
Buddleja jamesonii Benth.*	Ecuador: Azuay	*P. Jorgensen 92920* (MO)	KX827876	–	–	–	KX856134	KX828098	–
Buddleja japonica Hemsl.	Cultivated, USA: Arnold Arboretum (acc. #: 7-92-B)	*J. Wood 124-2014* (A)	KX827877	KX827958	KX828022	KX833296	–	–	–
Buddleja kleinii E.M.Norman & L.B.Sm.	Brazil: Santa Catarina	*R. Olmstead 2010-220* (WTU, ICN)	KX827878	KX827959	KX828023	KX833297	KX856135	KX828099	–
Buddleja lindleyana Fortune (A)	China: Hubei	*G. Chen 053* (KUN)	KX827879	–	KX828024	KX833298	KX856136	KX828100	KX855312
Buddleja lindleyana Fortune (B)	Cultivated, USA: R. Olmstead garden	*R. Olmstead 2009-51* (WTU)	KX827880	KX827960	–	–	–	–	–
Buddleja longiflora Brade	Cultivated, USA: University of Washington Biology greenhouse	*J. Chau 308* (WTU)	KX827881	KX827961	KX828025	KX833299	–	–	–
Buddleja loricata Leeuwenberg [*Chilianthus corrugatus* Benth.]	Cultivated, USA: University of California Botanical Garden (acc. #: 2006.0671)	*R. Welch s.n.* (UC)	KX827882	KX827962	KX828026	KX833300	KX856137	KX828101	–
Buddleja macrostachya Benth. (A)	China: Yunnan	*J. Chau 159* (WTU)	KX827883	KX827963	KX828027	–	–	–	–
Buddleja macrostachya Benth. (B)	China: Yunnan	*G. Chen 045* (KUN)	KX827884	–	–	KX833333	KX856138	KX828102	KX855313
Buddleja madagascariensis Lam. [*Nicodemia madagascariensis* (Lam.) R.Parker]	Cultivated, USA: Los Angeles County Arboretum (acc. #: 20050221)	*J. Chau 256* (WTU)	KX827885	KX827964	KX828028	KX833301	KX856139	KX828103	KX855314
Buddleja marrubiifolia Benth.	Cultivated, USA: University of California-Davis Arboretum (acc. #: A85.0360)	*J. Chau 40* (WTU)	KX827886	KX827965	KX828029	KX833302	KX856140	KX828104	–
Buddleja megalocephala Donn. Sm.+	Guatemala: Huehue tenango	*M. Christe nhusz et al. 5266* (MO)	KX827887	–	KX828030	KX833303	KX856141	KX828105	–
Buddleja mendozensis Gillies ex Benth.*	Argentina: Catamarca	*F. Zuloaga 12016* (SI)	KX827888	–	–	–	KX856142	KX828106	KX855315
Buddleja montana Britton	Bolivia: La Paz	*J. Chau 186* (WTU, LPB)	KX827889	KX827966	KX828031	KX833304	KX856143	KX828107	KX855316
Buddleja myriantha Diels (A)	China: Yunnan	*J. Chau 158* (WTU)	KX827890	KX827967	KX828032	KX833305	–	–	–
Buddleja myriantha Diels (B)	China: Yunnan	*W. Sun 033* (KUN)	KX827891	–	–	–	KX856144	KX828108	KX855317
Buddleja nitida Benth. (A)	Costa Rica: Cartago	*J. Chau 150* (WTU)	KX827892	KX827968	KX828033	–	–	–	–

Table A1. *Continued*

Taxon	Collection locality	Voucher	ETS	PPR24	PPR97	PPR123	*rpoA*	*trnD-trnT*	*trnS-trnfM*
Buddleja nitida Benth. (B)	Cultivated, USA: University of California Botanical Garden (acc. #: 87.0253)	*M. Grayum 8188* (CR)	KX827893	–	–	KX833306	KX856145	KX828109	–
Buddleja nivea Duthie	Cultivated, USA: University of Washington Botanic Gardens (acc. #: 396-61*A)	*R. Olmstead 2010-47* (WTU)	KX827894	KX827969	KX828034	KX833307	KX856146	KX828110	–
Buddleja officinalis Maxim. (A)	China: Yunnan	*J. Chau 179* (WTU)	KX827895	KX827970	KX828035	KX833308	–	–	–
Buddleja officinalis Maxim. (B)	China: Yunnan	*G. Chen 012* (KUN)	KX827896	–	–	–	KX856147	KX828111	KX855318
Buddleja polystachya Fresen.*	Tanzania: Arusha	*G. Simon 308* (MO)	KX827897	KX827971	KX828036	KX833309	KX856148	KX828112	–
Buddleja pulchella N.E.Br.*	South Africa: KwaZulu-Natal	*I. Nanni 319* (NBG)	KX827898	KX827972	KX828037	KX833310	KX856149	–	–
Buddleja racemosa Torr.*	USA: Texas	*G. Webster & B. Westlund 32714* (DAV)	KX827899	KX827973	KX828038	KX833311	KX856150	KX828113	KX855319
Buddleja rufescens Willd. ex Schultes & Schultes	Peru: Cajamarca	*J. Chau 99* (HAO)	KX827900	KX827974	KX828039	KX833312	KX856151	KX828114	KX855320
Buddleja saligna Willd. [*Chilianthus oleaceus* Burch.]	South Africa: Western Cape	*R. Olmstead 99-20*	KX827901	KX827975	KX828040	KX833313	KX856152	KX828115	–
Buddleja salviifolia (L.) Lam.	Cultivated, USA: San Francisco Botanical Garden (acc. #: XY-1999)	*J. Chau 43* (WTU)	KX827902	KX827976	KX828041	KX833314	KX856153	KX828116	–
Buddleja scordioides Kunth (A)*	Mexico: Sonora	*T. Van Devender 2007-744* (CAS)	KX827903	–	–	–	KX856154	KX828117	–
Buddleja scordioides Kunth (B)	Mexico: Coahuila	*M. Moore 2560* (WTU)	KX827904	KX827977	KX828042	KX833315	–	–	–
Buddleja sessiliflora Kunth*	USA: Texas	*G. Webster 31455* (DAV)	KX827905	KX827978	KX828043	KX833316	KX856155	KX828118	–
Buddleja skutchii C.V.Morton	Costa Rica: San José	*J. Chau 152* (WTU)	KX827906	KX827979	KX828044	KX833317	KX856156	KX828119	KX855321
Buddleja speciosissima Taub. (A)	Brazil: Rio de Janeiro	*F. Salimena 2980* (CESJ)	KX827907	–	–	KX833318	KX856157	KX828120	–
Buddleja speciosissima Taub. (B)	Cultivated, USA: University of Washington Biology greenhouse	*J. Chau 259* (WTU)	KX827908	KX827980	KX828045	–	–	–	–
Buddleja stachyoides Cham. & Schltdl. (A)	Brazil: Minas Gerais	*F. Salimena 2947* (CESJ)	KX827909	KX827981	KX828046	KX833319	–	–	–

Table A1. *Continued*

Taxon	Collection locality	Voucher	ETS	PPR24	PPR97	PPR123	*rpoA*	*trnD-trnT*	*trnS-trnfM*
Buddleja stachyoides Cham. & Schltdl. (B)*	Argentina: Jujuy	*F. Zuloaga 11630* (SI)	KX827910	–	–	–	KX856158	KX828121	KX855322
Buddleja tubiflora Benth.	Cultivated, USA: E. Norman garden	*Norman s.n.* (WTU)	KX827911	KX827983	KX828048	KX833321	KX856160	KX828123	KX855324
Buddleja tucumanensis Griseb.	Bolivia: Chuquisaca	*J. Chau 212* (WTU, LPB)	KX827912	KX827984	KX828049	KX833322	KX856161	KX828124	KX855325
Buddleja utahensis Coville	Cultivated, USA: Rancho Santa Ana Botanic Garden (acc. #: 17353)	*J. Chau 39* (WTU)	KX827913	KX827985	KX828050	KX833323	KX856162	KX828125	–
Buddleja vexans Kraenzl. & Loes. ex E.M.Norman	Peru: Huancavelica	*J. Chau 136* (HAO)	KX827914	–	–	–	KX856163	KX828126	KX855326
Buddleja yunnanensis L.F.Gagnep. (A)	Cultivated, China: Kunming Botanical Garden	*J. Chau 178* (WTU)	KX827915	KX827986	KX828051	–	–	–	–
Buddleja yunnanensis L.F.Gagnep. (B)	China: Yunnan	*W. Sun 028* (KUN)	KX827916	–	–	KX833324	KX856164	KX828127	KX855327
Emorya suaveolens Torr.*	Mexico: Coahuila	*D. Riskind 23860* (TEX)	KX827917	KX827987	KX828052	KX833325	KX856165	KX828128	–
Gomphostigma incomptum (L.f.) N.E.Br.+	South Africa: Northern Cape	*P. Goldblatt & L. Porter 12664* (NBG)	KX827918	KX827988	KX828053	KX833326	KX856166	KX828129	KX855328
Gomphostigma virgatum (L.f.) Baill.	Cultivated, USA: University of California-Davis Arboretum (acc. #: M06.9257)	*J. Chau 180* (WTU)	KX827919	KX827989	KX828054	KX833327	KX856167	KX828130	KX855329
Oftia africana (L.) Bocq.	South Africa: Western Cape	–	KX827920	KX827990	KX828055	KX833328	KX856168	KX828131	KX855330
Phygelius capensis E.Mey. ex Benth.	Cultivated, USA: R. Olmstead garden	*R. Olmstead 07-153* (WTU)	KX827921	KX827991	KX828056	KX833329	KX856169	KX828132	KX855331
Scrophularia nodosa L.	Cultivated, USA: University of Washington Medicinal Herb Garden	*J. Chau 228* (WTU)	KX827922	KX827992	–	KX827816	KX856170	KX828133	KX855332
Nemesia fruticans Benth.	Cultivated, USA: R. Olmstead garden	*R. Olmstead 07-107* (WTU)	KX827923	KX827993	KX828057	KX833330	–	–	–
Nuxia floribunda Benth.	Cultivated, USA: Los Angeles County Arboretum	*J. Chau 258* (WTU)	KX827924	KX827994	KX828058	KX833331	KX856171	KX828134	KX855333
Lantana depressa Small	Cultivated, USA: Fairchild Tropical Botanic Garden	*P. Lu-Irving 12-1* (WTU)	KX827925	KX827995	KX828059	KX833332	KX856172	KX828135	KX855334

Specimens with DNA extracted from herbarium specimen tissue indicated with asterisk (*). Specimens with DNA from DNA banks indicated with a cross (+). All other specimens have DNA extracted from silica-preserved leaf tissue. For species with multiple specimens, concatenated dataset always included ETS sequence from specimen (A). For species that have been included in *Buddleja* and *Chilianthus* or *Nicodemia*, accepted names in both genera are listed where available.

Table A2. Sequences of new primers used in this study

Name	Sequence (5′→3′)
PPR24-140F	CACGTACCCKTTTGTKTTTAAGGC
PPR24-1354R	ACTMAGCAAAGCACCRTAAAGTGG
PPR24-310F-Bud	GATGAGGCTACRGTTGTTAGTAC
PPR24-600R-Bud	GATACCATAMTTGTCCAACAAATAACATTCTT
PPR24-950F-Bud	CTTACAGGRTGTGCYCAATTAGG
PPR24-970R	TCTAAGMAACCACATTTTGCRTACAT
PPR97-781F	CTTGTRGATTTGGGTGCWARGTGGTT
PPR97-1585R	TTTTTCACATAAGCWGTYACAAGAAT
PPR97-1165F	AACACAATGATCACTGGAYATGGGA
PPR97-1351R	AAGTTTGAYGAATTRGGCTTAAA
PPR123-820F	ATGATTAAYGTGTTTGGAAAGGC
PPR123-1370F-Bud	GGAAAGTTAGATCGTGCAGC
PPR123-1500R-Bud	GAGCAACCAAACCAGCCCTCTC

Table A3. Trait states for taxa included in study as determined from published species descriptions in monographs and floras

Taxon	Fruit type	Stamen	Corolla shape	Corolla colour	Inflorescence type
Buddleja albiflora	Capsule	Included	Tube-shaped	Purple	Thyrsoid
Buddleja alternifolia	Capsule	Included	Tube-shaped	Purple	Paniculate
Buddleja americana	Capsule	Included	Cup-shaped	Yellow	Paniculate
Buddleja anchoensis	Capsule	Included	Tube-shaped	Orange	Capitate
Buddleja araucana	Capsule	Included	Tube-shaped	Orange	Capitate
Buddleja aromatica	Capsule	Included	Tube-shaped	White	Capitate
Buddleja asiatica	Capsule	Included	Tube-shaped	White	Spiciform
Buddleja auriculata	Capsule	Included	Tube-shaped	White	Paniculate
Buddleja axillaris	Berry	Included	Tube-shaped	White	Thyrsoid
Buddleja bhutanica	Capsule	Included	Tube-shaped	White	Spiciform
Buddleja blattaria	Capsule	Included	Tube-shaped	White	Capitate
Buddleja bullata	Capsule	Included	Tube-shaped	Yellow	Paniculate
Buddleja candida	Capsule	Included	Tube-shaped	Purple	Spiciform
Buddleja cardenasii	Capsule	Included	Tube-shaped	Orange	Paniculate
Buddleja caryopteridifolia	Capsule	Included	Tube-shaped	Purple	Spiciform
Buddleja cestriflora	Capsule	Included	Tube-shaped	Orange	Thyrsoid
Buddleja colvilei	Capsule	Included	Tube-shaped	Purple	Thyrsoid
Buddleja cordata	Capsule	Included	Cup-shaped	Yellow	Paniculate
Buddleja cordobensis	Capsule	Included	Tube-shaped	Yellow	Capitate
Buddleja coriacea	Capsule	Included	Tube-shaped	Orange	Paniculate
Buddleja corrugata	Capsule	Included	Tube-shaped	Orange	Spiciform
Buddleja crispa	Capsule	Included	Tube-shaped	Purple	Paniculate
Buddleja crotonoides	Capsule	Included	Cup-shaped	Yellow	Paniculate
Buddleja curviflora	Capsule	Included	Tube-shaped	Purple	Thyrsoid
Buddleja cuspidata	Berry	Included	Tube-shaped	Yellow	Spiciform
Buddleja davidii	Capsule	Included	Tube-shaped	Purple	Thyrsoid
Buddleja delavayi	Capsule	Included	Tube-shaped	Purple	Paniculate
Buddleja diffusa	Capsule	Included	Tube-shaped	Yellow	Paniculate
Buddleja dysophylla	Capsule	Exserted	Cup-shaped	White	Paniculate
Buddleja elegans subsp. *angustata*	Capsule	Included	Tube-shaped	White	Thyrsoid
Buddleja elegans subsp. *elegans*	Capsule	Included	Tube-shaped	White	Thyrsoid
Buddleja fallowiana	Capsule	Included	Tube-shaped	Purple	Thyrsoid

Table A3. *Continued*

Taxon	Fruit type	Stamen	Corolla shape	Corolla colour	Inflorescence type
Buddleja forrestii	Capsule	Included	Tube-shaped	Purple	Thyrsoid
Buddleja fragifera	Berry	Included	Tube-shaped	Orange	Capitate
Buddleja fusca	Berry	Included	Tube-shaped	Orange	Thyrsoid
Buddleja globosa	Capsule	Included	Tube-shaped	Orange	Capitate
Buddleja glomerata	Capsule	Exserted	Cup-shaped	Yellow	Paniculate
Buddleja grandiflora	Capsule	Included	Tube-shaped	Yellow	Capitate
Buddleja hieronymi	Capsule	Included	Tube-shaped	Yellow	Capitate
Buddleja incana	Capsule	Included	Tube-shaped	Orange	Paniculate
Buddleja indica	Berry	Included	Tube-shaped	Yellow	Thyrsoid
Buddleja interrupta	Capsule	Included	Cup-shaped	Yellow	Paniculate
Buddleja jamesonii	Capsule	Included	Tube-shaped	Yellow	Spiciform
Buddleja japonica	Capsule	Included	Tube-shaped	Purple	Thyrsoid
Buddleja kleinii	Capsule	Included	Tube-shaped	White	Capitate
Buddleja lindleyana	Capsule	Included	Tube-shaped	Purple	Thyrsoid
Buddleja longiflora	Capsule	Included	Tube-shaped	Orange	Thyrsoid
Buddleja loricata	Capsule	Exserted	Cup-shaped	White	Paniculate
Buddleja macrostachya	Capsule	Included	Tube-shaped	Purple	Thyrsoid
Buddleja madagascariensis	Berry	Included	Tube-shaped	Orange	Thyrsoid
Buddleja marrubiifolia	Capsule	Included	Tube-shaped	Orange	Capitate
Buddleja megalocephala	Capsule	Included	Tube-shaped	Orange	Capitate
Buddleja mendozensis	Capsule	Included	Tube-shaped	Orange	Spiciform
Buddleja montana	Capsule	Included	Cup-shaped	Orange	Paniculate
Buddleja myriantha	Capsule	Included	Tube-shaped	Purple	Thyrsoid
Buddleja nitida	Capsule	Included	Cup-shaped	Yellow	Paniculate
Buddleja nivea	Capsule	Included	Tube-shaped	Purple	Thyrsoid
Buddleja officinalis	Capsule	Included	Tube-shaped	Purple	Paniculate
Buddleja polystachya	Capsule	Included	Tube-shaped	Orange	Spiciform
Buddleja pulchella	Capsule	Included	Tube-shaped	White	Paniculate
Buddleja racemosa	Capsule	Included	Tube-shaped	Yellow	Capitate
Buddleja rufescens	Capsule	Included	Cup-shaped	Yellow	Paniculate
Buddleja saligna	Capsule	Exserted	Cup-shaped	White	Paniculate
Buddleja salviifolia	Capsule	Included	Tube-shaped	White	Paniculate
Buddleja scordioides	Capsule	Included	Cup-shaped	Yellow	Spiciform
Buddleja sessiliflora	Capsule	Included	Tube-shaped	Yellow	Spiciform
Buddleja skutchii	Capsule	Included	Cup-shaped	Orange	Paniculate
Buddleja speciosissima	Capsule	Included	Tube-shaped	Orange	Thyrsoid
Buddleja stachyoides	Capsule	Included	Tube-shaped	Yellow	Capitate
Buddleja tubiflora	Capsule	Included	Tube-shaped	Orange	Capitate
Buddleja tucumanensis	Capsule	Included	Tube-shaped	Orange	Capitate
Buddleja utahensis	Capsule	Included	Tube-shaped	Yellow	Spiciform
Buddleja vexans	Capsule	Included	Tube-shaped	Orange	Paniculate
Buddleja yunnanensis	Capsule	Included	Tube-shaped	Purple	Spiciform
Emorya suaveolens	Capsule	Exserted	Tube-shaped	Yellow	Thyrsoid
Gomphostigma incomptum	Capsule	Exserted	Cup-shaped	White	Racemose
Gomphostigma virgatum	Capsule	Exserted	Cup-shaped	White	Racemose

For fruit type, fleshy fruits are coded as 'berry' and dry fruits are coded as 'capsule'. For stamens, those that extended outside the corolla tube are coded as 'exserted' and those that are hidden inside the corolla tube are coded as 'included'. For corolla shape, those with a corolla tube length to lobe length ratio < 1.8 are coded as 'cup-shaped' and those with a ratio > 1.8 are coded as 'tube-shaped'. For corolla colour, the colour of the majority of the corolla is considered. In many species, the throat has a different colour which is not considered. For inflorescence type, those with peduncled cymes and sessile flowers are considered 'capitate', those with sessile cymes and sessile flowers are considered 'spiciform', those with pedunculate cymes and pedicellate flowers are considered 'thyrsoid', those with more than one order of branching are considered 'paniculate' and those with single-flowered cymes in a raceme are considered 'racemose'. All taxa were coded as having a single state for each trait, although in some cases polymorphism exists.

Table A4. Revised classification and list of species in Buddlejeae

Genus *Buddleja* L. [108]
- Section *Salviifoliae* J.H.Chau [1]
  - *B. salviifolia* (L.) Lam.
- Section *Gomphostigma* (Turcz.) J.H.Chau [2]
  - *B. incompta* L.f.
  - *B. virgata* L.f.
- Section *Chilianthus* (Burch.) Leeuwenberg [4]
  - *B. auriculata* Benth.
  - *B. dysophylla* (Benth.) Radlk.
  - *B. loricata* Leeuwenberg
  - *B. saligna* Willd.
- Section *Pulchellae* J.H.Chau [1]
  - *B. pulchella* N.E.Br.
- Section *Nicodemia* (Ten.) Leeuwenberg [9]
  - *B. acuminata* Poir.
  - *B. axillaris* Willd.
  - *B. cuspidata* Baker
  - *B. indica* Lam.
  - *B. fragifera* Leeuwenberg
  - *B. fusca* Baker
  - *B. madagascariensis* Lam.
  - *B. polystachya* Fresen.
  - *B. sphaerocalyx* Baker
- Section *Alternifoliae* Kränzl. [24]
  - *B. albiflora* Hemsl.
  - *B. alternifolia* Maxim.
  - *B. asiatica* Lour.
  - *B. bhutanica* Yamazaki
  - *B. brachystachya* Diels
  - *B. candida* Dunn
  - *B. caryopteridifolia* W.W.Sm.
  - *B. colvilei* Hook.f.
  - *B. crispa* Benth.
  - *B. davidii* Franch.
  - *B. delavayi* L.F.Gagnep.
  - *B. fallowiana* Balf.f. & W.W.Smith
  - *B. forrestii* Diels
  - *B. japonica* Hemsl.
  - *B. jinsixiaensis* R.B.Zhu
  - *B. lindleyana* Fortune
  - *B. macrostachya* Benth.
  - *B. microstachya* E.D.Liu & H.Peng
  - *B. myriantha* Diels
  - *B. nivea* Duthie
  - *B. officinalis* Maxim.
  - *B. paniculata* Wall.
  - *B. subcapitata* E.D.Liu & H.Peng
  - *B. yunnanensis* L.F.Gagnep.
- Section *Buddleja* [66]
  - *B. americana* L.
  - *B. anchoensis* Kuntze
  - *B. araucana* Phil.
  - *B. aromatica* J.Rémy
  - *B. blattaria* J.F.Macbr.
  - *B. brachiata* Cham. & Schltdl.
  - *B. bullata* Kunth
  - *B. cardenasii* Standl. ex E.M.Norman
  - *B. cestriflora* Cham.
  - *B. chapalana* B.L.Rob.
  - *B. chenopodiifolia* Kraenzl.
  - *B. cordobensis* Griseb.
  - *B. cordata* Kunth
  - *B. coriacea* J.Rémy
  - *B. corrugata* M.E.Jones
  - *B. crotonoides* A.Gray
  - *B. cuneata* Cham.
  - *B. diffusa* Ruíz & Pav.
  - *B. domingensis* Urb.
  - *B. elegans* Cham. & Schltdl.
  - *B. euryphylla* Standl. & Steyerm.
  - *B. filibracteolata* J.A.González & J.F.Morales
  - *B. globosa* Hope
  - *B. grandiflora* Cham. & Schltdl.
  - *B. hatschbachii* E.M.Norman & L.B.Sm.
  - *B. hieronymi* R.E.Fr.
  - *B. ibarrensis* E.M.Norman
  - *B. incana* Ruiz & Pav.
  - *B. interrupta* Kunth
  - *B. iresinoides* (Griseb.) Hosseus
  - *B. jamesonii* Benth.
  - *B. kleinii* E.M.Norman & L.B.Sm.
  - *B. lanata* Benth.
  - *B. lojensis* E.M.Norman
  - *B. longiflora* Brade
  - *B. longifolia* Kunth
  - *B. marrubiifolia* Benth.
  - *B. megalocephala* Donn.Sm.
  - *B. mendozensis* Gillies ex Benth.
  - *B. misionum* Kraenzl.
  - *B. montana* Britton
  - *B. multiceps* Kraenzl.
  - *B. nitida* Benth.
  - *B. normaniae* J.H.Chau
  - *B. oblonga* Benth.
  - *B. parviflora* Kunth
  - *B. perfoliata* Kunth
  - *B. pichinchensis* Kunth
  - *B. polycephala* Kunth
  - *B. racemosa* Torr.
  - *B. ramboi* L.B.Sm.
  - *B. rinconensis* (Mayfield) J.H.Chau
  - *B. rufescens* Willd. ex Schultes & Schultes
  - *B. scordioides* Kunth
  - *B. sessiliflora* Kunth
  - *B. simplex* Kraenzl.
  - *B. skutchii* C.V.Morton
  - *B. soratae* Kraenzl.

Table A4. *Continued*

B. speciosissima Taub.
B. stachyoides Cham. & Schltdl.
B. suaveolens Kunth & Bouché
B. thyrsoides Lam.
B. tubiflora Benth.
B. tucumanensis Griseb.
B. utahensis Coville
B. vexans Kraenzl. & Loes. ex E.M.Norman
Incertae sedis
B. glomerata H.Wendl.

Number of species in each taxon indicated in brackets.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Majority-rule consensus phylograms from Bayesian analyses of individual locus datasets, excluding 25% burn-in. Values at nodes indicate support: maximum likelihood bootstrap percentage (BP)/Bayesian posterior probability (PP), if > 50% BP or 0.5 PP. Nodes with > 70% BP and 0.9 PP support are highlighted with thicker branches. Letter after species name indicates species that has also been considered a member of *Chilianthus* (C) or *Nicodemia* (N). (A) ETS, (B) PPR24, (C) PPR97, (D) PPR123, (E) plastid, consisting of partitioned concatenated dataset with *trnD-trnT*, *trnS-trnfM* and *rpoA*.

Figure S2. Majority-rule consensus tree from Bayesian analysis of ETS dataset with expanded specimen sampling, excluding 25% burn-in. Values at nodes indicate support: maximum likelihood bootstrap percentage (BP)/Bayesian posterior probability (PP), if > 70% BP or 0.9 PP. Nodes with > 70% BP and 0.9 PP support are highlighted with thicker branches.

CHAPTER 2:

Origins and timing of intercontinental disjunctions in the widely-distributed plant genus *Buddleja* (Scrophulariaceae)

Abstract

Aim: Intercontinental disjunctions are a commonly observed pattern in the distributions of many taxa. Various mechanisms can result in the pattern, including vicariance by continental drift, long-distance dispersal, and the formation of temporary migration corridors. We reconstructed a time-calibrated phylogeny for the widely distributed plant genus *Buddleja*, which is found on five major land masses, and inferred ancestral ranges to evaluate possible mechanisms for the intercontinental disjunctions in this group. We also estimated diversification rates and assessed whether changes in distribution were associated with diversification rate shifts.

Location: Tropical montane and subtropical Africa, Madagascar, Asia, North America, and South America

Methods: Bayesian relaxed clock analyses were used to infer phylogenetic relationships and divergence dates from four nuclear and three plastid loci with samples representing the global distribution of *Buddleja*. Secondary calibrations from two broad-scale angiosperm dating studies were used. Ancestral ranges were estimated under several biogeographical models in BioGeoBEARS and with the statistical dispersal-extinction-cladogenesis and Bayesian binary methods in RASP. Diversification rate shifts were estimated with Bayesian approaches in BAMM, BayesRate, and RevBayes.

Results: *Buddleja* most likely originated in the early Miocene in southern Africa. Single transitions from Africa to the New World, Asia, and Madagascar were inferred to have occurred in the mid to late Miocene. In the New World, *Buddleja* most likely arrived first in North America and then expanded to South America several times. A speciation rate increase was inferred early in the diversification of the New World clade.

Main conclusions: We propose long-distance dispersal and past migration corridors through northern high-latitude pathways as possible mechanisms for the expansion of *Buddleja* from

Africa to the New World, Asia, and Madagascar. A possible diversification rate increase near the origin of the New World clade suggests the extraordinary potential for speciation found in the American cordilleras.

INTRODUCTION

Intercontinental disjunctions in the distribution of closely related species have been a long-documented and commonly observed pattern in plants (Willdenow, 1792; Humboldt & Bonpland, 1805; Gray, 1846; Hooker, 1853; Thorne, 1972). The processes leading to these distributions are an important topic of biogeography (Raven & Axelrod, 1974; Ronquist & Sanmartín, 2011) and can be broadly categorized into mechanisms involving vicariance, the fragmentation of a once widespread ancestral population, or the dispersal and establishment of populations from one area to another. Each mechanism offers predictions about divergence ages and pattern of relationships with respect to intervening barriers and species distributions (Morrone & Crisci, 1995).

Vicariance hypotheses posit that a population can be divided by the formation of a barrier that prevents interbreeding between populations on either side of the barrier, leading to the formation of separate species. With vicariance, speciation events are expected to occur concurrently with or prior to the formation of the barrier, but cannot be any younger. In addition, the order of branching events in species relationships should follow the order of barrier formation (Morrone & Crisci, 1995). One possible mechanism for the formation of barriers is continental drift. Plate tectonic theory explains how landmasses now widely separated by oceans and distance were contiguous in the past (Dietz & Holden, 1970). A species that originated during the existence of a larger landmass and achieved a wide range could be separated by vicariance when the landmass breaks apart. This mechanism is most often invoked when explaining the distribution of groups encompassing several continents of the Southern Hemisphere (e.g., ratite birds, Nothofagaceae), which were previously connected in the supercontinent Gondwana (Haddrath & Baker, 2001; Swenson et al., 2001). The modern position of continents had been established by the middle Tertiary (ca. 25 Ma),

though the separation of some landmasses began much earlier (e.g., Africa and South America separated ca. 96 Ma; Pitman et al., 1993).

Dispersal allows individuals of a population to establish new populations in remote areas, which may become separate species if gene flow is low. Long-distance dispersal of plant propagules is facilitated by various mechanisms, including wind, ocean currents, and transport on and in animals (Carlquist, 1967; Renner, 2004; Nathan et al., 2008). Dispersal is generally considered to be a stochastic process and may have large variance in frequency and distance (Nathan & Muller-Landau, 2000). Dispersal has been invoked to explain the colonization of oceanic islands (Cowie & Holland, 2006) and the distribution of taxa on separate continents (e.g. Särkinen et al., 2007; Olmstead, 2013; Dupin et al., 2016). If a barrier exists between species, colonization and speciation are inferred to have occurred more recently than barrier formation in this scenario. Time-calibrated phylogenies have helped support dispersal as the most likely explanation for the disjunct distributions of many groups by showing that speciation events postdate the formation of geographical barriers (de Queiroz, 2005; Yoder & Nowak, 2006; Christenhusz & Chase, 2013).

A combination of dispersal and vicariance can result in disjunct patterns of distribution when there is episodic expansion and contraction of biomes due to climate and other large-scale changes. With changes in the global climate regime, biomes restricted to discrete areas may increase their extent to connect with one another, allowing species to disperse among areas. Subsequent changes may cause biomes that stretched across large regions to fragment when intervening areas are converted to other biome types. For example, according to the boreotropical hypothesis, tropical biomes extended across high northern latitudes during periods of warmer global temperatures in the Paleogene (e.g., early Eocene climatic optimum, ca. 49 Ma). Subsequent cooling of global temperatures caused tropical lineages to be restricted to equatorial latitudes that were widely separated from each other on

different continents (Wolfe, 1975; Tiffney, 1985). This explanation has been proposed for the distribution of some plant taxa (e.g., *Magnolia*, Malpighiaceae) found widely across the tropical regions of Africa, Asia, and the Americas (Azuma et al., 2001; Davis et al., 2002).

Changes in distribution, whether by splitting through vicariance or expansion through dispersal, can result in changes in diversification dynamics (Moore & Donoghue, 2007; Uribe-Converse & Tank, 2015). Variation across clades in species richness and diversification rate is widespread across the tree of life, and various mechanisms have been proposed as explanations, including age, morphological and physiological innovations, and habitat characteristics (Linder, 2008; Heard & Hauser; 1995, Hughes & Atchison, 2015). Shifts in diversification rate have also been associated with movements into new geographic areas, which may present new ecological opportunities, allow for ecological release from competitive or inhibitory elements, or expand the group's distribution to include habitats that promote speciation (von Hagen & Kadereit, 2003; Moore & Donoghue, 2007; Uribe-Converse & Tank, 2015; Ogutcen et al., 2017).

The genus *Buddleja* comprises 108 species of shrubs and trees with a broad distribution extending across montane tropical, subtropical, and warm-temperate areas of continental Africa, Madagascar, Asia, and North and South America (Leeuwenberg, 1979; Norman, 2000; Chau et al., 2017; Fig. 1). On Africa, ten species in four sections have a combined range extending from southern Africa to montane regions of eastern Africa and the Arabian peninsula. Eight species of *Buddleja* in section *Nicodemia* are found on the island of Madagascar. *Buddleja* has its second highest concentration of taxonomic diversity in Asia, where 24 species are found. The eastern Himalaya-Hengduan Mountains region is especially rich in diversity, but the distribution extends from the western Himalayas to Japan and montane tropical areas of southeast Asia (Leeuwenberg, 1979). *Buddleja* is most diverse in the Americas, where about two-thirds of all species occur. The distribution extends from the

deserts of the southwestern United States to the Andean foothills of central Chile and Argentina. In the New World, about two-thirds of species occur in South America and the rest in North and Central America. Centers of diversity and endemism are found in the central Andes, southeastern Brazil, and the central Mexican highlands (Norman, 2000). The distribution of *Buddleja* includes five common patterns of intercontinental disjunctions (Thorne, 1972; Yoder & Nowak, 2006): African-Madagascan, African-Eurasian, North Temperate (Asia-North America), American-African, and North American-South American.

Determining the mechanisms behind the distribution of a group can have important implications for understanding its evolutionary history. The genus *Buddleja*, with a distribution and centers of diversity spread across five major landmasses, is a prime candidate for the study of broad-scale biogeographic processes and the origin of intercontinental disjunctions. In this study, we use a time-calibrated phylogeny of *Buddleja* to address the following questions: (1) What is the ancestral distribution? (2) When did intercontinental disjunctions arise? (3) What mechanisms best explain intercontinental disjunctions given the inferred timing and direction of divergence events? and (4) What effect have broad-scale changes in distribution had on species diversification?

MATERIALS AND METHODS

Taxonomic sampling

Our dataset comprised 79 species of *Buddleja*, representing the full geographic range of the group. We included eight species (100% of regional species diversity) from continental Africa, six species (75%) from Madagascar, 20 species (83%) from Asia, 15 species (71%) from North America, and 29 species (63%) from South America. For two species, *B. cordata* and *B. elegans*, we included accessions of two different subspecies, so the dataset included 81 taxa of *Buddleja*. We used sequence data from a previous study of the phylogeny of tribe Buddlejeae (Chau et al., 2017), supplemented by data for three additional taxa from North America (*B. cordata* subsp. *cordata*, *B. parviflora*, *B. perfoliata*) and one species from South America (*B. pichinchensis*). We also added a North American accession of *B. americana*, a widespread species whose range encompasses both North and South America, to complement the accession from South America in the previous dataset. Outgroups included *Teedia lucida* and *Phygelius capensis* from the sister group to *Buddleja*, *Scrophularia nodosa* and *Nemesia fruticosa* in Scrophulariaceae, and *Lantana depressa* in Verbenaceae. Voucher information for all accessions is provided in Appendix 1.

Sequence data

Sequences from seven loci were used: ETS, PPR24, PPR97, and PPR123 from the nuclear genome, and rpoA, *trnD-trnT*, and *trnS-trnfM* from the plastid genome. Molecular methods followed those in Chau et al. (2017). Additional sequencing of nuclear loci was done to augment data available for each accession. Amplification of PPR97 in some species was difficult, possibly due to mutations in annealing site of previously used primers (PPR97-781F and PPR97-1585R). Primers internal to these were designed for use in difficult taxa

(PPR97-800R-Bud: 5'-CAGGATGATGTTAGATTATAAGGT-3' and PPR97-1550R-Bud: 5'-ATGATCATACATGGAACCTTCCAGAAC-3').

Newly generated sequences were added to the alignments from the previous study on Buddlejeae (Chau et al., 2017). For each taxon, sequences for all loci were concatenated for analyses. All sequences for nuclear loci were from the same accession. Plastid sequences were from the same or a different accession.

Phylogenetic inference and divergence dating

Phylogenetic and divergence time estimation was performed in BEAST 2 (Bouckaert et al., 2014) on CIPRES Science Gateway (http://www.phylo.org). Loci were partitioned, and the GTR+gamma substitution model was used for each locus based on model testing using the AIC criterion in jModeltest 2.1.4 (Guindon & Gascuel, 2003; Darriba et al., 2012). Loci shared clock and tree models. The clock model was set as an uncorrelated lognormal relaxed clock with the mean modeled as a gamma distribution (alpha: 0.001, beta: 1000). The tree prior was modeled as a birth-death process. We also ran analyses using an uniform prior to model the mean of the clock and the Yule process for the tree prior. Settings for further analyses (gamma distribution, birth-death process) were chosen because they had the lowest AICM value (Baele et al., 2012) as calculated in Tracer v1.6 (Rambaut et al., 2014). Analyses were done with two replicate runs, each conducted for 300 million generations with sampling every 30,000 generations. Convergence was checked by examining estimated sample sizes (>200) and the trace plot for all parameters in Tracer v1.6. The initial 25% of trees were discarded as burnin, and post-burnin trees from replicate runs were combined in LogCombiner v2.4.3 before getting the maximum clade credibility (MCC) tree with median node heights in TreeAnnotator v2.4.1.

We conducted four different dating analyses using two different data sources for secondary calibration of nodes and two different distributions for priors. Magallón et al. (2015) and Tank and Olmstead (in prep) are two studies dating major components of the angiosperm phylogeny that were used for secondary calibration of node ages in our analyses. Normal and lognormal priors are frequently used for secondary calibrations, but uniform priors have been shown to reduce error compared to normal priors in estimates of node ages using secondary calibrations (Schenk, 2016). We refer to these analyses as follows: MLN = secondary calibrations from Magallón et al. (2015) and lognormal priors for nodes, MU = secondary calibrations from Magallón et al. (2015) and uniform priors for nodes, TLN = secondary calibrations from Tank and Olmstead (in prep) and lognormal priors for nodes, TU = secondary calibrations from Tank and Olmstead (in prep) and uniform priors for nodes.

Magallón et al. (2015) conducted an angiosperm-wide study using 792 angiosperm taxa, five plastid and nuclear gene markers, and 151 fossil calibrations distributed throughout the tree, including fifteen within Lamiidae, a large clade containing *Buddleja*. The results of their Bayesian analysis under an uncorrelated lognormal relaxed clock were used to calibrate the root node of our tree, which represents the most recent common ancestor (MRCA) of Scrophulariaceae and Verbenaceae. Their age estimate for this node has a median of 52.93 and 95% highest posterior density (HPD) interval from 44.18 to 66.11. In the analysis using a lognormal prior, we calibrated this node with a mean of 3.97 and standard deviation of 0.12, which corresponds to a median age of 53.0 and 95% HPD interval from 41.9 to 67.0. In the analysis with an uniform prior, the distribution covered the 95% HPD from Magallón et al. (2015).

In the other dating study, Tank and Olmstead (in prep) inferred a dated tree for Lamiidae using Bayesian analysis under an uncorrelated lognormal relaxed clock. They included 195 angiosperm taxa, including 129 lamiid species, and data from nine plastid

markers. Fifteen fossils were used for age calibrations, including eight within Lamiidae. All calibrations were different from those in Magallón et al. (2015), with five fossils being completely different and the others calibrated with a different age distribution and/or on a different node. Inferred node ages were used to calibrate four nodes on our tree. Tank and Olmstead (in prep) inferred the MRCA of *Buddleja* and *Phygelius* to have a median age of 17.83 and 95% HPD of 4.61-36.5, the MRCA of *Buddleja* and *Scrophularia* to have a median age of 25.37 and 95% HPD of 9.89-45.15, the crown age of Scrophulariaceae to have a median age of 53.63 and 95% HPD of 32.24-68.79, and the MRCA of Scrophulariaceae and Verbenaceae to have a median age of 65.34 and 95% HPD of 57.39-76.77. In our analysis with lognormal priors, the age of the MRCA of *Buddleja* and its sister group comprising *Teedia* and *Phygelius* was modeled with a mean of 2.88 and standard deviation of 0.39, which corresponds to a median of 17.8 and 95% HPD from 8.29 to 38.3. The age of the MRCA of *Buddleja* and *Scrophularia* was modeled with a mean of 3.23 and standard deviation of 0.31, which corresponds to a median of 25.28 and 95% HPD of 13.77-46.41. The MRCA of *Buddleja* and *Nemesia*, representing the crown node of Scrophulariaceae, was modeled with a mean of 3.98 and standard deviation of 0.16, corresponding to a median of 53.52 and 95% HPD from 39.11 to 73.23. The age of the root of our tree, representing the MRCA of Scrophulariaceae and Verbenaceae, was modeled with a mean of 4.18 and standard deviation of 0.08, corresponding to a median of 65.37 and 95% HPD of 55.88-76.46. In the analysis with uniform priors, the distributions encompassed the 95% HPD estimates from Tank & Olmstead (in prep) for the four nodes.

Biogeographic analyses

Five major geographic regions were defined based on tectonic history and distribution patterns in *Buddleja*: continental Africa, Madagascar and surrounding islands (Mascarene and

Comoros), Asia, North America including Central America, and South America. Species distributions were coded according to their native distribution as described in taxonomic treatments (Leeuwenberg, 1977, 1979; Norman, 2000). One species, *B. americana*, has a range that substantially encompasses more than one area, North America and South America. For the two accessions of this species included in the study, the distribution was coded according to the accession's provenance. Two other species include a second area as a minor part of its range, *B. crotonoides* whose range is mostly in North America but extends into northern South America, and *B. axillaris*, which is found mainly on Madagascar but also has a few records from eastern continental Africa. Our only accessions for these two species were from the main part of each of its range, but they were coded as polymorphic. Additional analyses were conducted in which *B. crotonoides* and *B. axillaris* were coded according to only the main part of its range. *Teedia lucida* and *Phygelius capensis*, members of *Buddleia*'s sister group which consists entirely of African endemics, were coded as distributed in continental Africa. Remaining outgroups were removed prior to biogeographic analyses since they represent large clades whose members have varied distributions.

We used the R package 'BioGeoBEARS' 0.2.1 (Matzke, 2013) to estimate ancestral distributions under the DEC model (Ree et al., 2005), DIVALIKE model based on dispersal-vicariance analysis (Ronquist, 1997), and BAYAREALIKE model based on Bayesian Inference of Historical Biogeography for Discrete Areas (Landis et al., 2013), with and without the founder-event speciation ("j") parameter (Matzke, 2014). The MCC tree with median branch lengths from the BEAST analysis using MU settings was used.

Ancestral distributions were also inferred using Bayes-Lagrange/Statistical DEC (S-DEC) and the Bayesian binary method (BBM) in RASP 3.2 (Yu *et al*., 2015). For both analyses, trees from the BEAST analysis using MU settings were used. The S-DEC method uses the dispersal-extinction-cladogenesis biogeographical model (Ree & Smith, 2008) and

accounts for topological uncertainty by using a distribution of trees. No range combinations were excluded from the analysis, and the maximum number of areas for the analysis was set at two since that is the maximum for any species' current distribution. Analyses were performed with 5,000 random post-burnin trees from the posterior distribution. The BBM analysis was run for 1,000,000 generations with sampling every 1000 samples, and the initial 100 samples were discarded as burnin. A F81+gamma model was used, and the maximum number of areas was set at 2. Other parameters were left at default values.

Diversification rate analyses

We inferred diversification rate parameters on the phylogenetic trees to determine if changes in distribution were associated with changes in diversification processes. Because the reliability of diversification rate estimation methods is controversial, we used several programs and assessed consistency of results.

We first used BAMM v2.5 (Rabosky, 2014), a Bayesian method that uses reversible jump Markov chain Monte Carlo to explore different models of diversification dynamics, to infer the number and location of shifts in diversification rate dynamics. The reliability of BAMM has been debated because its likelihood function excludes the possibility of unobserved rate shifts on extinct lineages and the compound Poisson process prior used may make the posterior distribution on the number of rate shifts overly sensitive to the prior and diversification rate estimates unreliable (Moore et al., 2016). The authors of the program counter that these issues are irrelevant or unsubstantiated (Rabosky et al., 2017). We used the MCC tree from the BEAST analysis using MU settings from which all taxa outside *Buddleja* were pruned. Evolutionary rate priors were set using the BAMMtools function BAMMpriors. BAMM allows users to specify the sampling fraction of clades on the phylogeny to account for incomplete taxon sampling. We specified sampling fractions of 0.83 (20 taxa sampled/24

taxa total) for the Asian clade, 0.75 (6/8) for the Madagascan clade, 0.61 (46/76) for the New World clade, and 1.0 for all remaining clades. Total numbers of taxa, including subspecies and varieties, in each clade were based on the most recent monographic works for each group and more recent publications of new and resurrected species (Leeuwenberg, 1977, 1979; Norman, 2000; Liu & Peng, 2004, 2006; Morales & González, 2007; Zhang et al., 2014). *Buddleja americana* was counted twice for the New World clade for both sampled and total taxa since our results suggest that it may occur in two separate clades. We ran the analysis for 10 million generations, with sampling every 5000 generations and a acceptance rate reset frequency of 1000 generations. Remaining parameters were set at default values. Analyses of BAMM output was conducted after removing the initial 10% of samples as burnin. Convergence was checked using the R package 'coda' to calculate ESS. We used BAMMtools to evaluate models with different numbers of rate shifts by computing posterior probabilities and Bayes factors comparing models to a zero rate shift model. We also used BAMMtools to compute the shift configuration with the highest posterior probability, the 95% credible set of rate shift configurations, the marginal probabilities of a shift occurring on each branch, and clade-specific speciation rates.

We also used RevBayes v1.0.4 (Landis et al., 2016) to infer the number of diversification rate categories in branch-specific diversification rate models. RevBayes corrects for unobserved rate shifts by using discrete rate categories rather than a continuous distribution. We used the MCC tree from the BEAST analysis using MU settings from which all taxa outside *Buddleja* were pruned. Empirical taxon sampling cannot be used with branch-specific diversification rate estimation to account for incomplete taxon sampling, so we used an uniform taxon sampling fraction of 0.759 (82 taxa sampled/108 species total) for the entire tree. Remaining parameters matched those in the tutorial for Branch-Specific Diversification Rate Estimation. We used the distributions dnBirthDeathMultiRate and

dnHeterogeneousBirthDeath to model a branch-specific birth-death process for models with multiple rate categories. We estimated marginal log likelihoods of two, three, four, and five rate models using stepping-stone sampling and path sampling, and estimated branch-specific diversification rates in the two rate model. We also used the distribution dnBDP to model a constant (one rate) birth-death process and estimated the marginal log likelihood using stepping-stone sampling and path sampling. Log marginal likelihoods were compared using Bayes factors.

We independently tested the location of a diversification rate shift inferred by BAMM using BayesRate v1.65 (Silvestro et al., 2011). BayesRate allows users to specify clade-specific diversification rate models and tests them on a distribution of trees. We used 500 trees from the posterior distribution from the BEAST analysis using MU settings from which all taxa outside *Buddleja* were pruned. We divided our taxa into two partitions according to the results of BAMM (see Results). One partition contained all New World species except *B. normaniae*. The sampling proportion for this partition was set at 0.61 (45 taxa sampled/74 taxa total). The other partition contained all other taxa, and the sampling proportion was set at 0.84 (37/44). The pure-birth model was used, and all other parameters were set at default values. We compared a one-rate model in which both partitions had the same rate with a two-rate model in which the two partitions had different rates. Marginal likelihoods were calculated using thermodynamic integration, and then compared using Bayes factor. Analyses were run with 6 scaling classes and an uniform temperature distribution for 200,000 iterations with sampling every 100 iterations. The burnin was set at 10000 generations. Convergence was checked by calculating ESS in Tracer v1.6.

RESULTS

Sequence data and phylogenetic analyses

We generated forty-eight new sequences for this study, including eight for ETS, four for PPR24, twenty-one for PPR97, and fifteen for PPR123. Genbank numbers for prior sequences used are listed in Appendix 1.

Phylogenetic relationships

Relationships in our trees are congruent with those inferred by Chau et al. (2017) using a similar dataset. Major well-supported lineages include a clade of New World species (section *Buddleja*), a clade of Asian species (section *Alternifoliae*), and a clade of Malagasy species sister to *B. polystachya* (section *Nicodemia*). These are nested within a grade of African species, comprising sections *Chilianthus*, *Gomphostigma*, and the monotypic *Pulchellae* and *Salviifoliae*.

The positions of taxa new to this study are mostly consistent with previous taxonomic classifications (Norman, 2000). *Buddleja pichinchensis* is in a clade with other South American members of series *Cordatae*, and *B. parviflora* is in a clade with other North American *Cordatae*. *Buddleja cordata* subsp. *cordata* is also in the North American *Cordatae* clade, though it is resolved to be more closely related to taxa other than *B. cordata* subsp. *tomentella*. *Buddleja perfoliata* is sister to *B. scordioides*, and together they form series *Scordioides*. Our phylogenetic results suggest that the widespread species *B. americana* may not be monophyletic. Our newly added accession from Mexico is sister to *B. crotonoides*, consistent with the grouping of both species in series *Buddleja*. This clade is sister to the North American *Cordatae*, composing a larger clade of North American polyploid species. The other accession of *B. americana*, from Peru, is separate and part of a clade of South American species in series *Cordatae* and *Lanatae* (Fig. 2).

Divergence dating

Estimates from dating analyses using different data sources for secondary calibrations and different prior distributions were very similar with overlapping 95% HPD intervals (Table 1). Estimated ages tended to be older when using a uniform distribution instead of a lognormal distribution for node age calibrations. The 95% HPD interval tended to be wider when using a single calibration point from Magallón et al. versus four calibration points from Tank & Olmstead. The results of the analysis using calibrations from Magallón et al. and uniform priors (MU) had the lowest AICM value and were chosen for use in further analyses.

The crown age of *Buddleja* indicates that the group most likely originated and began diversifying in the early Miocene (median: 21.7-19 Ma). Section *Buddleja*, the clade of New World species, arose in the mid-Miocene (median: 15-13.2 Ma), as did section *Alternifoliae*, the clade of Asian species (median: 10.4-9.2 Ma). The clade of Madagascan species, comprising most of section *Nicodemia*, has a crown age in the late Miocene (median: 7-6.2 Ma). The time-tree with estimated median ages from MU settings is shown in Fig. 2.

Historical biogeography reconstruction

For the models compared in BioGeoBEARS, inclusion of the founder-event speciation parameter "j" always increased model fit. AIC scores for the three models DEC, DIVALIKE, and BAYAREALIKE including "j" were nearly identical (Table 2), but DEC+J had the highest likelihood and lowest AIC scores, so we focus on comparing the results of this analysis with the results from S-DEC and BBM in RASP.

Results from all biogeographic analyses were similar (Table 3), with or without polymorphic states for two taxa (Supplementary Tables 1 and 2). In general, S-DEC analyses inferred more inclusive ancestral distributions compared with BBM and DEC+J analyses. Reconstructions in BioGeoBEARS without the jump-dispersal parameter were also more

inclusive. An African distribution was reconstructed with the highest support in all three analyses for the most recent common ancestor of *Buddleja*. S-DEC and DEC+J analyses reconstructed an ancestral distribution including North America and Africa with lower probability. The ancestral distribution persisted until there were three separate transitions to North America, Asia, and Madagascar. Only one transition from Africa to each of these areas was inferred. In the New World, a North American, rather than South American, origin is reconstructed as most likely. Given this, two transitions to a South American distribution were reconstructed.

Diversification rate analysis

Diversification rate analyses were determined to have reached convergence based on ESS values (>800 in BAMM, >250 in BayesRate). BAMM analyses showed support for one or two diversification rate shifts in the *Buddleja* phylogeny. Bayes factors provided some support to both one (8.20) and two (9.74) rate shift configurations when compared against a zero rate shift configuration (Table 4). The one rate shift configuration had the highest posterior probability (0.47). In the 95% credible set of shift configurations, four different configurations have a single rate shift. The set also includes a zero rate shift configuration and a two rate shift configuration (Supplementary Fig. 1). The configuration with the highest posterior probability (0.51) showed an increase in speciation rate on the branch leading to the clade of New World species except *B. normaniae*. This branch also had the highest marginal probability of containing a rate shift (Fig. 3).

RevBayes analyses supported a diversification rate model with multiple rate categories. Marginal log likelihoods computed by stepping-stone sampling and path sampling were nearly identical. Bayes factors strongly supported models with two to five rate

categories versus a single rate categoy (Table 5). Bayes factors did not indicate preference for models with more than two rate categories versus two rates (Bayes factor < 0.1).

BayesRate analyses provided strong support (Bayes factor = 19.39) for a two-rate model versus a one-rate model of diversification when partitioned at the branch leading to all New World species except *B. normaniae*, which was found to contain a rate shift in the highest posterior probability configuration from BAMM. Speciation rate in this clade was roughly twice as high as the background rate (Table 6).

DISCUSSION

The results of our phylogenetic, dating, and biogeographic analyses provide a framework for understanding the evolution of the geographic distribution of *Buddleja* in the context of Earth's history. *Buddleja* most likely originated in the early Miocene (~20 Ma) on the African continent, in southern Africa. In the mid to late Miocene, *Buddleja* expanded its range to include the New World, Asia, and Madagascar, mostly likely through single dispersal events from Africa to each of these areas. In the New World, speciation rate likely increased early in the diversification of the group. Although the focus of our study was on broad-scale biogeographic patterns, our framework is also valuable for testing species-level biogeographic and phylogeographic hypotheses in *Buddleja* (e.g. Yue et al., 2012).

Origins in Africa

Biogeographic analyses infer an African origin for *Buddleja* (Fig. 2). Our phylogenetic results show that within *Buddleja*, southern African lineages, including *B. salviifolia*, *B. glomerata*, section *Gomphostigma* comprising *B. incompta* and *B. virgata*, and section *Chilianthus* including *B. auriculata*, *B. dysophylla*, *B. loricata*, and *B. saligna*, form a basal grade. This result supports the hypothesis of Moore (1947), who proposed that southern Africa was the center of origin because of the high diversity of closely related genera there. Although several of the genera he considered are now known to be distantly related (*Nuxia*) or have been combined with *Buddleja* (*Chilianthus*, *Gomphostigma*), higher-level relationships still support an African origin (Oxelman et al., 2005). The sister group to *Buddleja*, comprising tribe Teedieae and *Phygelius*, has an exclusively African distribution. Infrageneric sectional diversity is also still highest in Africa (Chau et al., 2017). The origin of *Buddleja* in Africa fits with the wider pattern of geographic distribution in the family

Scrophulariaceae, which has mostly taxa endemic to southern hemisphere continents and is especially diverse in Africa (Tank et al., 2006).

Species in the basal grade of *Buddleja* occur in southern and southeastern Africa in a variety of habitats, including forest, grassland, desert, and river edges but are especially common in montane forests. Two other species in Africa, *B. pulchella*, which occurs in montane forests in southeastern and east Africa, and *B. polystachya*, which occurs in montane areas of east Africa and the Arabian peninsula, are probably derived from southern African ancestors. This pattern of Afromontane taxa in the highlands of tropical East Africa and northeastern Africa having ancestors from southern Africa has been observed in several other groups of plants (Galley et al., 2007; Devos et al., 2010; Galbany-Casals et al., 2014; Kandziora et al., 2016).

Biogeographic connections between Africa and other regions are common (Linder, 2014). Since the mid-Cretaceous, interchanges with the northern continents are most frequently observed. Studied groups show an asymmetry in movement, with dispersal from the Holarctic into Africa more common (Gheerbrant & Rage, 2006; Gehrke & Linder, 2009). Movement out of Africa, as occurred in *Buddleja*, is much more rare, but has been shown in a few groups (e.g. *Senecio*; Kandziora et al., 2017).

Vicariance hypotheses

From Africa, the tribe spread to the New World, Asia, and Madagascar. One possible explanation for the presence of *Buddleja* in many of the southern continents is that the group may have achieved a widespread distribution when these areas (Africa, South America, Madagascar, India) were part of a continuous Gondwanan landmass in the Mesozoic, and later lineages became isolated by vicariance when the continents separated. India and Madagascar, as part of East Gondwana, had begun to break away from Africa by the late

Jurassic (~160 Ma). India and Madagascar then separated from Antarctica-Australia and then each other, and became isolated islands by the late Cretaceous (~85 Ma). However, there is evidence for biogeographic connections among India, Madagascar, and Africa, possibly through oceanic islands, until the late Cretaceous (Briggs, 2003; Ali & Aitchison, 2008). South America had separated from Africa by the mid-Cretaceous (~96 Ma; Pitman et al., 1993).

Tectonic movements may also be responsible for the biogeographic connections between Africa and northern continents. Africa and South America, as part of Gondwana, had been separated from Europe and North America, as part of Laurasia, by the late Jurassic (145 Ma; Bortolotti & Principi, 2005; Frison de Lamotte et al., 2015). Southeast Asia and the Hengduan-Himalaya region, where most *Buddleja* species in Asia are currently distributed, have never been contiguous with Africa, but rafting on the Indian subcontinent as it moved from Africa to Asia may have provided a connection (~160-35 Ma; Ali & Aitchison, 2008).

Our dating analyses inferred Miocene ages for the origin of the New World (~14 Ma), Asian (~10 Ma), and Madagascan clades (~6.5 Ma) from African ancestors, making them much too young for Gondwanan vicariance or other tectonic movements to be a factor. Dating studies in other groups have also shown that, even when their pattern of relationships corresponds well to the sequence of Gondwanan breakup, divergences between disjunct groups often occurred much more recently than the separation of landmasses (Yoder & Nowak, 2006; Christenhusz & Chase, 2013).

Past connection routes

A second mode by which *Buddleja* may have attained its present disjunct distribution is by migration through intermittent dispersal routes in the Tertiary. An overland route connecting Africa and Asia may have existed in the Miocene through the Arabian peninsula.

At around 20 Ma, the Afro-Arabian plate collided with Eurasia, closing the marine barrier between Africa and Asia formed by the Tethys Sea (Rögl, 1998). Around the same time, the Middle Miocene thermal maximum (17-15 Ma) allowed tropical and subtropical vegetation to expand across Africa and southern Eurasia (Morley, 2000). This landbridge through Arabia has been proposed as an important connection route for several tropical and subtropical plants (Zhou et al., 2012; Yu et al., 2014) and animals (Bernor et al., 1987). Notably in *Buddleja*, the most closely related African species to the Asian clade is *B. polystachya*, which is distributed in central east Africa, the Horn of Africa, and the Arabian peninsula.

Connecting routes between the Old World and New World have also existed. During warm periods in the Tertiary (e.g. early Eocene climatic optimum), the spread of tropical forests at high latitudes allowed for the migration of tropical lineages between Eurasia and North America (Lavin et al., 2000). The North Atlantic landbridge connecting North America with Europe was traditionally thought to have existed only during periods of warmer climates in the Eocene (Tiffney, 1985), which is too early for our reconstructed date of the transition from African to American distributions in *Buddleja*. More recent studies suggest that the landbridge may have continued to exist or reopened as a corridor for temperate vegetation until the Late Miocene (Tiffney, 2008; Denk et al., 2010). This connecting route may have been important for the dispersal of some warm-temperate lineages with distributions in Africa and the New World (e.g. *Asclepias*, Fishbein et al., 2011). Although there are no native species or known fossils of *Buddleja* in Europe, the reconstructed sequence of migration from Africa to North America and then to South America is consistent with a scenario of dispersal from Africa through Europe to North America across the North Atlantic landbridge.

The Bering land bridge is another important migration route for groups between the Old and New Worlds (Raven & Axelrod, 1974). This connection between Asia and North America existed for much of the Tertiary between the middle Eocene and middle Pliocene, and major faunal interchanges occurred in the late Miocene, among other times (Hopkins, 1959). However, by the middle to late Miocene, cooler temperature at the high latitude of the Bering land bridge may not have supported warm-temperate and subtropical groups like *Buddleja* (Tiffney & Manchester, 2001). Biogeographic analyses do not support the hypothesis that Asian ancestors gave rise to the clade of New World *Buddleja*. We inferred an African rather than Asian ancestor for the New World clade, although it is possible that an Asian lineage giving rise to the New World clade went extinct in Asia and the extant clade of Asian species originated separately.

Long-distance dispersal

A third hypothesis explaining intercontinental disjunctions is long-distance transoceanic dispersal. Dispersal may occur any time after formation of the barrier, so it cannot be rejected based on our divergence dating. Long-distance dispersal has been implicated in many groups with disjunct distributions whose divergence times are younger than those required to be explained by vicariance (Yoder & Nowak, 2006; Christenhusz & Chase, 2013). Many members of *Buddleja* have small, light seeds, often with wings, that may be especially conducive to long-distance dispersal by air currents. Their small, light stature may also lend themselves to passive attachment to migrating animals. Many species have unspecialized ecological preferences, typically growing in open, disturbed habitats, which may allow them to more easily establish in new areas (Norman, 2000). The establishment and spread of several *Buddleja* species, including *B. davidii* and *B. madagascariensis*, outside

their native range after anthropogenic transport is suggestive of their opportunistic abilities (Tallent-Halsell & Watt, 2009).

Although general mechanisms for long-distance dispersal in plants have been proposed (Nathan, 2006; Nathan et al., 2008), the nature of long-distance dispersal events between specific landmasses is not well-known. However, some patterns have emerged from previous studies. Many examples of trans-Atlantic distributions in plants are known, and phylogenetic studies show dispersal in both directions between Africa and South America. Sea surface currents may carry propagules in either direction, but winds tend to blow from South America to Africa (Renner, 2004). Long-distance dispersals between tropical Africa and Madagascar and tropical Asia have been proposed (Yuan et al., 2005; Li et al., 2009; Schaefer & Renner, 2010), and possible mechanisms include transport by birds, wind, or ocean currents (Zhou et al., 2012). A stepping-stone route between Africa and Asia may have existed involving Madagascar, the Comoros, and Seychelles islands (Schatz, 1996). Dispersal from Africa to Madagascar has been very important for the formation of the Malagasy flora (Yoder & Nowak, 2006). In *Buddleja*, the distribution of some species (*B. acuminata*, *B. axillaris*) in both Madagascar and East Africa may be an indication of frequent recent interchange between the two areas. The species on Madagascar, in section *Nicodemia*, have fleshy fruits (Leeuwenberg, 1979), which may facilitate dispersal via endozoochory by birds.

Increased diversification rate in the New World

Buddleja has its highest taxonomic diversity in the New World and then Asia, not its area of origin in Africa. Thus, the time-for-speciation effect (Stephens & Wiens, 2003) cannot explain current distribution of species richness in the group. Our diversification rate analyses indicate that an increase in speciation rate likely occurred on the branch leading to the clade of New World species except *B. normaniae*. This fits the idea of "dispersification,"

whereby movement into a new geographic area results in a shift to a higher diversification rate (Moore & Donoghue, 2007). In the New World, *Buddleja* occurs in a variety of biomes, but is most diverse in mountainous regions, in particular the Andes and Mexican highlands (Norman, 2000; Fig. 1). These areas were actively undergoing uplift during the period *Buddleja* was diversifying in the region from the late Miocene to Pliocene (Hoorn et al., 2010). The orogeny created a topographically complex landscape that may have formed new habitats, empty niches, and vicariant events, which would promote speciation (Antonelli & Sanmartín, 2011; Badgley et al., 2017). Uplift of the Hengduan Mountains, where *Buddleja* in Asia is most diverse, was also occurring during this period (Clark et al., 2005; Sun et al., 2011), but our analyses did not support an increase in diversification rate specifically in this clade. Differences between the New World and Asian mountain systems, e.g. total area and climate, may account for the difference in diversification dynamics. A smaller clade size for the Asian taxa may also have constrained the ability of analyses to find statistical support for differences in diversification rate (Xing & Ree, 2017).

Conclusions

Buddleja expanded its range from Africa to the New World, Asia, and Madagascar in the mid to late Miocene, most likely through long-distance dispersal or past connecting routes. These findings add support to the idea that wide distributions in organisms are often the result of relatively recent events. Through the stochastic process of long-distance dispersal and during geologically short periods of climate change, taxa can expand their range and diversify in new areas, with potentially strong impacts on the community composition and dynamics of different regions.

ACKNOWLEDGEMENTS

We thank L. Frost, M. Moore, and the Missouri Botanical Garden herbarium for providing the specimens new to this study; the Index Seminum program and K. Possee of the University of Washington Department of Biology Medicinal Herb Garden for providing seeds of study taxa; W. Rahfeldt for generating the species richness heatmap; and H. Chau for assistance with map graphic design. This study was supported by National Science Foundation (NSF) grants DEB-0542493, DEB-1020369 to RGO and DEB-1311111 to RGO and JHC; a NSF Graduate Research Fellowship to JHC; graduate student grants from the Botanical Society of America, American Society of Plant Taxonomists and the Society of Systematic Biologists to JHC; the Giles Botanical Field Research Award, Sargent Award, Washington Research Foundation and Benjamin Hall Fellowship, and Melinda Denton Writing Fellowship from the University of Washington Department of Biology to JHC; Agencia Nacional de Promoción Científica y Tecnológica grant PICT 2010-0648 to NOL; and National Natural Science Foundation of China (NSFC) grant 30970192 to W-BS.

REFERENCES

Ali, J.R. & Aitchison, J.C. (2008) Gondwana to Asia: Plate tectonics, paleogeography and connectivity of the Indian sub-continent from the Miocene through latest Eocene (166-35 Ma). *Earth-Science Reviews*, **88**, 145-166.

Antonelli, A. & Sanmartín, I. (2011) Why are there so many plant species in the Neotropics? *Taxon*, **60**, 403-414.

Azuma, H., García-Franco, J.G., Rico-Gray, V & Thien, L.B. (2001) Molecular phylogeny of the Magnoliaceae: The biogeography of tropical and temperate disjunctions. *American Journal of Botany*, **88**, 2275-2285.

Badgley, C., Smiley, T.M., Terry, R., Davis, E.B., DeSantis, L.R.G., Fox, D.L., Hopkins, S.S.B., Jezkova, T., Matocq, M.D., Matzke, N., McGuire, J.L., Mulch, A., Riddle, B.R., Roth, V.L., Samuels, J.X., Strömberg, C.A.E. & Yanites, B.J. (2017) Biodiversity and topographic complexity: modern and geohistorical perspectives. *Trends in Ecology and Evolution*, **32**, 211-226.

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M.A. & Alekseyenko, A.V. (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, **29**, 2157-2167.

Bernor, R.L., Brunet, M., Ginsburg, L., Mein, P., Pickford, M., Rögl, F., Sen, S., Steininger, F. & Thomas, H. (1987). A consideration of some major topics concerning Old World Miocene mammalian chronology, migrations and paleogeography. *Geobios*, **20**, 431– 439.

Bortolotti, V. & Principi, G. (2005) Tethyan ophiolites and Pangea break-up. *The Island Arc*, **14**, 442-470.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond, A.J. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Briggs, J.C. (2003) The biogeographic and tectonic history of India. *Journal of Biogeography*, **30**, 381-388.

Carlquist, S. (1967) The biota of long-distance dispersal. V. Plant dispersal to Pacific islands. *Bulletin of the Torrey Botanical Club*, **94**, 129-162.

Chau, J.H., O'Leary, N., Sun, W.B. & Olmstead, R.G. (2017) Phylogenetic relationships in the tribe Buddlejeae (Scrophulariaceae) based on multiple nuclear and plastid markers. *Botanical Journal of the Linnean Society*, **184**, 137-166.

Christenhusz, M.J.M. & Chase, M.W. (2013) Biogeographical patterns of plants in the Neotropics - dispersal rather than plate tectonics is most explanatory. *Botanical Journal of the Linnean Society*, **171**, 277-286.

Clark, M.K., House, M.A., Royden, L.H., Whipple, K.X., Burchfiel, B.C., Zhang, X. & Tang, W. (2005) Late Cenozoic uplift of southeastern Tibet. *Geology*, **33**, 525-528.

Cowie, R.H. & Holland, B.S. (2006) Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. *Journal of Biogeography*, **33**, 193-198.

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

Davis, C.C., Bell, C.D., Mathews, S. & Donoghue, M.J. (2002) Laurasian migration explains Gondwanan disjunctions: Evidence from Malpighiaceae. *Proceedings of the National Academy of Sciences*, **99**, 6833-6837.

de Queiroz, A. (2005) The resurrection of oceanic dispersal in historical biogeography. *Trends in Ecology and Evolution*, **20**, 68-73.

Denk, T., Grímsson, F. & Zetter, R. (2010) Episodic migration of oaks to Iceland: evidence for a North Atlantic "land bridge" in the latest Miocene. *American Journal of Botany*, **97**, 276-287.

Devos, N., Barker, N.P., Nordenstam, B. & Mucina, L. (2010) A multi-locus phylogeny of *Euryops* (Asteraceae, Senecioneae) augments support for the “Cape to Cairo” hypothesis of floral migrations in Africa. *Taxon*, **59**, 57-67.

Dietz, R.S. & Holden, J.C. (1970) Reconstruction of Pangaea: Breakup and dispersion of continents, Permian to present. *Journal of Geophysical Research*, **75**, 4939-4956.

Dupin, J., Matzke, N.J., Särkinen, T., Knapp, S., Olmstead, R.G., Bohs, L. & Smith, S.D. (2016) Bayesian estimation of the global biogeographical history of the Solanaceae. *Journal of Biogeography*, **44**, 887-899.

Fishbein, M., Chuba, D., Ellison, C., Mason-Gamer, R.J. & Lynch, S.P. (2011) Phylogenetic relationships of *Asclepias* (Apocynaceae) inferred from non-coding chloroplast DNA sequences. *Systematic Botany*, **36**, 1008-1023.

Frizon de Lamotte, D., Fourdan, B., Leleu, S., Leparmentier, F. & de Clarens, P. (2015) Style of rifting and the stages of Pangea breakup. *Tectonics*, **34**, 1009-1029.

Galbany-Casals, M., Unwin, M., Garcia-Jacas, N., Smissen, R.D., Susanna, A. & Bayer, R.J. (2014) Phylogenetic relationships in *Helichrysum* (Compositae: Gnaphalieae) and related genera: incongruence between nuclear and plastid phylogenies, biogeographic and morphological patterns, and implications for generic delimitation. *Taxon*, **63**, 608-624.

Galley, C., Bytebier, B., Bellstedt, D.U. & Linder, H.P. (2007) The Cape element in the Afrotemperate flora: from Cape to Cairo? *Proceedings of the Royal Society B*, **274**, 535-543.

Gehrke, B & Linder, H.P. (2009) The scramble for Africa: pan-temperate elements on the African high mountains. *Proceedings of the Royal Society B*, **276**, 2657-2665.

Gheerbrant, E. & Rage, J.-C. (2006) Paleobiogeography of Africa: how distinct from Gondwana and Laurasia. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **241**, 224-246.

Gray, A. (1846) Analogy between the flora of Japan and that of the United States. *American Journal of Science and Arts, series 2*, **2**, 135-136.

Guindon, S. & Gascuel, O. (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology*, **52**, 696-704.

Haddrath, O. & Baker, A.J. (2001) Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proceedings of the Royal Society of London B*, **268**, 939-945.

Heard, S.B & Hauser, D.L. (1995) Key evolutionary innovations and their ecological mechanisms. *Historical Biology*, **10**, 151-173.

Höhna, S., Landis, M.J., Heath, T.A., Boussau, B., Lartillot, N., Moore, B.R., Huelsenbeck, J.P. & Ronquist, F. (2016) RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, **65**, 726-736.

Hooker, J.D. (1853) *The botany of the Antarctic voyage of H.M. Discovery ships "Erebus" and "Terror" in the years 1839-1843. II. Flora Novae-Zelandiae. Part I. Flowering plants*. Lovell Reeve, London.

Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J., Sanmartín, I., Sanchez-Meseguer, A, Anderson, C.L., Figueiredo, J.P., Jaramillo, C., Riff, D., Negri, F.R., Hooghiemstra, H., Lundberg, J., Stadler, T., Särkinen, T. & Antonelli, A. (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, **330**, 927-931.

Hopkins, D.M. (1959) Cenozoic history of the Bering land bridge. *Science*, **129**, 1519-1528.

Hughes, C.E. & Atchison, G.W. (2015) The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. *New Phytologist*, **207**, 275-282.

Humboldt, A. & Bonpland, A. (1805) *Essai sur la géographie des plantes*. Levrault Schoell, Paris.

Kandziora, M., Kadereit, J.W. & Gehrke, B. (2016) Frequent colonization and little in situ speciation in *Senecio* in the tropical alpine-like islands of eastern Africa. *American Journal of Botany*, **103**, 1483-1498.

Kandziora, M., Kadereit, J.W. & Gehrke, B. (2017) Dual colonization of the Palaearctic from different regions in the Afrotropics by *Senecio*. *Journal of Biogeography*, **44**, 147-157.

Landis, M.J., Matzke, N.J., Moore, B.R. & Huelsenbeck, J.P. (2013) Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology*, **62**, 789-804.

Leeuwenberg, A.J.M. (1977) The Loganiaceae of Africa XVI. *Gomphostigma* Turcz. *Mededelingen Landbouwhogeschool Wageningen*, **77**, 15-30.

Leeuwenberg, A.J.M. (1979) The Loganiaceae of Africa XVIII. *Buddleja* II. Revision of the African and Asiatic species. *Mededelingen Landbouwhogeschool Wageningen*, **79**, 1-163.

Li, Y., Dressler, S., Zhang, D. & Renner, S.S. (2009) More Miocene dispersal between Africa and Asia−the case of *Bridelia* (Phyllanthaceae). *Systematic Botany*, **34**, 521-529.

Linder, H.P. (2008) Plant species radiations: where, when, why? *Philosophical Transactions of the Royal Society B*, **363**, 3097-3105.

Linder, H.P. (2014) The evolution of African plant diversity. *Frontiers in Ecology and Evolution*, **2**, 1-14.

Liu, E.-D. & Peng, H. (2004) *Buddleja subcapitata* (Buddlejaceae), a new species from SW Sichuan, China. *Annales Botanici Fennici*, **41**, 467-469.

Liu, E.-D. & Peng, H. (2006) *Buddleja microstachya* (Buddlejaceae), a new species from SW Yunnan, China. *Annales Botanici Fennici*, **43**, 463-465.

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., & Hernández-Hernández, T. (2015) A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist*, **207**, 437-453.

Matzke, N.J. (2013) BioGeoBEARS: biogeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. R package, version 0.2.1. Available at: http://CRAN.R-project.org/package=BioGeoBEARS (accessed x).

Matzke, N.J. (2014) Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, **63**, 951-970.

Moore, B.R. & Donoghue, M.J. (2007) Correlates of diversification in the plant clade Dipsacales: geographic movement and evolutionary innovations. *The American Naturalist*, **170**, S28-S55.

Moore, B.R., Höhna, S., May, M.R., Rannala, B. & Huelsenbeck, J.P. (2016) Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 9569-9574.

Moore, R.J. (1947) Cytotaxonomic studies in the Loganiaceae. I. Chromosome numbers and phylogeny in the Loganiaceae. *American Journal of Botany*, **34**, 527-538.

Morales, J.F. & González, J. (2007) *Buddleja filibracteolata* (Buddlejaceae), una nueva especie para Costa Rica. *Anales del Jardín Botánico de Madrid*, **64**, 161-163.

Morley, R.J. (2000) *Origin and evolution of tropical rain forests*. John Wiley, New York.

Morrone, J.J. & Crisci, J.V. (1995) Historical biogeography: Introduction to methods. *Annual Review of Ecology and Systematics*, **26**, 373-401.

Nathan, R. (2006) Long-distance dispersal of plants. *Science*, **313**, 786-788.

Nathan, R. & Muller-Landau, H.C. (2000) Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology and Evolution*, **15**, 278-285.

Nathan, R., Schurr, F.M., Spiegel, O., Steinitz, O., Trakhtenbrot, A. & Tsoar, A. (2008) Mechanisms of long-distance seed dispersal. *Trends in Ecology and Evolution*, **23**, 638-647.

Norman, E. (2000) Buddlejaceae. *Flora Neotropica Monograph*, **81**, 1-225.

Ogutcen, E., Theriault, J., King, D.B. & Vamosi, J.C. (2017) Diversification rates in Antirrhineae (Plantaginaceae): The contribution of range shifts and pollination modes. *Perspectives in Plant Ecology, Evolution and Systematics*, **26**, 39-52.

Olmstead, R.G. (2013) Phylogeny and biogeography in Solanaceae, Verbenaceae and Bignoniaceae: a comparison of continental and intercontinental diversification patterns. *Botanical Journal of the Linnean Society*, **171**, 80-102.

Oxelman, B., Backlund, M. & Bremer, B. (1999) Relationships of the Buddlejaceae s. l. investigated using parsimony jackknife and branch support analysis of chloroplast ndhF and rbcL sequence data. *Systematic Botany*, **24**, 164-182.

Oxelman, B., Kornhall, P., Olmstead, R.G. & Bremer, B. (2005) Further disintegration of Scrophulariaceae. *Taxon*, **54**, 411-425.

Pitman, W.C.I., Cande, S., LaBreque, J. & Pindell, J. (1993) Fragmentation of Gondwana: the separation of Africa from South America. *Biological relationships between Africa and South America* (ed. by P. Goldblatt), pp. 202-213. Yale University Press, New Haven.

Rabosky, D.L. (2014) Automatic detection of key innovation, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE*, **9**, e89543.

Rabosky, D.L., Mitchell, J.S. & Chang, J. (2017) Is BAMM flawed? Theoretical and practical concerns in the analysis of multi-rate diversification models. *Systematic Biology*, doi: 10.1093/sysbio/syx037.

Rambaut, A., Suchard, M.A., Xie, D. & Drummond, A.J. (2014) Tracer v1.6. Available from http://beast.bio.ed.ac.uk/Tracer.

Raven, P.H & Axelrod, D.I. (1974) Angiosperm biogeography and past continental movements. *Annals of the Missouri Botanical Garden*, **61**, 539-673.

Ree, R.H. & Smith, S.A. (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, **57**, 4-14.

Renner, S. (2004) Plant dispersal across the tropical Atlantic by wind and sea currents. *International Journal of Plant Sciences*, **165**, S23-S33.

Rögl, F. (1998) Palaeogeographic considerations for Mediterranean and Paratethys seaways (Oligocene to Miocene). *Annalen des Naturhistorischen Museums in Wien*, 99, 279-310.

Ronquist, F. (1997) Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology*, **46**, 195-203.

Ronquist, F. & Sanmartín, I. (2011) Phylogenetic methods in biogeography. *Annual Review of Ecology, Evolution, and Systematics*, **42**, 441-464.

Särkinen, T.E., Newman, M.F., Maas, P.J.M., Maas, H., Poulsen, A.D., Harris, D., Richardson, J.E., Clark, A., Hollingsworth, M. & Pennington, R.T. (2007) Recent oceanic long-distance dispersal and divergence in the amphi-Atlantic rain forest genus *Renealmia* L.f. (Zingiberaceae). *Molecular Phylogenetics and Evolution*, **44**, 968-980.

Schaefer, H. & Renner, S.S. (2010) A three-genome phylogeny of *Momordica* (Cucurbitaceae) suggests seven returns from dioecy to monoecy and recent long-distance dispersal to Asia. *Molecular Phylogenetics and Evolution*, **54**, 553-560.

Schatz, G.E. (1996) Malagasy/Indo-Australo-Malesian phytogeographic connections. *Biogeography of Madagascar* (ed. by W.R. Lourenco), pp. 73-83. ORSTOM editions, Paris.

Schenk, J.J. (2016) Consequences of secondary calibrations on divergence time estimates. *PloS ONE*, **11**, e0148228.

Silvestro, D., Schnitzler, J. & Zizka, G. (2011) A Bayesian framework to estimate diversification rates and their variation through time and space. *BMC Evolutionary Biology*, **11**, 311.

Stephens, P.R. & Wiens, J.J. (2003) Explaining species richness from continents to communities: the time-for-speciation effect in emydid turtles. *The American Naturalist*, **161**, 112-128.

Sun, B.-N., Wu, J.-Y., Liu, Y.-S., Ding, S.-T., Li, X.-C., Xie, S.-P., Yan, D.-F. & Lin, Z.-C. (2011) Reconstructing Neogene vegetation and climates to infer tectonic uplift in western Yunnan, China. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **304**, 328-336.

Swenson, U., Hill, R.S. & McLoughlin, S. (2001) Biogeography of *Nothofagus* supports the sequence of Gondwana break-up. *Taxon*, **50**, 1025-1041.

Tallent-Halsell, N.G. & Watt M.S. (2009) The invasive *Buddleja davidii* (butterfly bush). *Botanical Review*, **75**, 292-325.

Tank, D., Beardsley, P.M., Kelchner, S.A. & Olmstead, R.G. (2006) Review of the systematics of Scophulariaceae s.l. and their current disposition. *Australian Systematic Botany*, **19**, 289-307.

Thorne, R.F. (1972) Major disjunctions in the geographic ranges of seed plants. *The Quarterly Review of Biology*, **47**, 365-411.

Tiffney, B.H. (1985) The Eocene North Atlantic Land Bridge: Its importance in Tertiary and modern phytogeography of the Northern Hemisphere. *Journal of the Arnold Arboretum*, **66**, 243-273.

Tiffney, B.H. (2008) Phylogeography, fossils, and Northern Hemisphere biogeography: the role of physiological uniformitarianism. *Annals of the Missouri Botanical Garden*, **95**, 135-143.

Tiffney, B.H. & Manchester, S.R. (2001) The use of geological and paleontological evidence in evaluating phylogeographic hypotheses in the Northern Hemisphere Tertiary. *International Journal of Plant Sciences*, **162**, S3-S17.

Uribe-Convers, S. &Tank, D.C. (2015) Shifts in diversification rates linked to biogeographic movement into new areas: an example of a recent radiation in the Andes. *American Journal of Botany*, **102**, 1854-1869.

von Hagen, K.G. & Kadereit, J.W. (2003) The diversification of *Halenia* (Gentianaceae): ecological opportunity versus key innovation. *Evolution*, **57**, 2507-2518.

Willdenow, K.L. (1792) *Grundriss der Kräuterkunde zu Vorlesungen*. Haude und Spener, Berlin.

Wolfe, J.A. (1975) Some aspects of plant geography of the Northern Hemisphere during the Late Cretaceous and Tertiary. *Annals of the Missouri Botanical Garden*, **62**, 264-279.

Xing. Y. & Ree, R.H. (2017) Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, E3444-E3451.

Yoder, A.D. & Nowak, M.D. (2006) Has vicariance or dispersal been the predominant biogeographic force in Madagascar? Only time will tell. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 405-431.

Yu, Y., Harris, A.J., Blair, C. & He, X. (2015) RASP (Reconstruct Ancestral State in Phylogenies): A tool for historical biogeography. *Molecular Phylogenetics and Evolution*, **87**, 46-49.

Yuan, Y.-M., Wohlhauser, S., Möller, M., Klackenberg, J., Callmander, M.W. & Küpper, P. (2005) Phylogeny and biogeography of *Exacum* (Gentianaceae): a disjunctive distribution in the Indian Ocean basin resulted from long distance dispersal and extensive radiation. *Systematic Biology*, **54**, 21-34.

Yue, L.-L., Chen, G., Sun, W.-B. & Sun, H. (2012) Phylogeography of *Buddleja crispa* (Buddlejaceae) and its correlation with drainage system evolution in southwestern China. *American Journal of Botany*, **99**, 1726-1735.

Zhang. X., Chen, G., Gong, W. & Sun, W. (2014) *Buddleja caryopteridifolia* (Scrophulariaceae), a species to be recognized based on morphology, floral scent, and AFLP data. *Phytotaxa*, **161**, 181-193.

TABLES

Table 1. Median ages and 95% highest posterior density (HPD) intervals in million years (Ma), inferred from Bayesian dating analyses under an uncorrelated lognormal relaxed clock in BEAST 2. Node ages were calibrated using dates from angiosperm or Lamiidae dating studies by Magallón et al. (2015) or Tank & Olmstead (in prep), respectively, and modeled with lognormal or uniform distributions. Node letters correspond to those in Fig. 2.

			Magallón et al.			Tank & Olmstead		
Node	Clade	Calibration Distribution	Median	95% HPD Minimum	95% HPD Maximum	Median	95% HPD Minimum	95% HPD Maximum
a	*Buddleja* Crown	LogNormal	19.36	10.29	30.33	19.03	11.54	27.14
		Uniform	20.28	10.64	30.74	21.65	14.01	29.23
f	Madagascan Stem	LogNormal	9.26	4.24	15.18	9.17	5.02	14.31
		Uniform	9.65	4.67	15.89	10.39	5.82	15.55
g	Madagascan Crown	LogNormal	6.25	2.77	11.15	6.20	3.11	10.41
		Uniform	6.53	2.85	11.36	7.03	3.44	11.39
e	Section *Alternifoliae* (Asia) Stem	LogNormal	11.58	5.71	18.53	11.50	6.91	17.28
		Uniform	12.10	6.18	19.12	13.02	7.92	18.73
h	Section *Alternifoliae* (Asia) Crown	LogNormal	9.25	4.40	14.96	9.19	5.28	13.91
		Uniform	9.66	4.79	15.43	10.43	6.07	15.22
c	Section *Buddleja* (New World) Stem	LogNormal	15.52	8.21	24.63	15.41	9.35	22.36
		Uniform	16.28	8.49	25.05	17.45	10.92	24.04
i	Section Buddleja (New World) Crown	LogNormal	13.31	6.92	21.44	13.22	7.84	19.75
		Uniform	13.96	7.45	22.14	14.97	9.15	21.22

Table 2. Log-likelihoods, AIC scores, and estimated parameter values (d=dispersal, e=extinction, j=jump-dispersal) for six models analyzed in BioGeoBEARS with MCC tree from BEAST analysis using calibrations from Magallón et al. (2015) and uniform distributions. Distribution codings included two taxa with polymorphic states. ΔAICc comparisons are within each modeltype (DEC, DIVALIKE, or BAYAREALIKE). AICc weight comparisons are among all six models. Model with highest likelihood and lowest AIC score highlighted in bold.

Model	LnL	AIC	AICc	ΔAICc	AICc weight	d	e	j
DEC	-45.88	95.76	95.91	3.61	0.079	0.0028	0	0
DEC+J	**-43**	**92**	**92.3**	**0**	**0.48**	**0.0013**	**0**	**0.0066**
DIVALIKE	-45.68	95.36	95.51	1.98	0.096	0.0037	0	0
DIVALIKE+J	-43.61	93.23	93.53	0	0.26	0.0017	0	0.006
BAYAREALIKE	-69.14	142.3	142.4	46.74	0	0.0039	0.015	0
BAYAREALIKE+J	-44.68	95.36	95.66	0	0.089	0.0011	0	0.0081

Table 3. Estimated ancestral distributions and their probabilities from historical biogeography analyses under DEC+J (dispersal-extinction-cladogenesis + jump-dispersal) model in BioGeoBEARS and S-DEC (statistical dispersal-extinction-cladogenesis) and BBM (binary Bayesian method) models in RASP. Analyses under DEC+J and BBM models used MCC tree from BEAST analysis using calibrations from Magallón et al. (2015) and uniform distributions. Analyses under S-DEC model used 5000 random trees from post-burnin posterior distribution of trees from BEAST analysis using calibrations from Magallón et al. (2015) and uniform distributions. Distribution codings included two taxa with polymorphic states. Only ancestral distributions with probabilities > 5% are shown. Node letters correspond to those in Fig. 2.

Node	Clade	BioGeoBEARS: DEC+J		RASP: S-DEC		RASP: BBM	
		Distribution 1 (%)	Distribution 2 (%)	Distribution 1 (%)	Distribution 2 (%)	Distribution 1 (%)	Distribution 2 (%)
a	*Buddleja* Crown	Africa (86.3)	Africa+North America (12.7)	Africa (62.1)	Africa+North America (36.4)	Africa (100)	-
b		Africa (85)	Africa+North America (13.9)	Africa (54)	Africa+North America (44.9)	Africa (99.8)	-
c	"Section *Buddleja* (New World) Stem"	Africa (79.8)	Africa+North America (17.7)	Africa+North America (84.3)	Africa (10.4)	Africa (94.3)	Africa+North America (5.1)
d		Africa (98.4)	-	Africa (95.9)	-	Africa (99.6)	-
e	"Section *Alternifoliae* (Asia) Stem"	Africa (86.6)	Africa+Asia (9.9)	Africa+Asia (100)	-	Africa (96.3)	-
f	Madagascan Stem	Africa (88.4)	Africa+Madagascar (10.7)	Africa (61.2)	Africa+Madagascar (38.8)	Africa (95.8)	-
g	Madagascan Crown	Madagascar (72.1)	Africa+Madagascar (27.2)	Africa+Madagascar (70.6)	Madagascar (29.5)	Madagascar (95.4)	-
h	Section *Alternifoliae* (Asia) Crown	Asia (100)	-	Asia (100)	-	Asia (93.2)	Africa+Asia (6.7)
i	Section Buddleja (New World) Crown	North America (93.7)	North America+ South America (5.4)	North America (83)	North America+ South America (17)	North America (89.2)	Africa+North America (9.6)

Table 4. Posterior probabilities and Bayes factors relative to a zero rate shift model for configurations with different numbers of rate shifts as inferred in BAMM and BAMMtools.

No. of shifts	Posterior probability	Bayes factor (compared to zero rate shift model)
0	0.120	-
1	0.470	8.20
2	0.280	9.74
3	0.093	6.45
4	0.027	3.71
5	0.012	3.25
6	0.003	1.55

Table 5. Marginal log likelihood of branch-specific diversification rate models with different numbers of rate categories, as computed by stepping-stone sampling in RevBayes, and Bayes factors relative to a one rate category model.

No. of rate categories	Marginal Log Likelihood	Bayes Factor (compared to one rate category model)
1	-440.234	-
2	-434.016	6.218
3	-434.097	6.137
4	-434.028	6.206
5	-434.188	6.046

Table 6. Mean speciation rate (and 95% highest posterior density) for clade comprising all New World *Buddleja* except *B. normaniae*, and group comprising all other taxa (Background) as inferred in BAMM and BayesRate.

	BAMM	BayesRate
New World (except B. normaniae)	0.371 (0.221-0.556)	0.346 (0.185-0.548)
Background	0.204 (0.122-0.312)	0.145 (0.077-0.229)

FIGURES

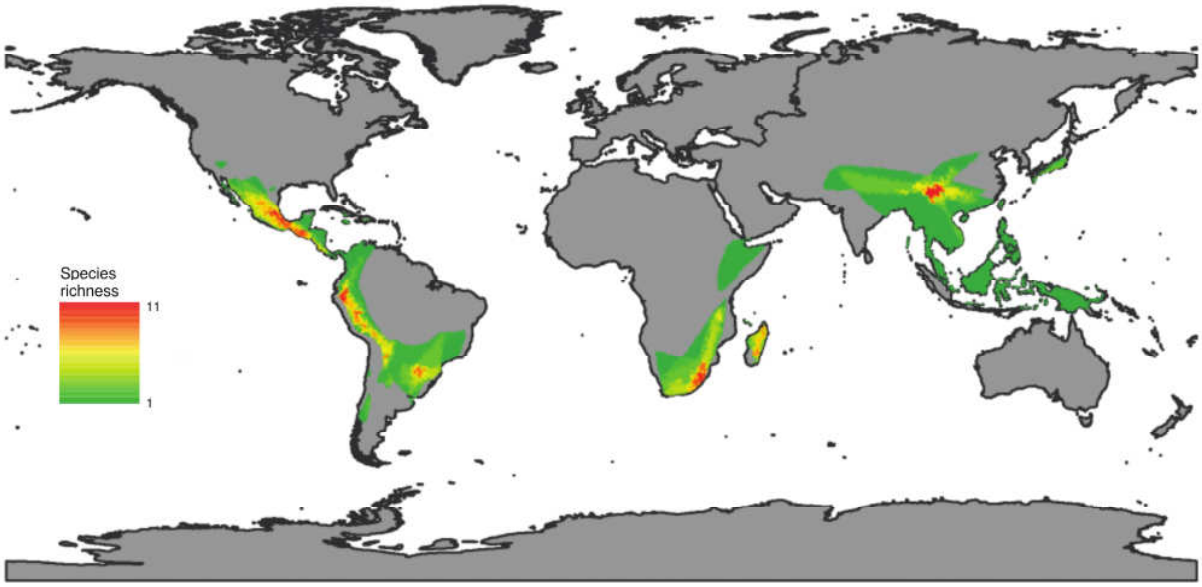


Figure 1. Distribution and species diversity of *Buddleja*. Colors indicate species richness. Maps of individual species distributions generated from occurrence records from the Global Biodiversity Information Facility and descriptions in taxonomic treatments were combined in ArcMap to create a heat map showing distribution of species richness.

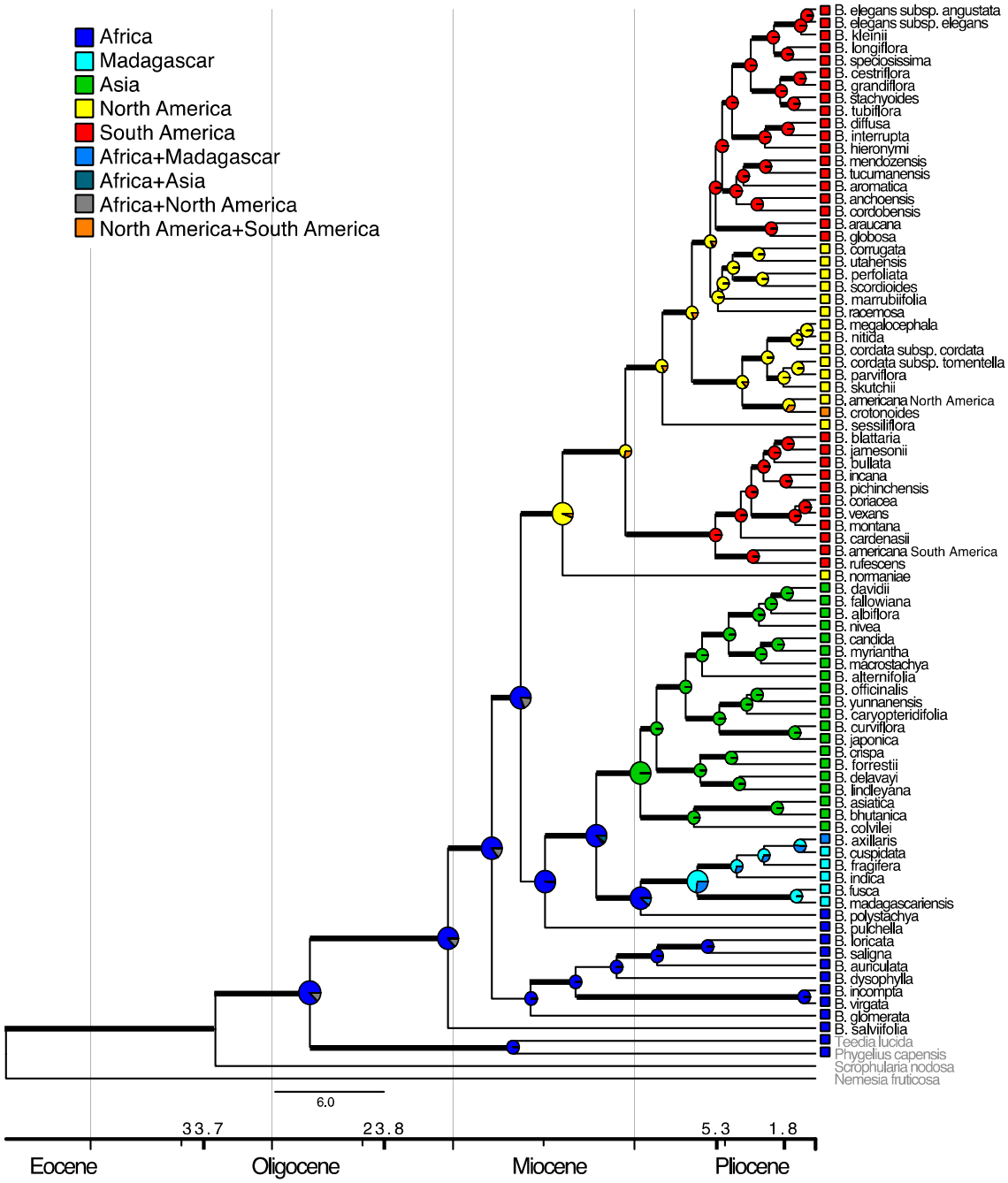


Figure 2. MCC chronogram of *Buddleja* inferred from Bayesian analysis in BEAST using secondary calibrations of node ages from Magallón et al. (2015) and uniform priors. Branches with posterior probability support ≥0.90 are highlighted with thicker branches. Pie charts indicate relative probability of ancestral distributions at nodes from BioGeoBEARS analysis using DEC+J model. Letters at nodes correspond to those in Tables 1 and 3.

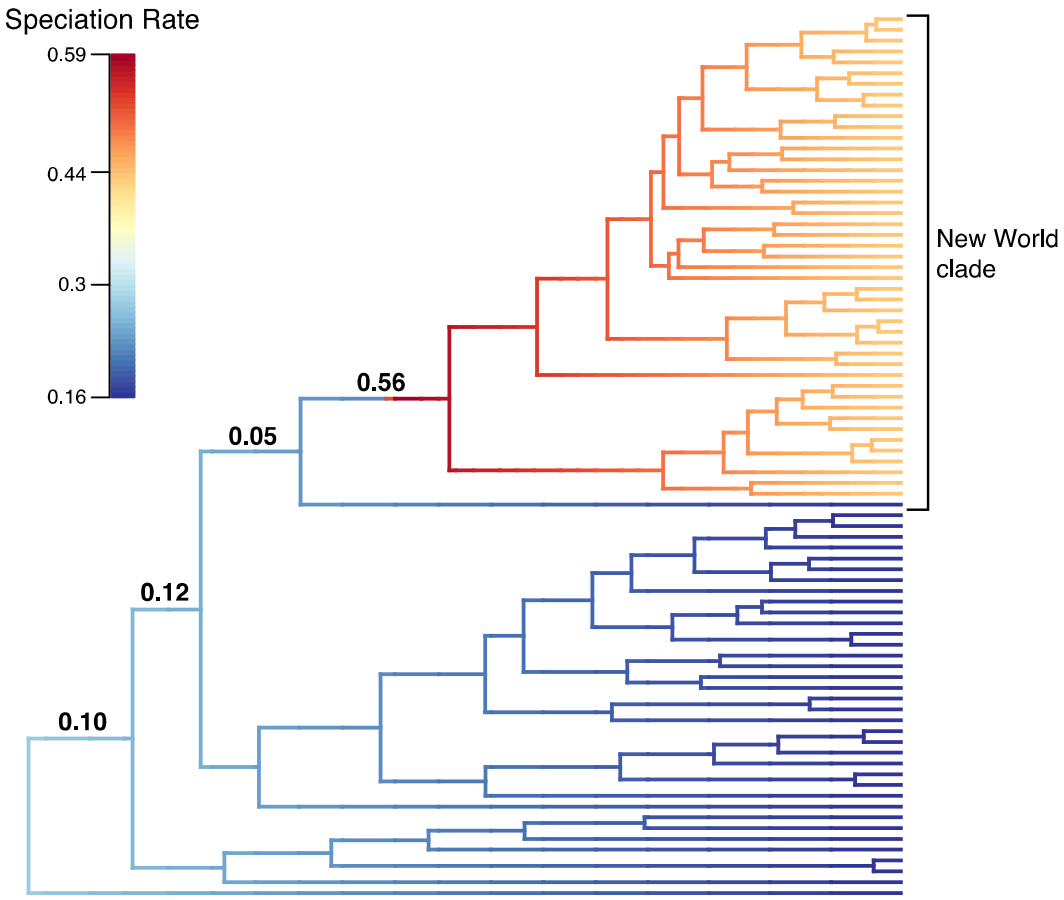


Figure 3. Cladogram of *Buddleja* with branch colors indicating speciation rates in shift configuration with highest posterior probability as inferred by BAMM. Numbers on branches indicate marginal probabilities of a rate shift occurring on that branch where marginal probability > 5%.

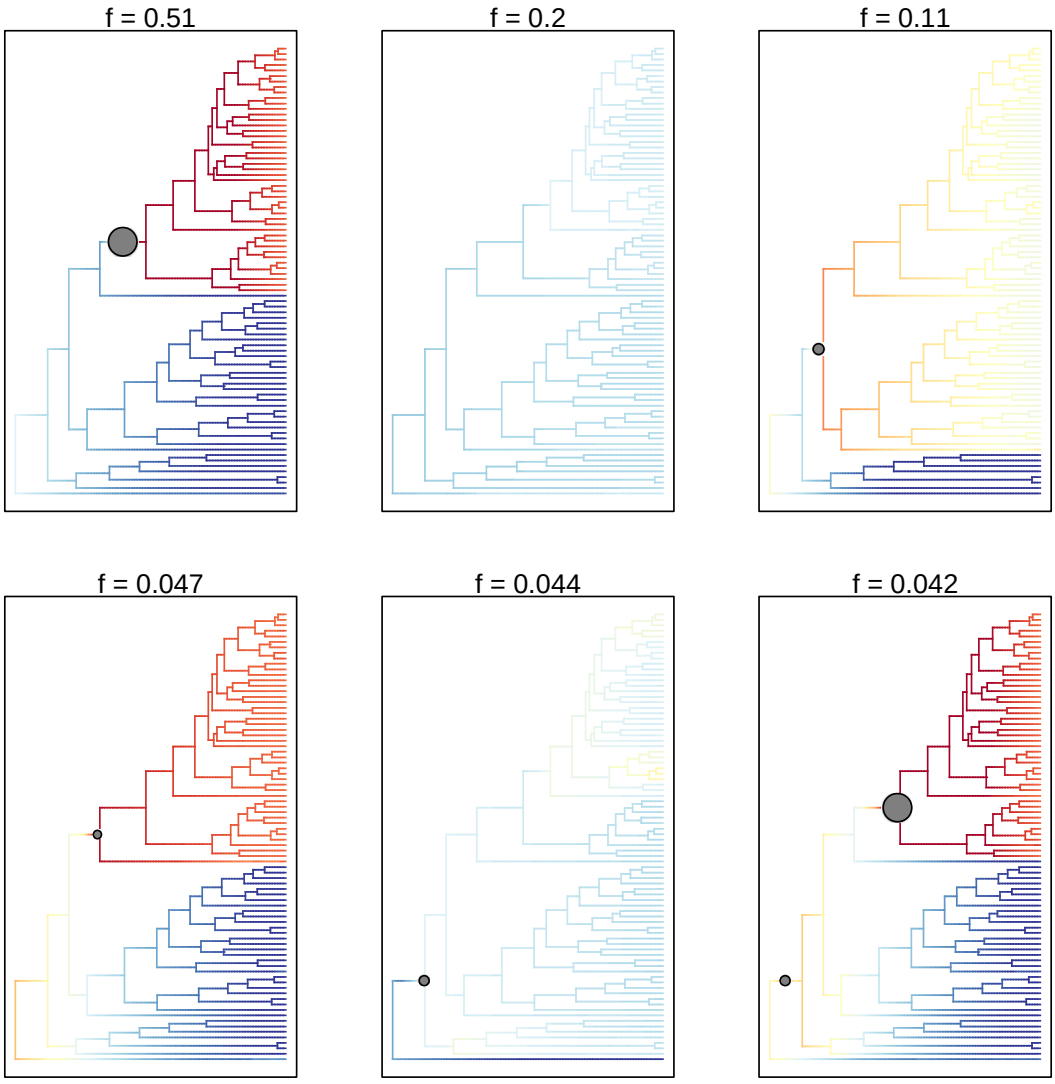
APPENDIX

Supplementary Table 1. Log-likelihoods, AIC scores, and estimated parameter values (d=dispersal, e=extinction, j=jump-dispersal) for six models analyzed in BioGeoBEARS with MCC tree from BEAST analysis using calibrations from Magallón et al. (2015) and uniform distributions. Distribution codings included only single states. ΔAICc comparisons are within each modeltype (DEC, DIVALIKE, or BAYAREALIKE). AICc weight comparisons are among all six models. Model (DEC+J) with highest likelihood and lowest AIC score highlighted in bold.

Model	LnL	AIC	AICc	ΔAICc	AICc weight	d	e	j
DEC	-36	76	76.15	11.62	0.0011	0.0019	0	0
DEC+J	**-29.11**	**64.23**	**64.53**	**0**	**0.36**	**0**	**0**	**0.0079**
DIVALIKE	-34.46	72.91	73.06	8.29	0.0050	0.027	0	0
DIVALIKE+J	-29.23	64.47	64.77	0	0.32	0	0	0.0081
BAYAREALIKE	-58.21	120.4	120.6	55.83	0	0.0028	0.014	0
BAYAREALIKE+J	-29.23	64.47	64.77	0	0.32	0	0	0.0080

Supplementary Table 2. Estimated ancestral distributions and their probabilities from historical biogeography analyses under DEC+J (dispersal-extinction-cladogenesis + jump-dispersal) model in BioGeoBEARS and S-DEC (statistical dispersal-extinction-cladogenesis) and BBM (binary Bayesian method) models in RASP. Analyses under DEC+J and BBM models used MCC tree from BEAST analysis using calibrations from Magallón et al. (2015) and uniform distributions. Analyses under S-DEC model used 5000 random trees from post-burnin posterior distribution of trees from BEAST analysis using calibrations from Magallón et al. (2015) and uniform distributions. Distribution codings included only single states. Only ancestral distributions with probabilities > 5% are shown. Node letters correspond to those in Fig. 2.

Node	Clade	BioGeoBEARS: DEC+J		RASP: S-DEC		RASP: BBM	
		Distribution 1 (%)	Distribution 2 (%)	Distribution 1 (%)	Distribution 2 (%)	Distribution 1 (%)	Distribution 2 (%)
a	*Buddleja* Crown	Africa (88.5)	Africa+North America (11.0)	Africa (56.0)	Africa+North America (42.0)	Africa (100)	-
b		Africa (88.5)	Africa+North America (10.9)	Africa (48.9)	Africa+North America (48.7)	Africa (99.7)	-
c	"Section *Buddleja* (New World) Stem"	Africa (87.8)	Africa+North America (10.5)	Africa+North America (85.2)	Africa (9.6)	Africa (93.9)	Africa+North America (5.6)
d		Africa (99.5)	-	Africa (95.6)	-	Africa (99.5)	-
e	"Section *Alternifoliae* (Asia) Stem"	Africa (97.2)	-	Africa/Asia (100)	-	Africa (94.6)	-
f	Madagascan Stem	Africa (98.3)	-	Africa/Madagascar (99.9)	-	Africa (95.1)	-
g	Madagascan Crown	Madagascar (100)	-	Madagascar (100)	-	Madagascar (95.7)	-
h	Section *Alternifoliae* (Asia) Crown	Asia (100)	-	Asia (100)	-	Asia (95.5)	-
i	Section Buddleja (New World) Crown	North America (99.2)	-	North America (83.3)	North America+ South America (16.7)	North America (93.3)	Africa+North America (6.2)



Supplementary Figure 1. 95% confidence set of diversification rate shift configurations inferred in BAMM.

CHAPTER 3:

Comparison of taxon-specific versus general locus sets for targeted sequence capture for plant phylogenomics

ABSTRACT

Targeted sequence capture is an effective method for efficiently and economically gathering sequence data for large numbers of loci when used in conjunction with multiplexing and high-throughput sequencing platforms. Since universal single-copy nuclear loci target sets do not exist for plants, target loci are often developed individually for a taxon using multiple genomic and/or transcriptomic resources. Another source of information for developing targets are large sets of loci that have been identified as putatively single-copy and having orthologs in a broad range of plants. In this study, we compare the utility for phylogenomics of targets developed for the genus *Buddleja* using a pipeline that identifies "taxon-specific" loci de novo using genome and transcriptome sequence information versus targets developed from three different sets of "general" loci previously identified in diverse taxa. The "taxon-specific" locus set had the greatest number and greatest total length of target loci. The percentage of target sequences with an assembled sequence in *Buddleja* was above 90% for all locus sets, but was highest for one "general" locus set, the pentatricopeptide repeat (PPR) gene family. The PPR loci and "taxon-specific" locus sets also had loci with the highest average variability. We suggest that researchers consider including "general" loci, especially PPR loci, as sequence capture targets for phylogenomic work, especially if genomic resources are not available for their clade of interest. Phylogenomic analyses resolved a well-supported tree for *Buddleja*, although the positions of several taxa remain uncertain.

INTRODUCTION

Recent and rapid diversifications are common in the tree of life (e.g. Hughes and Eastwood, 2006; Seehausen, 2006; Kozak et al., 2015) and often require the input of large amounts of data in order to resolve phylogenetic relationships (Rokas et al., 2003). Multiple unlinked loci are required to accurately reconstruct species trees (Doyle, 1992; Small et al., 2004; Leaché and Rannala, 2011) because gene trees from single loci may not reflect species relationships due to incomplete lineage sorting, unrecognized paralogy, and lateral gene transfer or hybridization between species (Maddison, 1997; Edwards, 2009).

Until recently, many phylogenetic studies in plants have relied on a few loci, particularly in the easily amplified and variable plastid and nuclear ribosomal RNA regions. However, each of these regions represents only a single gene history (Small et al., 2004). Additional nuclear loci have been targeted for development for phylogenetic studies because they are relatively fast-evolving and each nuclear locus potentially represents an independent gene history (Yuan et al., 2009). However, their traditional application using PCR and Sanger sequencing is often difficult because of the need to design primers on an individual basis for each group under study and to test each locus for phylogenetic utility in the group (Hughes et al., 2006; Zimmer and Wen, 2013). In addition, if there are gene duplications or length polymorphisms in the locus, obtaining phased sequence data usually requires the use of labor-intensive cloning techniques (Sang, 2002; Dufresne et al., 2014).

The development of next-generation sequencing (NGS) technologies allows for the efficient sequencing of huge numbers of loci, including in non-model taxa (Egan et al., 2012; Twyford and Ennos, 2012; Soltis et al., 2013). Next-generation sequencing also allows for the separation of sequences from different copies of a locus through bioinformatic techniques, rather

than additional labwork (Griffin et al., 2011; Krasileva et al., 2013; Brassac and Blattner, 2015). Further increases in efficiency in time and cost can be achieved by combining multiplexing techniques with target enrichment, which reduces the proportion of the genome sequenced to subsets that are more likely to be useful (Grover et al., 2012). Various techniques have been developed for target enrichment, including methods based on restriction enzymes, RNA, PCR, and sequence capture (Cronn et al., 2012; McCormack et al, 2012; Lemmon and Lemmon, 2013).

Targeted sequence capture, or hybridization, with oligonucleotide probes is used to isolate targeted sequences from fragmented genomic DNA, which can then be sequenced through NGS (Mamanova et al., 2010). Among the strengths of targeted sequence capture are the ability to target known loci, the ability to obtain sequences flanking both sides of the probe, lower stringency in matching of probes and targets compared to primers for PCR, and the ability to capture sequences even from degraded DNA (Cronn et al., 2012; Lemmon et al. 2012).

The process of developing target loci for sequence capture has been a deterrent to its use in plant phylogenetics because a broadly applicable target locus set, like ultraconserved elements in amniotes (Faircloth et al., 2012), has not been developed in plants since highly conserved sequences are rare in plants (Freeling and Subramanian, 2009; Zheng and Zhang, 2012). In order to identify single-copy nuclear loci with orthologs across a clade of interest, multiple genomic and/or transcriptomic resources for taxa in the clade and bioinformatic expertise are typically necessary (Mandel et al., 2014; Nicholls, 2015; Stephens et al., 2015; Heyduk et al., 2016), although several pipelines requiring minimal bioinformatic skills have been developed (Weitemier et al., 2014; Chamala et al., 2015; Schmickl et al., 2016). Target locus sets identified

with these methods are "taxon-specific" in that they contain loci that are putatively single-copy and have orthologs only in the clade comprising the species whose genomic data are used.

Because genomic resources are currently available for only a small number of taxa and can be expensive to generate, the development of more general target locus sets would facilitate the wider use of targeted sequence capture for plant phylogenetics. Several studies have identified loci that are putatively single-copy and have orthologs across large clades of plants by examining genome and transcriptome data from distantly related species. These include the conserved ortholog set (COSII) in euasterids (Wu et al., 2006), shared single copy nuclear genes (APVO SSC genes) in angiosperms (Duarte et al., 2010), the pentatricopeptide repeat (PPR) gene family in angiosperms (Yuan et al., 2009), other low-copy nuclear genes conserved across angiosperms (Zhang et al., 2012), and universal markers developed for individual families (Chapman et al., 2007; Curto et al., 2012). Utilizing "general" locus sets additionally facilitates the targeting of known loci, which enables the combination of data from different studies. The utility of these general locus sets in comparison with "taxon-specific" locus sets in targeted sequence capture and phylogenomics has not been evaluated (but see Granados Mendoza et al., 2015; Léveillé-Bourret et al., in press; Buddenhagen et al., in prep).

Buddleja section *Alternifoliae* (Scrophulariaceae) is a clade of 24 species of shrubs from Asia that began diversifying approximately 10 Ma (Chau et al., 2017, in prep). Ten species are known to be polyploid (Chen et al., 2007). Taxonomic treatments have varied, especially in the circumscription of species (Marquand, 1930; Leeuwenberg, 1979; Li, 1982). Sequence data from four nuclear loci and three plastid loci were generally insufficient for inferring relationships with good support, but the phylogenetic trees suggest that some previously proposed series (e.g. *Rectiflorae*) are not monophyletic (Chau et al., 2017). The radiation in this group presents an

opportunity to evaluate the utility of different locus sets for targeted sequence capture and phylogenomic analysis at the inter-species level.

In this study, we identified four sets of loci for targeted sequence capture, one "taxon-specific" set identified using genomic and transcriptomic data for *Buddleja* and three "general" sets, consisting of COSII, APVO SSC, and PPR loci. We evaluated the performance of the locus sets in the genus *Buddleja* and several outgroups in terms of assembly of target sequences and phylogenetic informativeness of assembled sequences. We also inferred phylogenetic relationships from assembled sequences for the recently diversified *Buddleja* section *Alternifoliae* and evaluate broader relationships in the genus against previous phylogenetic reconstructions using only a few loci.

METHODS

Whole-genome shotgun sequencing of Buddleja globosa–

One specimen of *Buddleja globosa* growing in the Washington Park Arboretum (WPA) of the University of Washington Botanic Gardens (WPA accession number: 179-99-A, herbarium voucher: R.G. Olmstead 2010-46 [WTU]) was selected for genome sequencing. This species has been shown to be diploid (2n=38; Moore, 1947). We confirmed the specimen's ploidy through chromosome counts from preparations of flower buds (Kato, 1999), and determined genome size using flow cytometry (Bino et al., 1993). Young leaves were picked from the plant and ground after freezing in liquid nitrogen. DNA was immediately extracted from ground tissue using a modified CTAB protocol (Doyle and Doyle, 1987) and purified through isopropanol precipitation. DNA was diluted to a concentration of 10 ng/μL, and 100 μL aliquots were sheared by sonication in a Bioruptor (Diagenode Inc., Denville, New Jersey, USA) with a target size of 300 bp. The sequencing library was prepared with the Illumina TruSeq v2 DNA sample preparation kit (Illumina, Inc., San Diego, California, USA), and quality was checked with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA). The library was sequenced with 100 bp paired-end reads on one lane of an Illumina HiSeq 2000 (Illumina, Inc., San Diego, California, USA) at the QB3 Genomics Sequencing Laboratory at the University of California, Berkeley. Reads were filtered and de novo assembled using CLC Genomics Server 5.0.2 (Qiagen Bioinformatics, Redwood City, California, USA).

Selection of loci for targeted sequence capture–

We took two approaches to selecting loci for sequence capture that are in the nuclear genome and are likely to be single-copy and have orthologs across the genus *Buddleja*. One

approach we term “taxon-specific” because it identifies single-copy loci with orthologs specifically in a group of interest by comparing multiple genomic and/or transcriptomic resources for that group. We used our genomic data for *B*. *globosa* and two transcriptomes for *B*. *davidii* (samples GRFT and XRLM) from the 1000 Plants (1KP) initiative (https://sites.google.com/a/ualberta.ca/onekp/). *Buddleja globosa* and *B*. *davidii* are in sections *Buddleja* and *Alternifoliae*, respectively (Chau et al., 2017), so loci found in both taxa are likely to have orthologs throughout the genus. To select loci, we utilized a modified version of the marker development pipeline Sondovac (Schmickl et al., 2016). Briefly, the pipeline takes genome read data and first removes any reads matching a plastome or mitochondriome reference. We used a plastome from Genbank for *Scrophularia takesimensis* (accession: NC_026202), which is from the same family (Scrophulariaceae) as *Buddleja*, and a mitochondriome for *Salvia miltiorrhiza* (accession: NC_023209), which is from the same order (Lamiales) as *Buddleja*. Sondovac then removes duplicated transcripts from the transcriptome, and finds genome reads matching the unique transcripts, which are then de novo assembled. Assembled contigs from genome reads are filtered for length (contig > 180 bp, total length of all contigs for a transcript > 600 bp) and uniqueness. Remaining contigs were compiled as target sequences.

Our other approach we term “general” because it targets loci that are putatively single-copy and have orthologs in large clades of plants. We selected three sets of loci for targeting. Single-copy orthologous genes (COSII) were identified for the euasterid clade by comparing expressed sequence tag databases for four species of euasterids (*Solanum lycopersicum* and *S. pennellii*, *S. tuberosum*, *Capsicum annuum*, *Coffea canephora*) and *Arabidopsis thaliana* (Wu et al., 2006). Sequences for a subset of 369 COSII genes in *Solanum lycopersicum* were downloaded from the Sol Genomics Network

(ftp://ftp.sgn.cornell.edu/COSII/Rasmus_s_cleantomatoseq.fasta). Duarte et al. (2010) identified a set of 959 single-copy nuclear genes (APVO SSC) shared broadly in angiosperms by comparing genome sequences of three eudicots (*Arabidopsis*, *Populus*, *Vitis*) and one monocot (*Oryza*). Yuan et al. (2009) identified 127 loci in the pentatricopeptide repeat (PPR) gene family that are single-copy and intronless in both *Arabidopsis thaliana* and *Oryza sativa*. Coding sequences for APVO SSC and PPR genes in *Arabidopsis thaliana* were downloaded from The Arabidopsis Information Resource (www.arabidopsis.org). Sequences for three additional genes (WAXY, LFY, CAL) were added for other projects in the group. Sequences in *Buddleja globosa* for genes in the three locus sets were compiled by conducting a BLASTN search of sequences from *Solanum* or *Arabidopsis* against the assembled *B*. *globosa* contigs. The top five hits with a bit score greater than 70 were retained. BLAST hits were then assembled by locus using de novo assembly in Geneious v9.1.6 (Biomatters, Auckland, New Zealand), and consensus sequences were saved if pairwise identity > 95%. For contigs with pairwise identity < 95%, all hits matching those loci were removed because this was inferred to be evidence of the presence of paralogs. Remaining hits and saved consensus sequences were filtered by length (individual sequence > 120 bp, total length of all sequences for a locus > 600 bp). Sequences with > 90% sequence similarity were identified using cd-hit-est (Li and Godzik, 2006), and the longest sequence in each cluster was retained. Remaining sequences were compiled as target sequences.

All target sequences for probe design were checked for duplicates by searching for sequences with > 90% sequence similarity using cd-hit-est (Li and Godzik, 2006). The longest sequence in each cluster was retained. Some sequences from different locus sets had significant overlap at the ends of the sequence and were assumed to be from the same locus. There were assembled by locus using de novo assembly in Geneious v9.1.6, and consensus sequences were

saved if pairwise identity > 95%. Consensus sequences and sequences unused in assembly were used as final target sequences for probe design.

Probe design and manufacture were done by RAPiD Genomics (Gainesville, Florida, USA). Biotinylated RNA probes were 120 bp with 2x tiling density over target sequences. Additional checks were performed to eliminate probes targeting multi-copy loci. Probes with more than ten hits to the assembled *B*. *globosa* genome or with more than 100 matching raw *B*. *globosa* reads were discarded.

Taxon sampling−

Fifty samples were chosen for sequencing (Appendix 1). We were interested in inferring relationships within the Asian clade of *Buddleja* (section *Alternifoliae*), which represents a recent diversification and had poor resolution in previous phylogenetic analyses with seven genetic markers (Chau et al., 2017). We sampled 21 of the 24 species. We were also interested in verifying broader relationships within *Buddleja*, so we sampled 25 other species of *Buddleja*, including at least one representative from each of the seven sections of *Buddleja* (Chau et al., 2017). Additionally, we wanted to test the performance of our probes, which were designed using *Buddleja* genome sequence data, in other taxa. We included *Teedia* (Scrophulariaceae), a member of the sister group to *Buddleja*; *Scrophularia* (Scrophulariaceae), in the same family as *Buddleja*; and *Parmentiera* (Bignoniaceae) and *Lantana* (Verbenaceae), in the same order as *Buddleja*. We also wanted to examine the effectiveness of this method for museum samples, so we included eight samples with DNA extracted from herbarium specimens. Samples from fourteen species known or expected to be polyploid were included to test the utility of this method in detecting and separating sequences from paralogs.

DNA extraction, sequence capture and sequencing–

DNA was extracted from dried leaf tissue, either silica gel-preserved or from an herbarium specimen, using a modified CTAB protocol and purified by isopropanol precipitation. DNA was run on 1% agarose gels to assess DNA quality. DNA concentration was measured with a Qubit (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Samples were diluted or concentrated to attain a concentration of 50 ng/μL, where possible, though some samples had a concentration as low as 2 ng/μL. Volumes of 35-50 μL were submitted for further processing.

Library preparation, sequence capture, and sequencing (Capture-Seq) were done by RAPiD Genomics (Gainesville, Florida, USA). For each sample, 250-1000 ng of genomic DNA, where available, was fragmented to a target size of 400 bp. DNA from herbarium specimens were not additionally fragmented if gel images showed that DNA was already degraded. DNA libraries were constructed by end-repairing the sheared DNA, A-tailing and adapter ligation, barcoding, and PCR amplification. Libraries were pooled by ploidy, and probes were hybridized to the pools to enrich for targets. Enriched pools were combined in equimolar ratios for sequencing, and 100 bp, paired-end reads were sequenced on ~16% of one lane of an Illumina HiSeq 3000 (Illumina, Inc., San Diego, California, USA).

Read processing and assembly–

De-multiplexed reads were provided by RAPiD Genomics. Sequence quality was checked using FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc). Using modified scripts from the pipeline SqCL (https://github.com/singhal/SqCL), Trimmomatic 0.36 (Bolger et al., 2014) was used to remove adapters, barcodes, and poor quality bases using the setting LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 MINLEN:36.

Remaining paired and unpaired reads were assembled using the pipeline HybPiper (Johnson et al., 2016). Briefly, the reads_first.py script sorts reads by target sequence using BWA (Li and Durbin, 2009), then assembles mapped reads for each sequence using the assembler SPAdes (Bankevich et al., 2012), and finally extracts the coding sequence from the assembled contig using Exonerate (Slater and Birney, 2005). For the target sequences used by BWA, we used our target sequences for probe design. In addition, we added sequences for the plastome and external (ETS) and internal (ITS) transcribed spacers of nuclear ribosomal DNA, since these high-copy regions are often sequenced with high coverage despite not being targets of sequence capture (Weitemier et al., 2014). The plastome sequence for *Buddleja globosa* was assembled in Geneious v9.1.6 by conducting a reference-guided assembly of whole genome shotgun sequencing reads against the plastome sequence of *Scrophularia takesimensis* (accession: NC_026202) from Genbank. The ETS sequence was obtained by Sanger sequencing after amplification of the region by PCR (Chau et al., 2017). The ITS region was difficult to sequence by Sanger sequencing because there were multiple peaks in the chromatograms throughout the sequenced region. Instead, we assembled the ITS region from *B*. *globosa* genome reads by first getting the sequences of the 18S, 5.8S, and 26S regions by BLASTing ITS sequences for *Buddleja* from Genbank (AF363671, AJ550579, AJ550577, AJ550578, JF421479) against the assembled contigs for *B*. *globosa* and aligning the query sequences and hits. The 18S, 5.8S, and 26S sequences for *B*. *globosa* were then concatenated and used for a reference-guided assembly of the *B*. *globosa* genome raw reads with ten iterations. The consensus sequence was trimmed to the ITS region between 18S and 26S.

Assembled coding sequences for each sample sorted by locus were compiled using the HybPiper script retrieve_sequences.py. When multiple long-length contigs were assembled by

SPAdes for a locus in a sample, a single contig was chosen based on higher sequencing coverage depth or higher percent identity to the reference sequence. Data on lengths of assembled coding sequences for each target sequence and statistics on assembly efficiency were calculated using the scripts get_seq_lengths.py and hybpiper_stats.py.

Multiple long assembled contigs for a locus in a sample, which may represent paralogs, were identified using the HybPiper script paralog_investigator.py. However, due to a bug in the script, sometimes if the target sequence had repetitive regions, a locus was identified as having paralogs even if only a single contig was assembled. We wrote a script to list loci identified by paralog_investigator.py that actually had multiple assembled contigs for any sample. All assembled coding sequences for every sample for these loci were compiled using the script paralog_retriever.py.

Assembly efficiency for each sample was examined, and any samples with assembled coding sequences for less than 50% of target sequences were excluded from further analyses. Differences among locus sets in the proportion of target sequences with an assembled sequence and in the proportion of the total target length assembled were evaluated with one-way analysis of variance tests blocked by sample. Tukey multiple comparison tests were performed to detect differences in the mean proportion and length of different locus sets. Statistical tests were conducted in R (R Core Team, 2015).

Phylogenetic analyses−

A custom script was used to filter out loci with missing data or with multiple long contigs in any sample. This created a final dataset that was complete for every sample and had no loci with evidence of paralogous sequences.

For each locus, sequences were aligned with MAFFT (Katoh and Standley, 2013) using default parameters. Sites with more than 50% missing data were removed from alignments using the "clean" function in Phyutility (Smith and Dunn, 2008). Concatenated alignments were generated for the different locus sets ("taxon-specific", COSII, APVO SSC, and PPR) using the "concat" function in Phyutility. Since a target locus might be composed of multiple target sequences, we also created concatenated alignments for each locus. Percentage of identical sites were calculated for each locus in Geneious v9.1.6 (Biomatters, Auckland, New Zealand). Differences among locus sets in the average percentage of identical sites were evaluated with one-way analysis of variance tests in R. Tukey multiple comparison tests were performed to detect significant differences in the averages.

Maximum likelihood (ML) trees were inferred for the concatenated alignments for each locus set and for all loci using RAxML v8.0.7 (Stamatakis, 2014). We searched for the best-scoring ML tree and conducted 100 rapid bootstraps. Datasets were unpartitioned, and we used the GTR + gamma model of rate heterogeneity (GTRGAMMA) to model nucleotide substitution.

Concatenated alignments were also used to infer species trees using SVD quartets (Chifman and Kubatko, 2014) in PAUP* v4.0a152 (Swofford, 2003). All possible quartets were evaluated. Trees were selected using QFM quartet assembly, and the multispecies coalescent tree model was used. Ambiguities were distributed. For each analysis, 100 bootstraps were performed.

To test for the effect of length differences among locus sets, we created a dataset using PPR sequence data to match the length of the COSII dataset by randomly sampling positions in the aligned PPR dataset without replacement. A ML tree was inferred in RAxML and a species tree was inferred using SVD quartets using the settings above.

RESULTS

Whole-genome shotgun sequencing of Buddleja globosa–

Our *Buddleja globosa* specimen was confirmed to be diploid (2n=38) in our chromosome counts (Fig. 1), and had a haploid genome size of approximately 996 Mbp, which corresponds well with previous measures of genome size in this species (1C: 0.858 pg = 840 Mbp; Hanson et al., 2001).

One lane of sequencing on an Illumina HiSeq 2000 produced 292,788,924 100-bp paired reads. Filtering removed 35,250,878 reads, and the remaining 257,538,046 reads (88%) were used in the de novo assembly. Of these, 10,569,650 reads were not mapped. The remaining 246,968,396 reads (84.4%) were assembled into 311,304 contigs that had a total length of 343,339,138 bp. Contigs ranged in length from 118 to 166,512 bp and had a N50 of 2,390 bp.

Locus sets for targeted sequence capture–

In total, 2,906 target sequences representing 1,049 loci with a total length of 1,010,028 bp were submitted for probe design (Table 1). Of these, 1,880 target sequences in 708 loci with a total length of 580,437 bp were identified in *Buddleja* using the "taxon-specific" method. The remaining 1,026 target sequences in 341 loci with a total length of 429,591 bp were in the three "general" locus sets. There were 67 COSII loci, 162 APVO SSC loci, and 112 PPR loci. The average locus length was higher in any of the three "general" sets (COSI: 1119 bp, APVO SSC: 1079 bp, PPR: 1605 bp) than in the "taxon-specific" set (820 bp).

***DNA extraction and sequencing*–**

All but one of our samples (B. *rinconensis*) had at least 250 ng DNA extracted for further processing (Appendix 2). All DNA from silica gel-preserved tissues was of high molecular weight. DNA from herbarium specimen tissue varied in quality, but most was degraded (Fig. 2).

Of our 50 samples, 48 were successfully sequenced (Appendix 2). One sample (*B*. *rinconensis*) had a very low starting amount of DNA and produced no mapped reads. Another sample (*B*. *macrostachya*) had sufficient starting DNA, but sequencing failed for unknown reasons. For the 48 remaining samples, between 372,898 and 4,963,618 paired reads were produced. There were no issues with read quality when checked in FastQC. On average, 96% of reads were retained after trimming for low-quality bases and adapter and barcode sequences.

***Read assembly*–**

For each sample, between 44% and 49% of total reads were mapped to the target sequences. For species of *Buddleja*, the HybPiper assembly pipeline produced assembled coding sequences for 91-99% of target sequences. For samples outside *Buddleja*, the number of target sequences without assemblies increased with phylogenetic distance from *Buddleja*. For other members of Scrophulariaceae, 76-90% of target sequences had assemblies. For members of other families in Lamiales, 24-47% of target sequences had assemblies. The sample with the lowest number of assemblies, *Lantana leonariorum*, had a low quantity of starting DNA and is in a different family from *Buddleja*.

All locus sets targeted with probes had high success in assembly. When considering all 48 samples with assemblies, significant differences between locus sets were found in the proportion of target sequences with assemblies ($p<0.01$) and the proportion of total target length

assembled ($p<0.01$). The PPR locus set had the highest success, with an average of 96% of target sequences having an assembly and an average of 97% of total target length assembled. The COSII locus set had the lowest success, with an average of 92% of sequences having an assembly and an average of 94% of total target length assembled.

For samples in *Buddleja*, success of assembly for all locus sets was high, with an average of 94-98% of target sequences with assemblies. In samples outside *Buddleja*, a difference between the "taxon-specific" and "general" locus sets was more apparent. For example, in *Parmentiera* 36% of "taxon-specific" target sequences had assemblies, whereas 65% of "general" target sequences had assemblies.

High-copy regions of the genome included in the assembly pipeline were not adequately assembled. Between 1% and 24% of the plastid genome was assembled for each sample. For ETS, 63% of samples did not have an assembly, and for ITS, 60% did not. These regions were not included in further analyses.

Multiple long contigs, i.e., putative paralogs, were assembled for only a small number of target sequences. In *Buddleja*, an average of eight target sequences had more than one long contig assembled, although the maximum was 126 target sequences in *B. americana*. All samples with a larger number of target sequences with paralogs are polyploids from section *Buddleja* (*B. americana*, *B. blattaria*, *B. coriacea*, *B. nitida*, *B. sessiliflora*). *Scrophularia nodosa* also had a high number of target sequences with paralogs.

Loci filtering–

We used only the 46 samples in Scrophulariaceae with successful sequencing in further analyses in order to have a complete data matrix for a greater number of target sequences. Of the

2906 target sequences, 1524 did not have an assembled coding sequence for at least one sample and 538 had paralogous sequences for at least one sample. These were removed from further analyses, and 1200 target sequences remained for phylogenetic analyses (Table 3). The PPR locus set had the largest percentage of target sequences (58%) remaining after filtering, whereas the COSII locus set had the smallest percentage (29%). The "taxon-specific" locus set had an intermediate percentage of target sequences (43%) retained, though the total number of target sequences (800) was higher than in the three "general" sets (400 total). The PPR locus set also had the longest average length of target sequences (1194 bp), whereas the "taxon-specific" locus set had the shortest (336 bp).

The trimmed alignment had a total length of 510,579 bp, which was split nearly evenly between the "taxon-specific" sequences and the three "general" sequences. The PPR and "taxon-specific" locus sets showed the greatest sequence variability (i.e. lowest percentage of identical sites). Variable sites comprised 35.17% and 36.07%, respectively, of the loci sequences on average, which was significantly higher than the average percentages in the other two "general" locus sets ($p < 0.01$). The COSII locus set had the lowest percentage of variable sites (27.96%).

Phylogenetic analyses−

The ML analysis in RAxML with all sequences concatenated produced a fully resolved and well-supported tree (Fig. 2). All nodes in the Asian clade (section *Alternifoliae*) had full support (bootstrap support [BP] = 100%), and in the tree overall, 37 of 43 nodes (86%) had bootstrap support ≥ 90%. Analyses with single locus sets produced trees with varying level of nodal support, from 93% of nodes with BP ≥ 90% in the tree from "taxon-specific" sequences to 60% of nodes with BP ≥ 90% in the tree from COSII sequences (Fig. 3).

The SVD quartets analyses produced trees with similar topologies to the ML analyses but with less support at nodes. The tree with all sequences concatenated had 63% of nodes with bootstrap support above 90% (Fig. 2). For trees from single locus sets, support varied from 65% of nodes with BP ≥ 90% in the tree from PPR sequences to 47% of nodes with BP ≥ 90% in the tree from COSII sequences (Fig. 4). Topological incongruencies between trees from ML and SVD quartets analyses generally occurred at nodes weakly supported in the SVD quartets tree. Among ML trees from different locus sets, there were several well-supported topological differences, including in the positions of *B*. *asiatica*, *B*. *alternifolia*, *B*. *crispa*, and *B*. *myriantha* in section *Alternifoliae* and multiple relationships in section *Buddleja*, the New World clade (Fig. 3).

When the PPR dataset was subsampled to have the shorter length of the COSII dataset, the topology remained mostly the same, but support values decreased (Fig. 6). In the ML tree, the shorter dataset had 51% of nodes with BP ≥ 90%, whereas the full PPR dataset had 65% of nodes with BP ≥ 90%. In the SVD quartets tree, the shorter dataset had 40% of nodes with BP ≥ 90%, whereas the full PPR dataset had 65% of nodes with BP ≥ 90%.

DISCUSSION

Comparison of locus sets–

We were able to develop a substantially greater number of loci with greater total length for the "taxon-specific" locus set versus any of the "general" locus sets. This was not surprising since closely related species, which were used to develop the "taxon-specific" locus set, are expected to share more loci than distantly related species, which were used to develop the "general" locus sets.

We found significant differences in the performance of the locus sets in assembly efficiency after hybridization and sequencing and in phylogenetic informativeness of loci, although the trend between "taxon-specific" and "general" locus sets was not consistent. The "taxon-specific" and PPR locus sets performed best overall, with a greater number of target sequences with assemblies and loci with more information content. The two other "general" locus sets, COSII and APVO SSC, had lower assembly efficiency and less variable loci.

Recovery of assembled coding sequences for our target sequences was high overall. In *Buddleja*, for which our probes were designed, no sample had less than 80% of target sequences with an assembled sequence in any locus set, and the average for the locus sets ranged from 94% to 98%. The PPR locus set had the highest average percentage of recovered sequences, followed by the "taxon-specific" locus set, then the APVO SSC locus set, and finally the COSII locus set. The proportion of total target length recovered followed the same trend.

Taxa outside *Buddleja* showed a different pattern in recovery efficiency. "General" locus sets consistently outperformed the "taxon-specific" locus set. This pattern is consistent with the fact that the "taxon-specific" locus set was designed using genomic resources in *Buddleja*, so it is unknown whether these loci are single-copy or even present in taxa outside *Buddleja*. On the

other hand, the "general" locus sets include loci which have a high probability of being single-copy and having orthologs in large clades of plants, regardless of phylogenetic distance from *Buddleja*. Recovery efficiency of "general" loci should be affected mostly by the ability of the *Buddleja*-designed probes to capture the target sequences, which depends on the amount of sequence divergence between them. Although recovery of assembled sequences was lower overall for the outgroup taxa, even for *Parmentiera aculeata*, a species in a different family that diverged from *Buddleja* approximately 53 Ma (Magallón et al., 2015), at least 56% of target sequences in the "general" locus sets were recovered. The other species in a different family, *Lantana leonariorum*, had a lower than recommended quantity of DNA available for library preparation, which may explain the overall lower assembly efficiency in this sample. Which "general" locus set performed best varied in the different taxa. The APVO SSC locus set had the highest percentage of target sequences with an assembly in three outgroup species, whereas the PPR locus set had the highest percentage in one species.

The loci in our "taxon-specific" and PPR locus sets had significantly higher average percentages of variable sites than the COSII and APVO SSC locus sets. The PPR loci also had the greatest average length, which in combination with its higher variability may support the inference of well-supported gene trees, which are necessary for a number of species tree methods (e.g., ASTRAL; Mirarab and Warnow, 2015). In our phylogenetic trees from ML analyses with concatenated data, the "taxon-specific" locus set produced the tree with the highest proportion of well-supported nodes, though this result likely was affected by the longer total sequence length for this locus set. The three "general" locus sets all produced trees with lower proportions of well-supported nodes. The APVO SSC locus set had the highest of the three with 74%. In the SVD quartets analyses, the PPR locus set produced the tree with the greatest proportion of well-

supported nodes. The COSII locus set produced trees with the lowest proportion of well-supported nodes in both analyses. Decreasing the size of the PPR dataset to match that of the smallest COSII dataset resulted in lower support in the tree. In fact, fewer nodes were well-supported in the tree from the reduced PPR dataset than in the tree from the COSII dataset.

Recommendations for use of locus sets in sequence capture for plant phylogenomics−

In groups where genomic resources do not exist to design a "taxon-specific" locus set for sequence capture, using "general" locus sets, and in particular the PPR loci, is a good alternative. In our ingroup, both our "taxon-specific" and "general" locus sets had high recovery of sequences, with the PPR locus set having the highest. In our outgroups, recovery of sequences was higher in all "general" locus sets than in the "taxon-specific" locus set. Information content of PPR loci was also high, due to their greater average length and high proportion of variable sites. Although the total number of loci and total sequence length will likely be lower in "general" locus sets than in a "taxon-specific" set, potentially dozens to hundreds of target loci can still be generated, which may be sufficient to resolve relationships.

Even in groups where "taxon-specific" locus sets can be designed, researchers may consider adding PPR loci to their sequence capture targets. In addition to having greater or comparable assembly efficiency and informativeness, PPR loci have other traits which make them desirable for phylogenetic analysis, including a lack of introns which facilitates unambiguous alignment (Yuan et al., 2009).

Designing a "general" target locus set for a group does not require as many genomic resources as designing a "taxon-specific" locus set. However, some source of genomic sequence data is still necessary to design probes with sequences that will adequately complement the

targets in the group of interest. In our study, sequence capture with probes designed for a different genus in the same family or different family in the same order was still able to recover 56-95% of sequences in a "general" locus set. Many genomic resources for plants are now publicly available, including genomes (e.g. Phytozome; https://phytozome.jgi.doe.gov/pz/portal.html) and transcriptomes (e.g. 1KP initiative; https://sites.google.com/a/ualberta.ca/onekp/).

Targeted sequence capture is a suitable method even for samples from herbarium specimens or otherwise have degraded DNA. In our study, sequence recovery was not significantly different in our seven samples from herbarium specimens with adequate DNA for normal library preparation. The average percentage of targets with assembled sequences was 96.1% in samples from herbarium specimens versus 97.6% in samples from silica-preserved tissue. For several of these samples (*B*. *brachystachya*, *B*. *microstachya*, *B*. *subcapitata*), PCR amplification of low-copy nuclear loci had not been successful, but the targeted sequence capture method generated large amounts of sequence data suitable for phylogenetic analysis.

***Phylogenetic relationships in* Buddleja**−

Species relationships in the Asian clade of *Buddleja* have been clarified with this massively larger dataset compared to previous work (Chau et al., 2017). *Buddleja asiatica*, with a wide range extending across montane regions of southern China and Southeast Asia, together with the closely related *B*. *bhutanica*, endemic to the eastern Himalayas (Chau et al., 2017), is sister to the rest of the species. The remaining species fall into two large clades. One consists of polyploids with distributions concentrated in the eastern Himalayas and Hengduan Mountains of southwest China. Species with higher ploidy, including the dodecaploid *B*. *colvilei* and

hexaploids *B*. *forrestii*, *B*. *nivea*, and *B*. *albiflora*, form a basal grade leading to a clade of tetraploids comprising *B*. *davidii*, *B*. *fallowiana*, and *B*. *myriantha*. The tetraploid *B*. *candida* and hexaploid *B*. *macrostachya* are expected to fall in this polyploid clade (Chau et al., 2017). The other large clade includes mostly diploids. The widespread species *B*. *crispa*, found in the western and eastern Himalayas and Hengduan Mountains, and *B*. *alternifolia*, found in montane areas of central China, may be sister species and together are sister to the remaining species in the clade. The recently described species *B*. *microstachya*, which had been described as morphologically similar to *B*. *yunnanensis* (Liu and Peng, 2006), is instead sister to the hexaploid species *B*. *delavayi*. Both species are found in Yunnan Province, China. *Buddleja subcapitata*, another recently described species from Sichuan Province, China, is sister to the morphologically similar *B*. *yunnanensis*, from Yunnan Province (Liu and Peng, 2004). This pair of species is sister to a clade of three other species from southwest China, *B*. *caryopteridifolia*, *B*. *brachystachya*, and *B*. *officinalis*. The three species that have been placed in series *Curviflorae*, *B*. *curviflora*, *B*. *japonica*, and *B*. *lindleyana*, form a monophyletic group. All have long, curved corolla tubes and are native to east Asia in China, Taiwan, and Japan (Leeuwenberg, 1979). Known natural hybrids are proposed to have parents which are sister species in our ML tree from the total concatenated dataset: *B*. *davidii* x *B*. *fallowiana*, *B*. *albiflora* x *B*. *nivea* (=*B*. x *alata*), and *B*. *alternifolia* x *B*. *crispa* (=*B*. x *wardii*). There are also natural hybrids of *B*. *candida* and *B*. *macrostachya* (=*B*. x *griffithii*) and *B*. *macrostachya* and *B*. *forrestii* (Li and Leeuwenberg, 1996), but neither *B*. *candida* nor *B*. *macrostachya* were included in our tree.

Many of the broader relationships in *Buddleja* found previously (Chau et al., 2017) are reflected in our phylogenetic trees. *Buddleja polystachya*, from east Africa, is sister to species from Madagascar. Most of the New World species form a clade, which includes a subclade of

diploid species. North American and South American species form several separate clades. Species from southern Africa form a basal grade. However, in our trees, *Buddleja virgata* in section *Gomphostigma* is sister to the rest of the genus, instead of *B*. *salviifolia*. *Buddleja auriculata* is sister to *B*. *salviifolia* and does not form a clade with the remaining species in section *Chilianthus*, where it had been placed. *Buddleja glomerata*, whose placement had been uncertain in previous molecular phylogenetic work, is supported as being close to species in section *Chilianthus*, where it was placed based on morphology. The position of *B*. *pulchella* is still uncertain; it may be sister to a clade of Old World species outside southern Africa. The position of *B*. *normaniae*, from northern Mexico, has changed most significantly. In our results, it is sister to a large clade of African, Madagascan, Asian, and New World species, rather than being sister to the remaining New World species. If this relationship is correct, it would represent a second New World clade.

Conclusions−

We show in this study that general locus sets, and in particular the PPR loci, are effective targets for sequence capture for phylogenomics. Utilizing general locus sets widens the opportunity to use targeted sequence capture, a method which works for degraded samples and allows for targeting known loci, to groups with few or no genomic resources. Assembly of sequencing reads can be accomplished with a number of different programs and pipelines. Although HybPiper successfully generated assembled coding sequences for the vast majority of target sequences, it did not assemble separate contigs for paralogs of target sequences where they were expected to occur in polyploid species. For groups where polyploidization or hybridization are important parts of the evolutionary history, testing of other assembly methods is suggested.

ACKNOWLEDGEMENTS

The authors thank A. Leaché, T. Gill, and J. Grummer for technical assistance in preparing genomic libraries, use of laboratory and computing equipment, and advice on genomic lab techniques and phylogenomic analyses; P. Lu-Irving for spearheading the initial foray into genomic work; W. Rahfeldt for assistance with programming; and M. Deyholos, D. Soltis, and the 1KP initiative for providing access to transcriptome data.

REFERENCES

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455-477.

Bino, R. J., S. Lanter, H. A. Verhoeven, and H. L. Kraak. 1993. Flow cytometric determination of nuclear replication stages in seed tissues. *Annals of Botany* 72: 181-187.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114-2120.

Brassac, J., and F. R. Blattner. 2015. Species-level phylogeny and polyploid relationships in *Hordeum* (Poaceae) inferred by next-generation sequencing and *in silico* cloning of multiple nuclear loci. *Systematic Biology* 64: 792-808.

Buddenhagen C., A. R. Lemmon, E. M. Lemmon, J. Bruhl, J. Cappa, W. L. Clement, M. J. Donoghue, et al. In prep. Anchored Phylogenomics of Angiosperms I: Assessing the Robustness of Phylogenetic Estimates. *bioRxiv* doi: 10.1101/086298.

Chamala, S., N. García, G. T. Godden, V. Krishnakumar, I. E. Jordon-Thaden, R. De Smet, W. B. Barbazuk, et al. 2015. MarkerMiner 1.0: a new application for phylogenetic marker development using angiosperm transcriptomes. *Applications in Plant Sciences* 3: 1400115.

Chapman, M. A., J. C. Chang, D. Weisman, R. V. Kesseli, and J. M. Burke. 2007. Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). *Theoretical Applied Genetics* 115: 747-755.

Chau, J. H., N. O'Leary, W.-B. Sun, and R. G. Olmstead. 2017. Phylogenetic relationships in tribe Buddlejeae (Scrophulariaceae) based on multiple nuclear and plastid markers. *Botanical Journal of the Linnean Society* 184: 137-166.

Chen, G., W.-B. Sun, and H. Sun. 2007. Ploidy variation in *Buddleja* L. (Buddlejaceae) in the Sino-Himalayan region and its biogeographical implications. *Botanical Journal of the Linnean Society* 154: 305-312.

Chifman, J., and L. Kubatko. 2014. Quartet inference from SNP data under the coalescent. *Bioinformatics* 30: 3317-3324.

Cronn, R., B. J. Knaus, A. Liston, P. J. Maughan, M. Parks, J. V. Syring, and J. Udall. 2012. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* 99: 291-311.

Curto, M. A., P. Puppo, D. Ferreria, M. Nogueira, and H. Meimberg. 2012. Development of phylogenetic markers from single-copy nuclear genes for multi locus, species level analyses in the mint family (Lamiaceae). *Molecular Phylogenetics and Evolution* 63: 758-767.

Doyle, J. J. 1992. Gene trees and species trees - molecular systematics as one-character taxonomy. *Systematic Botany* 17: 144-163.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin, Botanical Society of America* 19: 11-15.

Duarte, J. M., P. K. Wall, P. P. Edger, L. L. Landherr, H. Ma, J. C. Pire, J. Leebens-Mack, et al. 2010. Identifiction of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology* 10: 61.

Dufresne, F., M. Stift, R. Vergilino, and B. K. Mable. 2014. Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology* 23: 40-69.

Edwards, S. V. 2009. Is a new and general theory of molecular systematics emerging? *Evolution* 63: 1-19.

Egan, A. N., J. Schlueter, and D. M. Spooner. 2012. Applications of next-generation sequencing in plant biology. *American Journal of Botany* 99: 175-185.

Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfeld, and R. C. Glenn. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology* 61: 717-726.

Freeling, M., and S. Subramanian. 2009. Conserved noncoding sequences (CNSs) in higher plants. *Current Opinion in Plant Biology* 12: 126-132.

Granados Mendoza, C., J. Naumann, M.-S. Samain, P. Goetghebeur, Y. De Smet, and S. Wanke. 2015. A genomie-scale mining strategy for recovering novel rapidly-evolving nuclear single-copy genes for addressing shallow-scale phylogenetics in *Hydrangea*. *BMC Evolutionary Biology* 15: 132.

Griffin, P. C., C. Robin, and A. A. Hoffmann. 2011. A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to *Poa* grasses. *BMC Biology* 9: 19.

Grover, C. E., A. Salmon, and J. F. Wendel. 2012. Target sequence capture as a powerful tool for evolutionary analysis. *American Journal of Botany* 99: 312-319.

Hanson, L., K. A. McMahon, M. A. T. Johnson, and M. D. Bennett. 2001. First nuclear DNA C-values for 25 angiosperm families. Annals of Botany 87: 251-258.

Heyduk, K., D. W. Trapnell, C. F. Barrett, and J. Leebens-Mack. 2016. Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biological Journal of the Linnean Society* 117: 106-120.

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

Hughes, C. E., R. J. Eastwood, and C. D. Bailey. 2006. From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. *Philosophical Transactions of the Royal Society B* 361: 211-225.

Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. C. Zerega, et al. 2016. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4: 1600016.

Kato, A. 1999. Air drying method using nitrous oxide for chromosome counting in maize. *Biotechnic & Histochemistry* 74: 160-166.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772-780.

Kozak, K. M., N. Wahlberg, A. F. E. Neild, K. K. Dasmahapatra, J. Mallet, and C. D. Jiggins. 2015. Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius* butterflies. *Systematic Biology* 64: 505-524.

Krasileva, K. V., V. Buffalo, P. Bailey, S. Pearce, S. Ayling, F. Tabbita, M. Soria, et al. 2013. Separating homeologs by phasing in the tetraploic wheat transcriptome. *Genome Biology* 14: R66.

Leaché, A. D., and B. Rannala. 2011. The accuracy of species tree estimation under simulation: a comparison of methods. *Systematic Biology* 60: 126-137.

Leeuwenberg, A. J. M. 1979. The Loganiaceae of Africa XVIII. *Buddleja* II. Revision of the African and Asiatic species. *Mededelingen Landbouwhogeschool Wageningen* 79: 1-163.

Lemmon, A. R., S. A. Emme, S.A., E. M. Lemmon. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* 61: 727-744.

Lemmon, E. M., and A. R. Lemmon. 2013. High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44: 99-121.

Léveillé-Bourret, É., J. R. Starr, B. A. Ford, E. M. Lemmon, and A. R. Lemmon. In press. Resolving rapid radiations within angiosperm families using anchored phylogenomics. *Systematic Biology* doi: 10.1093/sysbio/syx050.

Li, P. 1982. A study on the genus *Buddleja* L. of China. *Acta Botanica Yunnanica* 4: 241-247.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754-1760.

Li, P., and A. J. M. Leeuwenberg. 1996. Buddleja. *In* C. Y. Wu, and P. H. Raven [eds.], Flora of China, vol. 15, 329-337. Science Press, Beijing, China.

Li, W., and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658-1659.

Liu, E.-D., and H. Peng. 2004. *Buddleja subcapitata* (Buddlejaceae), a new species from SW Sichuan, China. *Annales Botanici Fennici* 41: 467-469.

Liu, E.-D., and H. Peng. 2006. *Buddleja microstachya* (Buddlejaceae), a new species from SW Yunnan, China. *Annales Botanici Fennici* 43: 463-465.

Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523-536.

Mamanova, L., A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner, A. Kumar, E. Howard, et al. 2010. Target-enrichment strategies for next-generation sequencing. *Nature Methods* 7: 111-118.

Magallón, S., S. Gómez-Acevedo, L. L. Sánchez-Reyes, & T. Hernández-Hernández. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* 207: 437-453.

Mandel, J. R., R. B. Dikow, V. A. Funk, R. R. Masalia, S. E. Staton, A. Kozik, R. W. Michelmore, et al. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: an example from the Compositae. *Applications in Plant Sciences* 2: 1300085.

Marquand C. V. B. 1930. Revision of the Old World species of *Buddleja*. *Kew Bulletin of Miscellaneous Information* 5: 177-208.

McCormack, J. E., S. M. Hird, A. J. Zellmer, B. C. Carstens, and R. T. Brumfield. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* 66: 526-538.

Mirarab, S., and T. Warnow. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31: i44-i52.

Moore, R. J. 1947. Cytotaxonomic studies in the Loganiaceae. I. Chromosome numbers and phylogeny in the Loganiaceae. *American Journal of Botany* 34: 527-538.

Nicholls, J. A., R. T. Pennington, E. J. M. Koenen, C. E. Hughes, J. Hearn, L. Bunnefeld, K. G. Dexter, et al. 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Frontiers in Plant Science* 6: 710.

R Core Team. 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website http://www.R-project.org/ [accessed 26 May 2017].

Rokas, A., B. L. Williams, N. King, and S. B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenetics. *Nature* 425: 798- 804.

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

Schmickl, R., A. Liston, V. Zeisek, K. Oberlander, K. Weitemier, S. C. K. Straub, R. C. Cronn, et al. 2016. Phylogenetic marker development for target enrichment from transcriptomic and genome skim data: the pipeline and its application in southern African *Oxalis* (Oxalidaceae). *Molecular Ecology Resources* 16:1124-1135.

Seehausen, O. 2006. African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B* 273:1987-1998.

Slater, G., and E. Birney. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31.

Small, R. L., R. C. Cronn, and J. F. Wendel. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17: 145-170.

Smith, S. A., and C. W. Dunn. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24: 715-716.

Soltis, D. E., M. A. Gitzendanner, G. Stull, M. Chester, A. Chanderbali, S. Chamala, I. Jordon-Thaden, et al. 2013. The potential of genomics in plant systematics. *Taxon* 62: 886-898.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312-1313.

Stephens, J. D., W. L. Rogers, C. M. Mason, L. A. Donovan, and R. L. Malmberg. 2015. Species tree estimation of diploid *Helianthus* (Asteraceae) using target enrichment. *American Journal of Botany* 102: 910-920.

Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Twyford, A. D., and R. A. Ennos. 2012. Next-generation sequencing as a tool for plant ecology and evolution. *Plant Ecology and Diversity* 5: 411-413.

Weitemier, K., S. C. K. Straub, R. C. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2: 1400042.

Wu, F., L. A. Mueller, D. Crouzillat, V. Pétiard, and S. D. Tanksley. 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics* 174: 1407-1420.

Yuan, Y.-W., C. Liu, H. E. Marx, and R. G. Olmstead. 2009. The pentatricopeptide repeat (PPR) gene family, a tremendous resource for plant phylogenetic studies. *New Phytologist* 182: 272-283.

Zhang, N., L. Zeng, H. Shan, and H. Ma. 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytologist* 195: 923-937.

Zheng, W.-X., and C.-T. Zhang. 2012. Ultraconserved elements between the genomes of the plants *Arabidopsis thaliana* and rice. *Journal of Biomolecular Structure and Dynamics* 26: 1-8.

Zimmer, E. A., and J. Wen. 2012. Using nuclear gene data for plant phylogenetics: progress and prospects. *Molecular Phylogenetics and Evolution* 65: 774-785.

TABLES

Table 1. Characteristics of target locus sets for probe design.

Locus sets	Total target sequences	Total target loci	Total target length (bp)	Average target sequence length (bp)	Average target locus length (bp)
"Taxon-specific"	1880	708	580437	309	820
"General"	1034	344	431226	419	1260
COSII	280	67	74988	268	1119
APVO SSC	572	162	174848	306	1079
PPR	174	112	179755	1033	1605
TOTAL	2906	1049	1010028	348	963

Table 2. Sequencing success, assembly efficiency, and presence of paralogous sequences in each sample. Averages calculated for 44 *Buddleja* samples and for all 48 samples with successful sequencing. Superscripts show significant differences at 0.05 level among locus sets in average proportion of target sequences with assembly and in average total target length assembled using a Tukey multiple comparison test.

Sample	Total DNA (ng)	Total raw reads	Total trimmed reads (% of raw)	Total mapped reads (% of raw)	Total target sequences with assembled coding sequence (% of total target sequences)	Total length of target sequences with assembled coding sequence (% of total target sequence length)
Buddleja albiflora	2439	2327360	2237793 (96%)	1074864 (46%)	2867 (99%)	1005867 (100%)
Buddleja alternifolia	2635	1929508	1846378 (96%)	882612 (46%)	2814 (97%)	985218 (98%)
Buddleja americana	2097	1903722	1836865 (96%)	885319 (47%)	2817 (97%)	985491 (98%)
Buddleja anchoensis	1332	1541712	1485689 (96%)	714972 (46%)	2799 (96%)	981540 (97%)
Buddleja aromatica	2158	2039000	1974719 (97%)	954485 (47%)	2845 (98%)	999882 (99%)
Buddleja asiatica	2629	1783174	1711379 (96%)	820709 (46%)	2823 (97%)	987711 (98%)
Buddleja auriculata	1964	2157144	2065826 (96%)	988506 (46%)	2856 (98%)	992868 (98%)
Buddleja blattaria	1656	1447556	1386802 (96%)	663387 (46%)	2775 (95%)	979239 (97%)
Buddleja brachystachya	279	3683050	3638140 (99%)	1795017 (49%)	2820 (97%)	987081 (98%)
Buddleja caryopteridifolia	1446	2052498	1979901 (96%)	954372 (46%)	2834 (98%)	988725 (98%)
Buddleja colvilei	1944	1853250	1740264 (94%)	823349 (44%)	2817 (97%)	983865 (97%)
Buddleja coriacea	1690	1498552	1432719 (96%)	683814 (46%)	2789 (96%)	978015 (97%)
Buddleja crispa	2714	2530742	2444030 (97%)	1179451 (47%)	2866 (99%)	999414 (99%)
Buddleja curviflora	1163	1733384	1664926 (96%)	798999 (46%)	2815 (97%)	987141 (98%)
Buddleja davidii	1454	2063608	1966719 (95%)	935958 (45%)	2855 (98%)	992937 (98%)
Buddleja delavayi	2422	2207958	2111416 (96%)	1008612 (46%)	2856 (98%)	996837 (99%)
Buddleja dysophylla	1201	2937822	2830220 (96%)	1362500 (46%)	2860 (98%)	991176 (98%)
Buddleja elegans	2856	1197350	1160891 (97%)	561792 (47%)	2634 (91%)	947574 (94%)
Buddleja fallowiana	1916	2138936	2051696 (96%)	983118 (46%)	2847 (98%)	992382 (98%)
Buddleja forrestii	1292	2462226	2361212 (96%)	1131268 (46%)	2857 (98%)	998178 (99%)
Buddleja glomerata	528	2369896	2271643 (96%)	1088307 (46%)	2854 (98%)	988509 (98%)
Buddleja interrupta	1813	1267806	1206474 (95%)	572778 (45%)	2791 (96%)	977706 (97%)
Buddleja japonica	2759	2113824	2018901 (96%)	964282 (46%)	2813 (97%)	987528 (98%)
Buddleja lindleyana	2914	2451874	2376319 (97%)	1151239 (47%)	2836 (98%)	993366 (98%)
Buddleja loricata	1140	2786804	2670518 (96%)	1279501 (46%)	2877 (99%)	995751 (99%)
Buddleja madagascariensis	1851	2601922	2463259 (95%)	1165370 (45%)	2873 (99%)	996018 (99%)
Buddleja marrubiifolia	1338	1551386	1477662 (95%)	703133 (45%)	2835 (98%)	989523 (98%)
Buddleja microstachya	1854	4963618	4864924 (98%)	2381190 (48%)	2891 (99%)	1007031 (100%)
Buddleja myriantha	2604	2099840	2007289 (96%)	958589 (46%)	2850 (98%)	991965 (98%)
Buddleja nitida	2435	2274192	2189504 (96%)	1052472 (46%)	2852 (98%)	992940 (98%)
Buddleja nivea	3077	2043774	1961831 (96%)	941137 (46%)	2845 (98%)	993144 (98%)
Buddleja normaniae	555	1761248	1723674 (98%)	843232 (48%)	2715 (93%)	953595 (94%)
Buddleja officinalis	2421	2847288	2723927 (96%)	1302699 (46%)	2874 (99%)	998727 (99%)
Buddleja polystachya	1762	2852964	2753780 (97%)	1328296 (47%)	2855 (98%)	993105 (98%)
Buddleja pulchella	813	4035056	3975394 (99%)	1955353 (48%)	2855 (98%)	996246 (99%)
Buddleja racemosa	1934	1686714	1606473 (95%)	764316 (45%)	2852 (98%)	991569 (98%)
Buddleja saligna	1691	3294204	3184952 (97%)	1539951 (47%)	2874 (99%)	995121 (99%)
Buddleja salviifolia	1310	2476222	2361059 (95%)	1124310 (45%)	2855 (98%)	992979 (98%)
Buddleja sessiliflora	458	865438	836550 (97%)	404077 (47%)	2568 (88%)	930696 (92%)
Buddleja subcapitata	489	2879700	2796956 (97%)	1357879 (47%)	2844 (98%)	991950 (98%)
Buddleja tucumanensis	1704	2650974	2568142 (97%)	1242037 (47%)	2862 (98%)	1002648 (99%)
Buddleja utahensis	1306	1811458	1744501 (96%)	838026 (46%)	2840 (98%)	990624 (98%)
Buddleja virgata	1369	3910310	3770843 (96%)	1817916 (46%)	2862 (98%)	985062 (98%)
Buddleja yunnanensis	2234	2246266	2147602 (96%)	1025852 (46%)	2861 (98%)	995664 (99%)
Teedia lucida	2529	1669018	1606112 (96%)	773791 (46%)	2620 (90%)	915036 (91%)
Scrophularia nodosa	2099	1921010	1837964 (96%)	880448 (46%)	2210 (76%)	758007 (75%)
Parmentiera aculeata	2227	1271530	1232236 (97%)	598303 (47%)	1359 (47%)	469827 (47%)
Lantana leonariorum	391	372898	350065 (94%)	164457 (44%)	707 (24%)	254751 (25%)
AVERAGE (*Buddleja*)	1765	2302985	2218858 (96%)	1068297 (46%)	2829 (97%)	988468 (98%)
AVERAGE (All)	1769	2220121	2138669 (96%)	1029626 (46%)	2737 (94%)	956046 (95%)
Buddleja macrostachya	2729	7148	2191 (31%)	0 (0%)	0 (0%)	0 (0%)
Buddleja rinconensis	53	16406	7280 (44%)	0 (0%)	0 (0%)	0 (0%)

Sample	"Taxon-specific" target sequences with assembled coding sequence (% of total "taxon-specific" target sequences)	COSII target sequences with assembled coding sequence (% of total COSII target sequences)	APVO SSC target sequences with assembled coding sequence (% of total APVO SSC target sequences)	PPR target sequences with assembled coding sequence (% of total PPR target sequences)
Buddleja albiflora	1862 (99%)	271 (97%)	561 (98%)	173 (99%)
Buddleja alternifolia	1834 (98%)	263 (94%)	546 (95%)	171 (98%)
Buddleja americana	1844 (98%)	259 (93%)	543 (95%)	171 (98%)
Buddleja anchoensis	1842 (98%)	253 (90%)	534 (93%)	170 (98%)
Buddleja aromatica	1858 (99%)	264 (94%)	550 (96%)	173 (99%)
Buddleja asiatica	1840 (98%)	265 (95%)	547 (96%)	171 (98%)
Buddleja auriculata	1856 (99%)	268 (96%)	559 (98%)	173 (99%)
Buddleja blattaria	1812 (96%)	259 (93%)	534 (93%)	170 (98%)
Buddleja brachystachya	1851 (98%)	261 (93%)	537 (94%)	171 (98%)
Buddleja caryopteridifolia	1851 (98%)	260 (93%)	551 (96%)	172 (99%)
Buddleja colvilei	1840 (98%)	264 (94%)	545 (95%)	168 (97%)
Buddleja coriacea	1827 (97%)	257 (92%)	536 (94%)	169 (97%)
Buddleja crispa	1862 (99%)	269 (96%)	564 (99%)	171 (98%)
Buddleja curviflora	1837 (98%)	258 (92%)	550 (96%)	170 (98%)
Buddleja davidii	1858 (99%)	271 (97%)	553 (97%)	173 (99%)
Buddleja delavayi	1861 (99%)	267 (95%)	555 (97%)	173 (99%)
Buddleja dysophylla	1855 (99%)	270 (96%)	561 (98%)	174 (100%)
Buddleja elegans	1760 (94%)	229 (82%)	480 (84%)	165 (95%)
Buddleja fallowiana	1860 (99%)	267 (95%)	549 (96%)	171 (98%)
Buddleja forrestii	1857 (99%)	269 (96%)	558 (98%)	173 (99%)
Buddleja glomerata	1851 (98%)	271 (97%)	561 (98%)	171 (98%)
Buddleja interrupta	1836 (98%)	254 (91%)	529 (92%)	172 (99%)
Buddleja japonica	1838 (98%)	259 (93%)	545 (95%)	171 (98%)
Buddleja lindleyana	1845 (98%)	264 (94%)	557 (97%)	170 (98%)
Buddleja loricata	1860 (99%)	279 (100%)	564 (99%)	174 (100%)
Buddleja madagascariensis	1868 (99%)	271 (97%)	564 (99%)	170 (98%)
Buddleja marrubiifolia	1853 (99%)	268 (96%)	544 (95%)	170 (98%)
Buddleja microstachya	1874 (100%)	275 (98%)	568 (99%)	174 (100%)
Buddleja myriantha	1862 (99%)	266 (95%)	551 (96%)	171 (98%)
Buddleja nitida	1854 (99%)	268 (96%)	559 (98%)	171 (98%)
Buddleja nivea	1857 (99%)	267 (95%)	550 (96%)	171 (98%)
Buddleja normaniae	1767 (94%)	250 (89%)	533 (93%)	165 (95%)
Buddleja officinalis	1868 (99%)	272 (97%)	561 (98%)	173 (99%)
Buddleja polystachya	1856 (99%)	266 (95%)	560 (98%)	173 (99%)
Buddleja pulchella	1863 (99%)	267 (95%)	552 (97%)	173 (99%)
Buddleja racemosa	1856 (99%)	268 (96%)	555 (97%)	173 (99%)
Buddleja saligna	1862 (99%)	275 (98%)	565 (99%)	172 (99%)
Buddleja salviifolia	1855 (99%)	269 (96%)	560 (98%)	171 (98%)
Buddleja sessiliflora	1719 (91%)	227 (81%)	458 (80%)	164 (94%)
Buddleja subcapitata	1860 (99%)	265 (95%)	549 (96%)	170 (98%)
Buddleja tucumanensis	1867 (99%)	269 (96%)	556 (97%)	170 (98%)
Buddleja utahensis	1848 (98%)	266 (95%)	553 (97%)	173 (99%)
Buddleja virgata	1843 (98%)	278 (99%)	567 (99%)	174 (100%)
Buddleja yunnanensis	1856 (99%)	271 (97%)	562 (98%)	172 (99%)
Teedia lucida	1652 (88%)	263 (94%)	543 (95%)	162 (93%)
Scrophularia nodosa	1309 (70%)	229 (82%)	524 (92%)	148 (85%)
Parmentiera aculeata	684 (36%)	156 (56%)	409 (72%)	110 (63%)
Lantana leonariorum	322 (17%)	86 (31%)	225 (39%)	74 (43%)
AVERAGE (*Buddleja*)	1845 (98%)	264 (94%)	549 (96%)	171 (98%)
AVERAGE (All)	1774 (94%)[ab]	258 (92%)[c]	538 (94%)[b]	167 (96%)[a]
Buddleja macrostachya	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Buddleja rinconensis	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Sample	Total length of "taxon-specific" target sequences with assembled coding sequence (% of total "taxon-specific" target sequence length)	Total length of COSII target sequences with assembled coding sequence (% of total COSII target sequence length)	Total length of APVO SSC target sequences with assembled coding sequence (% of total APVO SSC target sequence length)	Total length of PPR target sequences with assembled coding sequence (% of total PPR target sequence length)
Buddleja albiflora	578979 (100%)	73416 (98%)	172050 (98%)	181422 (101%)
Buddleja alternifolia	564198 (97%)	72330 (96%)	169830 (97%)	178860 (100%)
Buddleja americana	566616 (98%)	71232 (95%)	168741 (97%)	178902 (100%)
Buddleja anchoensis	564861 (97%)	70035 (93%)	168069 (96%)	178575 (99%)
Buddleja aromatica	577983 (100%)	72150 (96%)	170835 (98%)	178914 (100%)
Buddleja asiatica	566256 (98%)	72336 (96%)	170622 (98%)	178497 (99%)
Buddleja auriculata	568677 (98%)	72948 (97%)	172215 (98%)	179028 (100%)
Buddleja blattaria	558126 (96%)	71358 (95%)	168093 (96%)	181662 (101%)
Buddleja brachystachya	569202 (98%)	71568 (95%)	167901 (96%)	178410 (99%)
Buddleja caryopteridifolia	567498 (98%)	71904 (96%)	170613 (98%)	178710 (99%)
Buddleja colvilei	563940 (97%)	72054 (96%)	169746 (97%)	178125 (99%)
Buddleja coriacea	560007 (96%)	70983 (95%)	168309 (96%)	178716 (99%)
Buddleja crispa	574776 (99%)	73164 (98%)	172704 (99%)	178770 (99%)
Buddleja curviflora	567492 (98%)	70995 (95%)	169938 (97%)	178716 (99%)
Buddleja davidii	570000 (98%)	73335 (98%)	170886 (98%)	178716 (99%)
Buddleja delavayi	573270 (99%)	72663 (97%)	171246 (98%)	179658 (100%)
Buddleja dysophylla	566532 (98%)	73176 (98%)	172419 (99%)	179049 (100%)
Buddleja elegans	545775 (94%)	66249 (88%)	158847 (91%)	176703 (98%)
Buddleja fallowiana	570099 (98%)	72942 (97%)	170547 (98%)	178794 (99%)
Buddleja forrestii	574725 (99%)	72873 (97%)	171513 (98%)	179067 (100%)
Buddleja glomerata	565074 (97%)	72657 (97%)	172119 (98%)	178659 (99%)
Buddleja interrupta	560952 (97%)	70728 (94%)	167193 (96%)	178833 (99%)
Buddleja japonica	566673 (98%)	71724 (96%)	169590 (97%)	179541 (100%)
Buddleja lindleyana	570303 (98%)	71826 (96%)	172080 (98%)	179157 (100%)
Buddleja loricata	568935 (98%)	74520 (99%)	172779 (99%)	179517 (100%)
Buddleja madagascariensis	571761 (99%)	72933 (97%)	172755 (99%)	178569 (99%)
Buddleja marrubiifolia	568347 (98%)	72963 (97%)	169788 (97%)	178425 (99%)
Buddleja microstachya	580857 (100%)	73947 (99%)	173415 (99%)	178812 (99%)
Buddleja myriantha	570192 (98%)	72762 (97%)	170481 (98%)	178530 (99%)
Buddleja nitida	569679 (98%)	72636 (97%)	171603 (98%)	179022 (100%)
Buddleja nivea	570342 (98%)	72723 (97%)	170832 (98%)	179247 (100%)
Buddleja normaniae	538533 (93%)	70212 (94%)	167562 (96%)	177288 (99%)
Buddleja officinalis	573645 (99%)	73590 (98%)	172356 (99%)	179136 (100%)
Buddleja polystachya	569733 (98%)	72453 (97%)	171951 (98%)	178968 (100%)
Buddleja pulchella	573279 (99%)	72552 (97%)	171012 (98%)	179403 (100%)
Buddleja racemosa	568218 (98%)	72747 (97%)	171498 (98%)	179106 (100%)
Buddleja saligna	569283 (98%)	73962 (99%)	172983 (99%)	178893 (100%)
Buddleja salviifolia	568737 (98%)	73218 (98%)	172296 (99%)	178728 (99%)
Buddleja sessiliflora	532899 (92%)	66048 (88%)	154200 (88%)	177549 (99%)
Buddleja subcapitata	571227 (98%)	72093 (96%)	169941 (97%)	178689 (99%)
Buddleja tucumanensis	580179 (100%)	72552 (97%)	171471 (98%)	178446 (99%)
Buddleja utahensis	567846 (98%)	72684 (97%)	171354 (98%)	178740 (99%)
Buddleja virgata	558453 (96%)	74376 (99%)	173133 (99%)	179100 (100%)
Buddleja yunnanensis	570924 (98%)	73509 (98%)	172338 (99%)	178893 (100%)
Teedia lucida	498321 (86%)	71223 (95%)	168546 (96%)	176946 (98%)
Scrophularia nodosa	378360 (65%)	58569 (78%)	157818 (90%)	163260 (91%)
Parmentiera aculeata	187875 (32%)	39657 (53%)	121932 (70%)	120363 (67%)
Lantana leonariorum	84819 (15%)	24111 (32%)	72294 (41%)	73527 (41%)
AVERAGE (*Buddleja*)	567161 (98%)	72207(96%)	170224 (97%)	178876 (100%)
AVERAGE (All)	543843 (94%)[b]	70223 (94%)[b]	166884 (95%)[b]	175097 (97%)[a]
Buddleja macrostachya	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Buddleja rinconensis	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Sample	Total length of assembled plastid sequence (% of total plastid target sequence length)	Total length of assembled ETS sequence (% of total ETS target sequence length)	Total length of assembled ITS sequence (% of total ITS target sequence length)	Sequences with paralogs [corrected] (% of total target sequences)
Buddleja albiflora	8508 (7%)	0 (0%)	0 (0%)	3 (0%)
Buddleja alternifolia	9084 (7%)	0 (0%)	0 (0%)	0 (0%)
Buddleja americana	1953 (2%)	0 (0%)	0 (0%)	126 (4%)
Buddleja anchoensis	3312 (3%)	0 (0%)	0 (0%)	0 (0%)
Buddleja aromatica	9648 (8%)	489 (98%)	0 (0%)	1 (0%)
Buddleja asiatica	10632 (8%)	489 (98%)	909 (99%)	1 (0%)
Buddleja auriculata	15972 (13%)	489 (98%)	669 (73%)	3 (0%)
Buddleja blattaria	3531 (3%)	0 (0%)	0 (0%)	39 (1%)
Buddleja brachystachya	7905 (6%)	489 (98%)	915 (99%)	0 (0%)
Buddleja caryopteridifolia	30558 (24%)	0 (0%)	909 (99%)	0 (0%)
Buddleja colvilei	987 (1%)	0 (0%)	0 (0%)	1 (0%)
Buddleja coriacea	17037 (13%)	0 (0%)	0 (0%)	27 (1%)
Buddleja crispa	26469 (21%)	489 (98%)	909 (99%)	3 (0%)
Buddleja curviflora	18972 (15%)	486 (97%)	0 (0%)	0 (0%)
Buddleja davidii	8073 (6%)	489 (98%)	0 (0%)	5 (0%)
Buddleja delavayi	20382 (16%)	0 (0%)	0 (0%)	3 (0%)
Buddleja dysophylla	14883 (12%)	489 (98%)	909 (99%)	5 (0%)
Buddleja elegans	11283 (9%)	0 (0%)	0 (0%)	0 (0%)
Buddleja fallowiana	12288 (10%)	489 (98%)	0 (0%)	4 (0%)
Buddleja forrestii	8241 (6%)	0 (0%)	0 (0%)	4 (0%)
Buddleja glomerata	4827 (4%)	0 (0%)	0 (0%)	5 (0%)
Buddleja interrupta	1245 (1%)	0 (0%)	0 (0%)	1 (0%)
Buddleja japonica	9240 (7%)	486 (97%)	0 (0%)	0 (0%)
Buddleja lindleyana	11310 (9%)	489 (98%)	909 (99%)	1 (0%)
Buddleja loricata	19935 (16%)	489 (98%)	0 (0%)	5 (0%)
Buddleja madagascariensis	11244 (9%)	489 (98%)	0 (0%)	3 (0%)
Buddleja marrubiifolia	9519 (8%)	489 (98%)	201 (22%)	0 (0%)
Buddleja microstachya	13284 (10%)	0 (0%)	129 (14%)	9 (0%)
Buddleja myriantha	7794 (6%)	0 (0%)	240 (26%)	2 (0%)
Buddleja nitida	28482 (22%)	489 (98%)	327 (36%)	46 (2%)
Buddleja nivea	3285 (3%)	0 (0%)	0 (0%)	1 (0%)
Buddleja normaniae	21129 (17%)	0 (0%)	810 (88%)	4 (0%)
Buddleja officinalis	7038 (6%)	489 (98%)	912 (99%)	2 (0%)
Buddleja polystachya	9138 (7%)	0 (0%)	291 (32%)	1 (0%)
Buddleja pulchella	9960 (8%)	0 (0%)	903 (98%)	6 (0%)
Buddleja racemosa	9681 (8%)	0 (0%)	0 (0%)	1 (0%)
Buddleja saligna	23904 (19%)	489 (98%)	906 (98%)	7 (0%)
Buddleja salviifolia	6435 (5%)	0 (0%)	0 (0%)	2 (0%)
Buddleja sessiliflora	9270 (7%)	0 (0%)	0 (0%)	44 (2%)
Buddleja subcapitata	7434 (6%)	0 (0%)	0 (0%)	0 (0%)
Buddleja tucumanensis	11145 (9%)	0 (0%)	0 (0%)	2 (0%)
Buddleja utahensis	26802 (21%)	0 (0%)	0 (0%)	1 (0%)
Buddleja virgata	27693 (22%)	486 (97%)	906 (98%)	2 (0%)
Buddleja yunnanensis	8289 (7%)	489 (98%)	0 (0%)	0 (0%)
Teedia lucida	12000 (9%)	486 (97%)	909 (99%)	8 (0%)
Scrophularia nodosa	22956 (18%)	486 (97%)	855 (93%)	166 (6%)
Parmentiera aculeata	6423 (5%)	0 (0%)	792 (86%)	-
Lantana leonariorum	5733 (5%)	0 (0%)	615 (67%)	-
AVERAGE (*Buddleja*)	12223 (10%)	211 (42%)	267 (29%)	8 (0%)
AVERAGE (All)	12186 (10%)	214 (43%)	311 (34%)	-
Buddleja macrostachya	0 (0%)	0 (0%)	0 (0%)	(0%)
Buddleja rinconensis	0 (0%)	0 (0%)	0 (0%)	(0%)

Table 3. Characteristics of assembled sequence datasets used for phylogenetic analyses. Target sequences with missing data or paralogous sequences in any sample were removed from datasets. In parentheses are percentages of total targets. Superscripts show significant differences at 0.05 level from a Tukey multiple comparison test.

Locus sets	Total sequences	Total loci	Average sequence length (bp)	Average locus length (bp)	Average total length - unaligned (bp)	Total length - aligned, trimmed (bp)	Average % variable sites
"Taxon-specific"	800 (43%)	511 (72%)	336	526	268710 (46%)	268603 (46%)	36.07%[a]
"General"	400 (39%)	261 (76%)	605	928	242161 (56%)	242359 (56%)	30.55%
COSII	82 (29%)	50 (75%)	346	567	28332 (38%)	28380 (38%)	27.96%[b]
APVO SSC	217 (38%)	128 (79%)	429	728	93194 (53%)	93253 (53%)	28.56%[b]
PPR	101 (58%)	83 (74%)	1194	1453	120635 (67%)	120726 (67%)	35.17%[a]
TOTAL	1200 (41%)	772 (74%)	425	661	510579 (51%)	510962 (51%)	34.20%

FIGURES

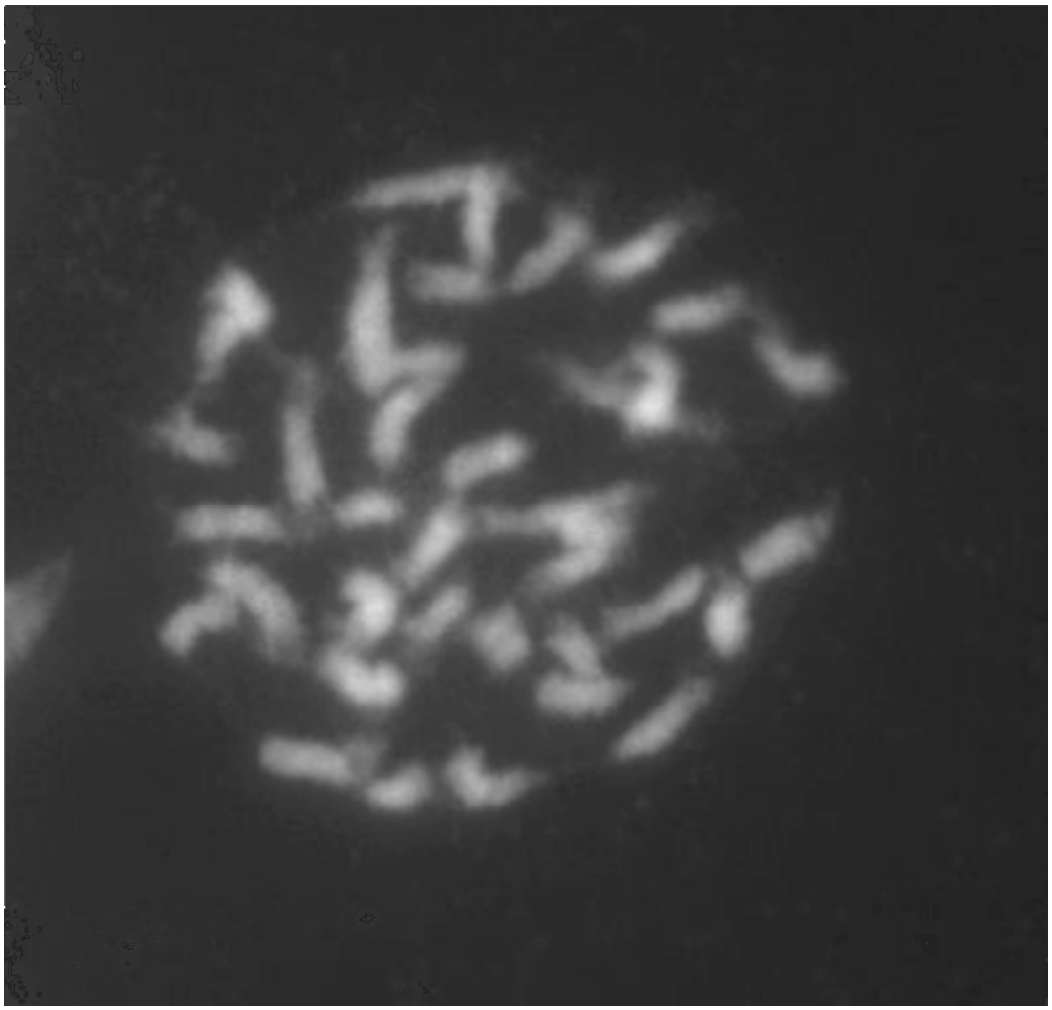
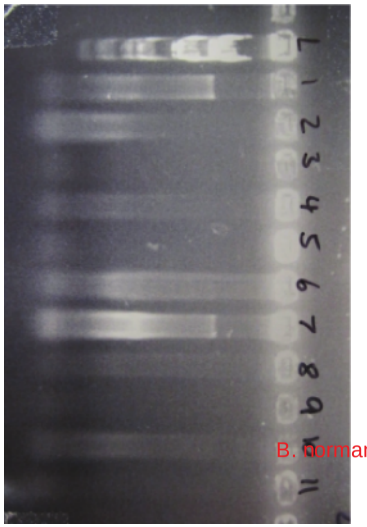
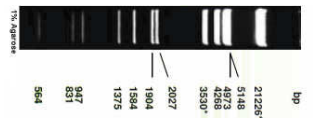


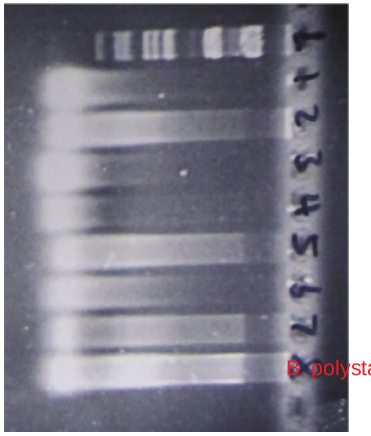
Figure 1. Micrograph of fixed cell used for chromosome counting. *Buddleja globosa* (R.G. Olmstead 2010-46 [WTU]) was confirmed to be diploid (2n=38).

Figure 2. Extracted DNA (2.5−5 μL) from select samples and DNA mass ladder run on 1% agarose gels, showing size distribution of DNA. Green labels indicate samples from silica gel-preserved tissue. Red labels indicate samples from herbarium specimen tissue.

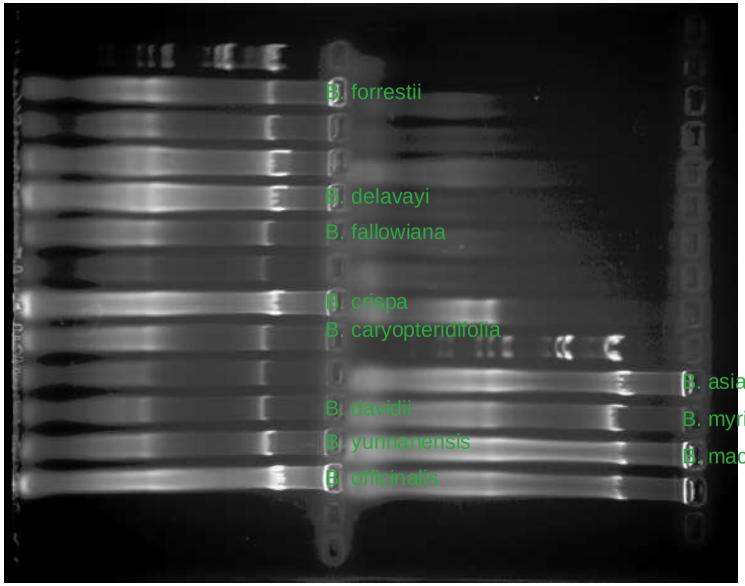


B. normaniae





B. polystachya



B. forrestii

B. delavayi

B. fallowiana

B. crispa

B. caryopteridifolia

B. asiatica

B. davidii

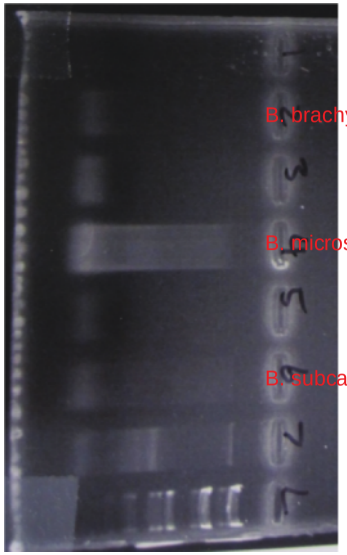
B. myriantha

B. yunnanensis

B. macrostachya

B. officinalis

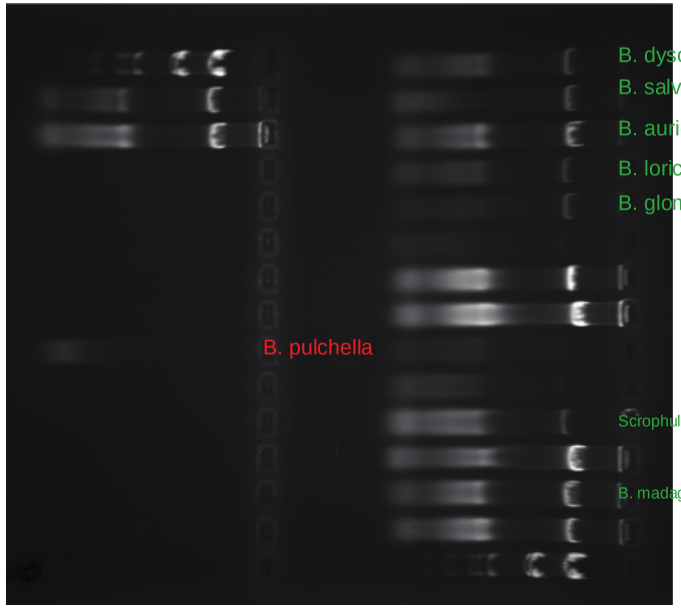
B. rinconensis



B. brachystachya

B. microstachya

B. subcapitata



B. dysophylla

B. salviifolia

B. auriculata

B. loricata

B. glomerata

B. pulchella

Scrophularia ndosoa

B. madagascariensis

Figure 3. Phylogenetic trees from analyses of concatenated sequences from all locus sets: a) maximum likelihood phylogram from RAxML analysis, and b) species tree from SVD quartets analysis. Values at nodes indicate bootstrap support from analyses with different locus sets: all/"taxon-specific"/COSII/APVO SSC/PPR. Dashed lines highlight taxa with incongruent relationships between trees.

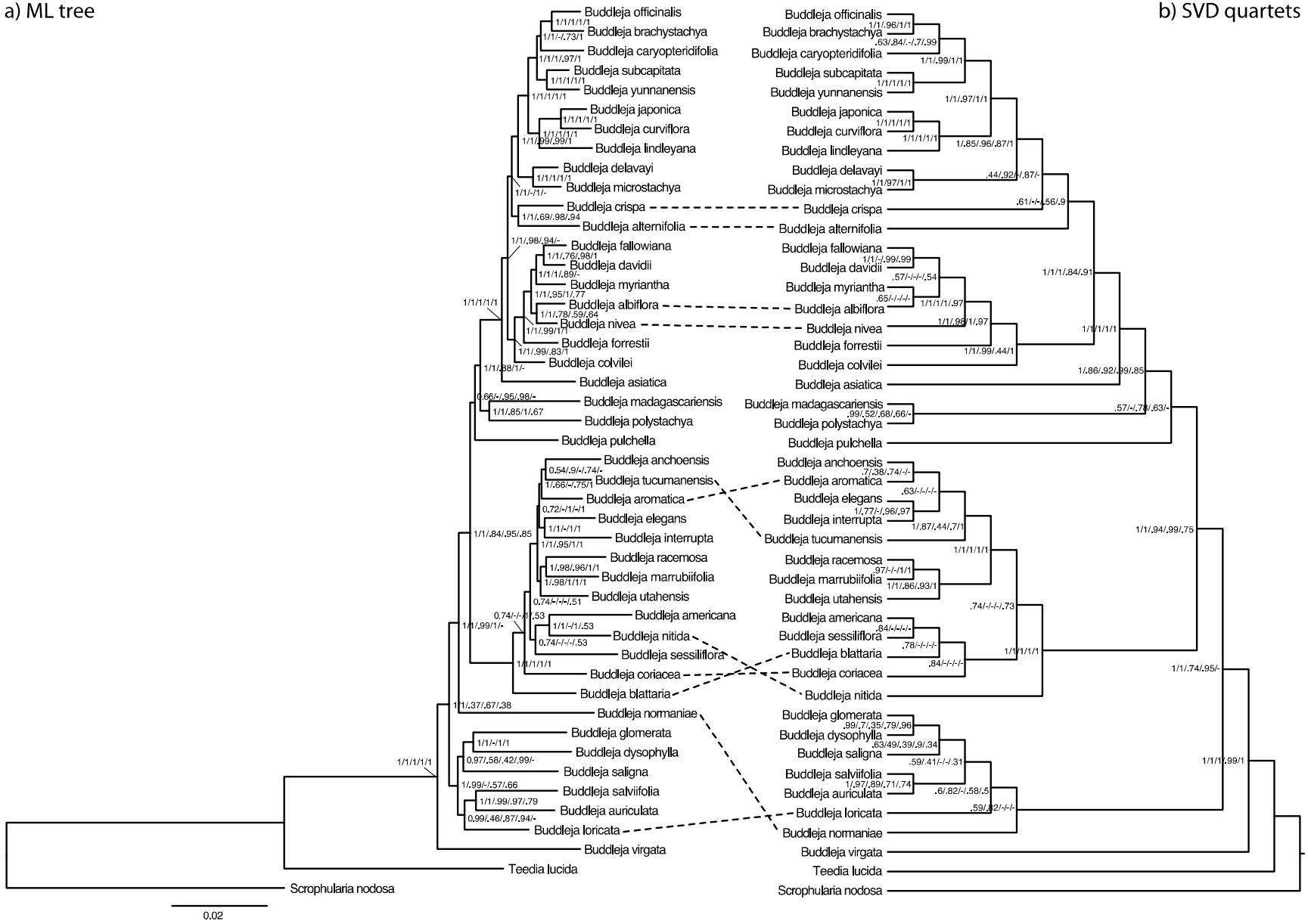
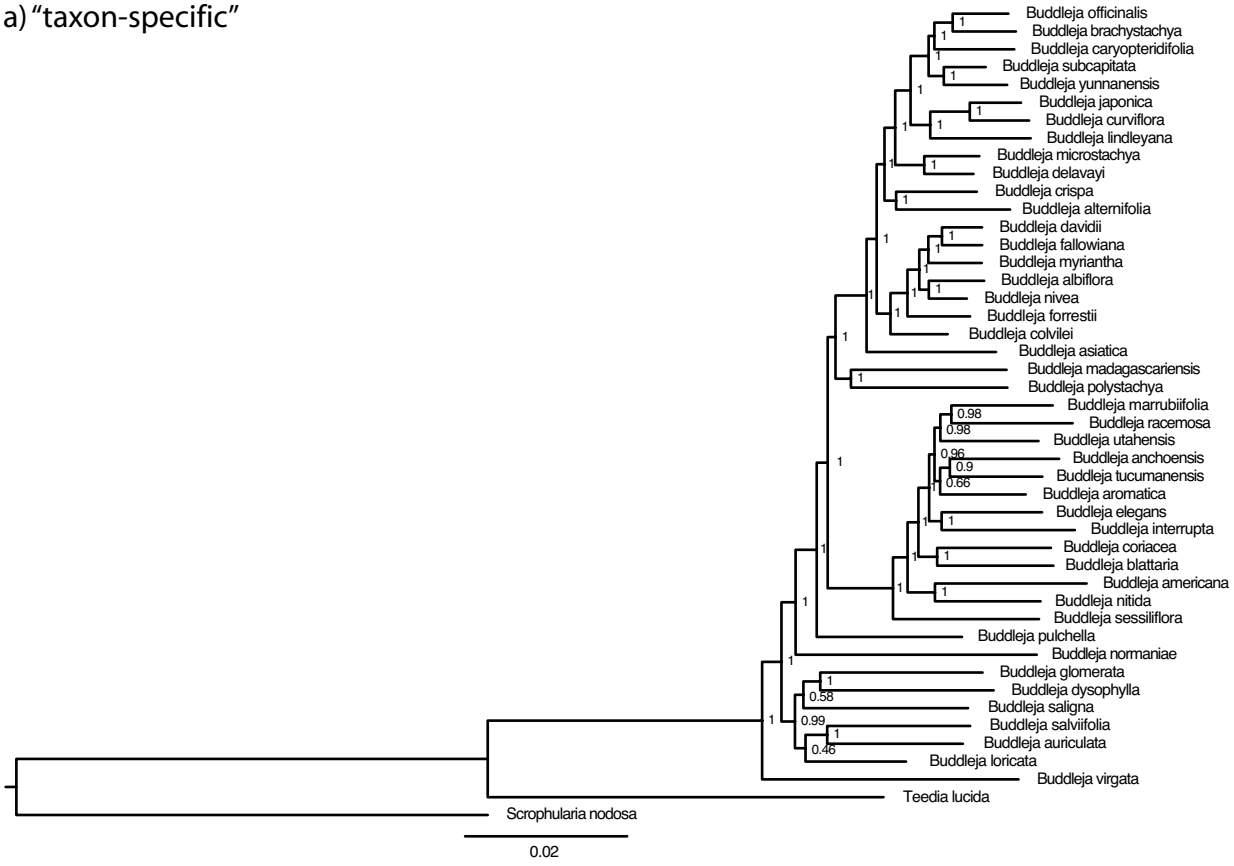
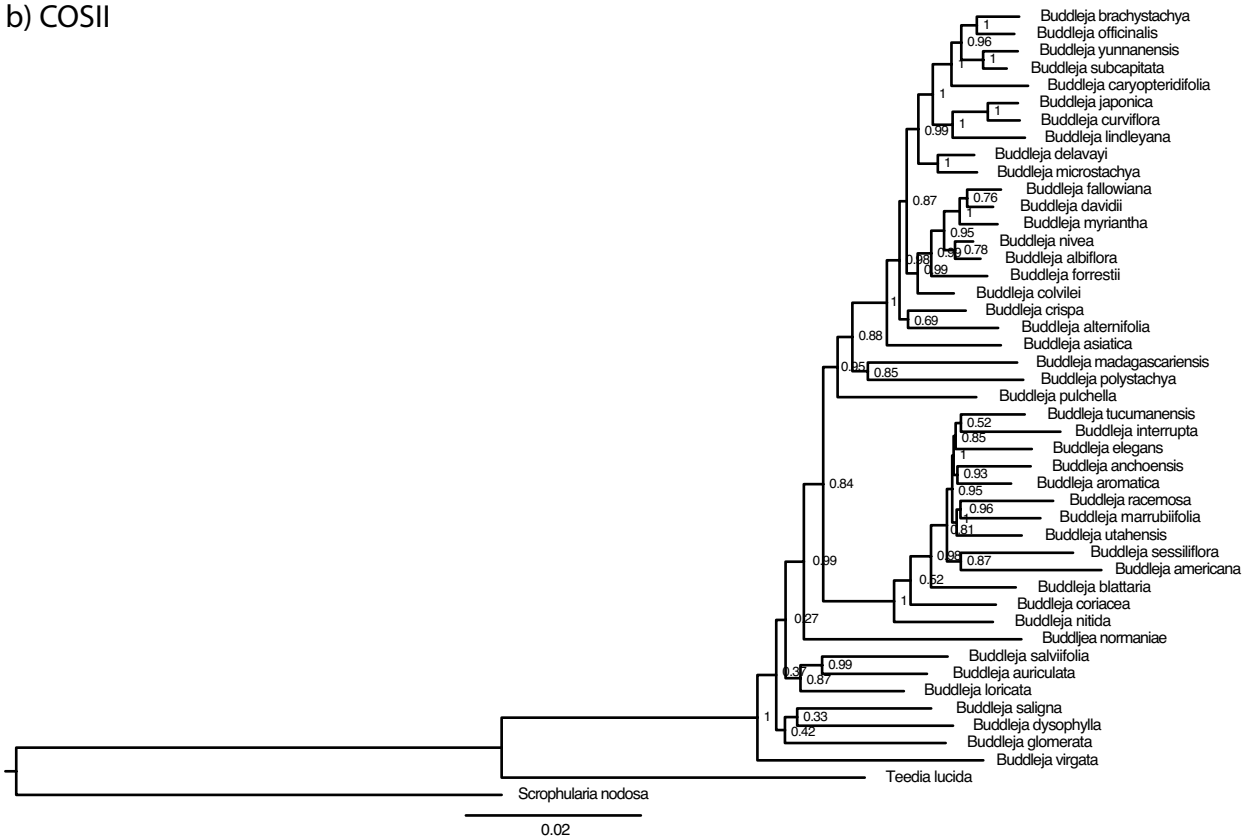

a) ML tree
b) SVD quartets
Buddleja officinalis
Buddleja brachystachya
Buddleja caryopteridifolia
Buddleja subcapitata
Buddleja yunnanensis
Buddleja japonica
Buddleja curviflora
Buddleja lindleyana
Buddleja delavayi
Buddleja microstachya
Buddleja crispa
Buddleja alternifolia
Buddleja fallowiana
Buddleja davidii
Buddleja myriantha
Buddleja albiflora
Buddleja nivea
Buddleja forrestii
Buddleja colvilei
Buddleja asiatica
Buddleja madagascariensis
Buddleja polystachya
Buddleja pulchella
Buddleja anchoensis
Buddleja tucumanensis
Buddleja aromatica
Buddleja elegans
Buddleja interrupta
Buddleja racemosa
Buddleja marrubiifolia
Buddleja utahensis
Buddleja americana
Buddleja nitida
Buddleja sessiliflora
Buddleja coriacea
Buddleja blattaria
Buddleja normaniae
Buddleja glomerata
Buddleja dysophylla
Buddleja saligna
Buddleja salviifolia
Buddleja auriculata
Buddleja loricata
Buddleja virgata
Teedia lucida
Scrophularia nodosa
0.02

Figure 4. Maximum likelihood phylograms from RAxML analyses with concatenated sequences from different locus sets: A) "taxon-specific", B) COSII, C) APVO SSC, and D) PPR. Values at nodes indicate bootstrap support.

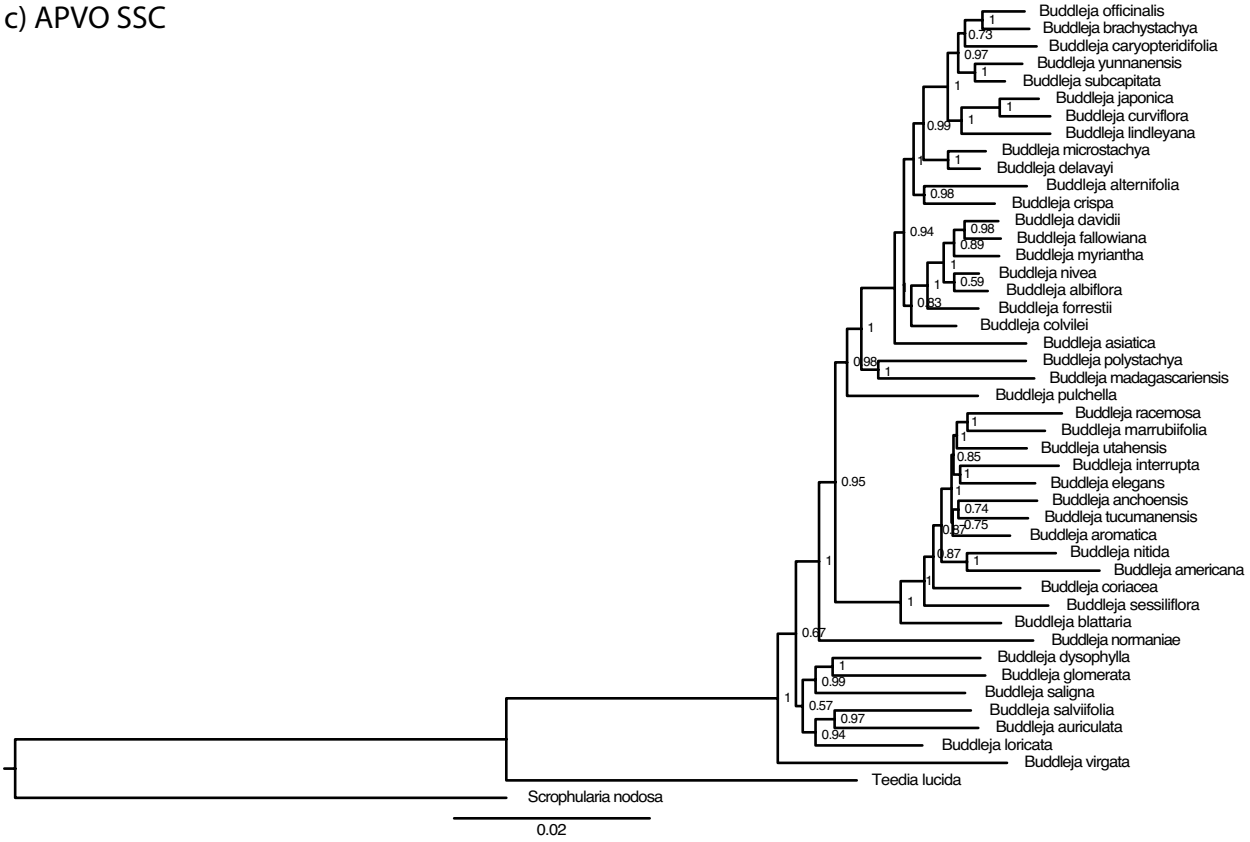
a) "taxon-specific"



b) COSII



c) APVO SSC



d) PPR

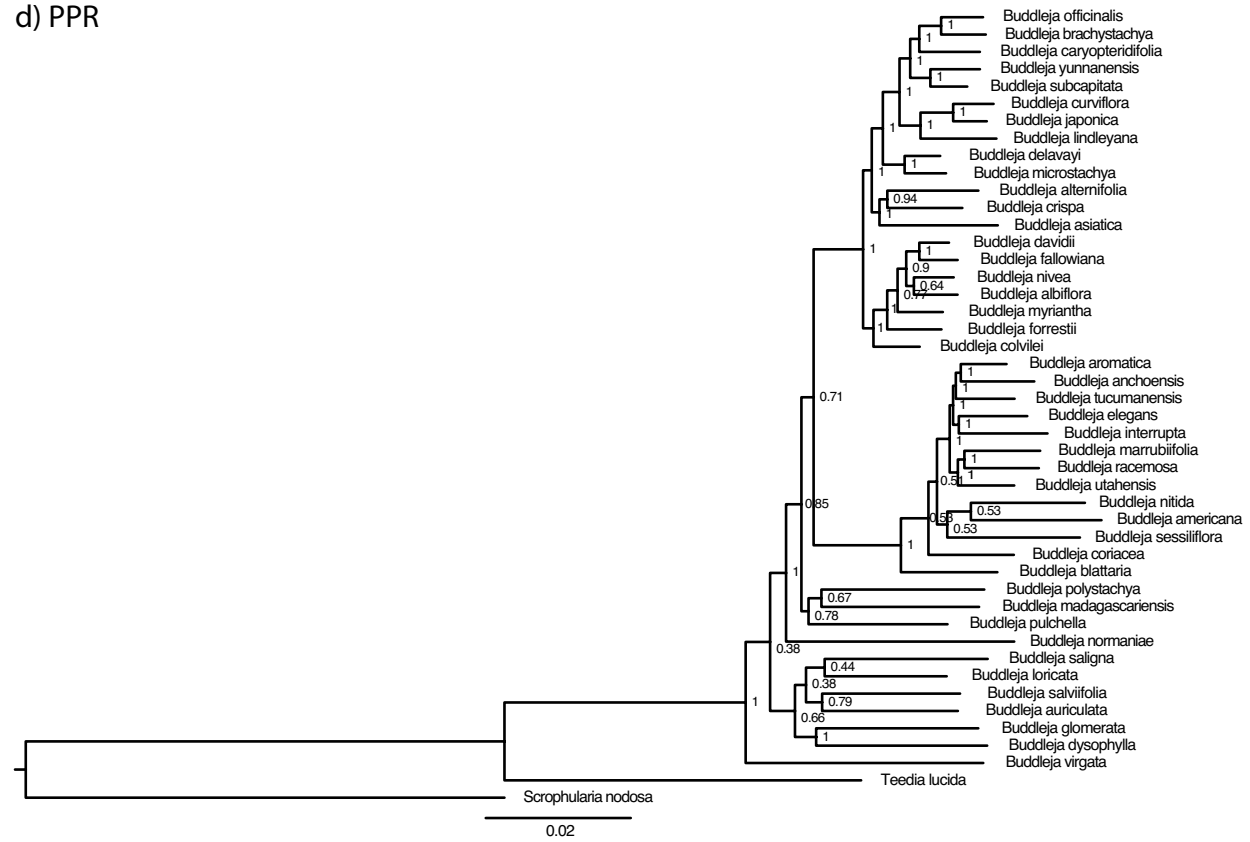
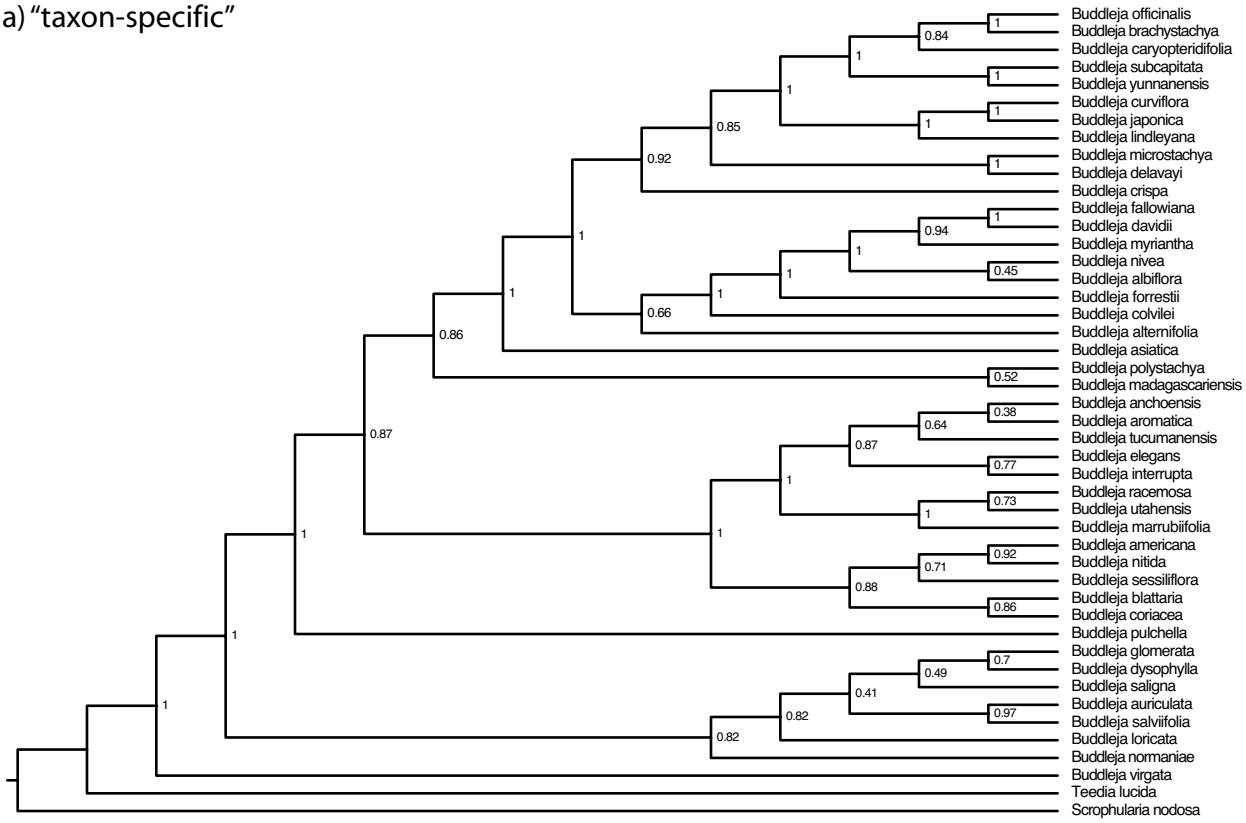
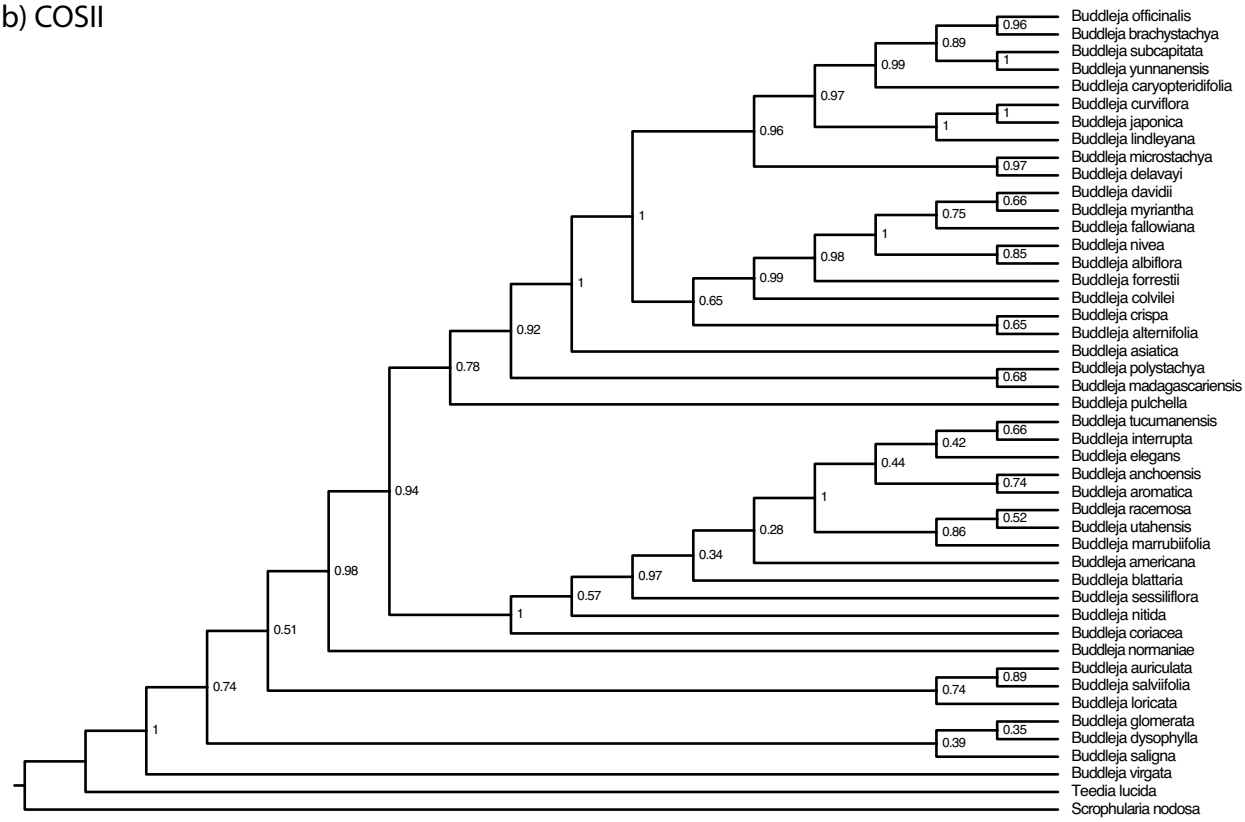


Figure 5. Trees from SVD quartets analyses with concatenated sequences from different locus sets: A) "taxon-specific", B) COSII, C) APVO SSC, and D) PPR. Values at nodes indicate bootstrap support.

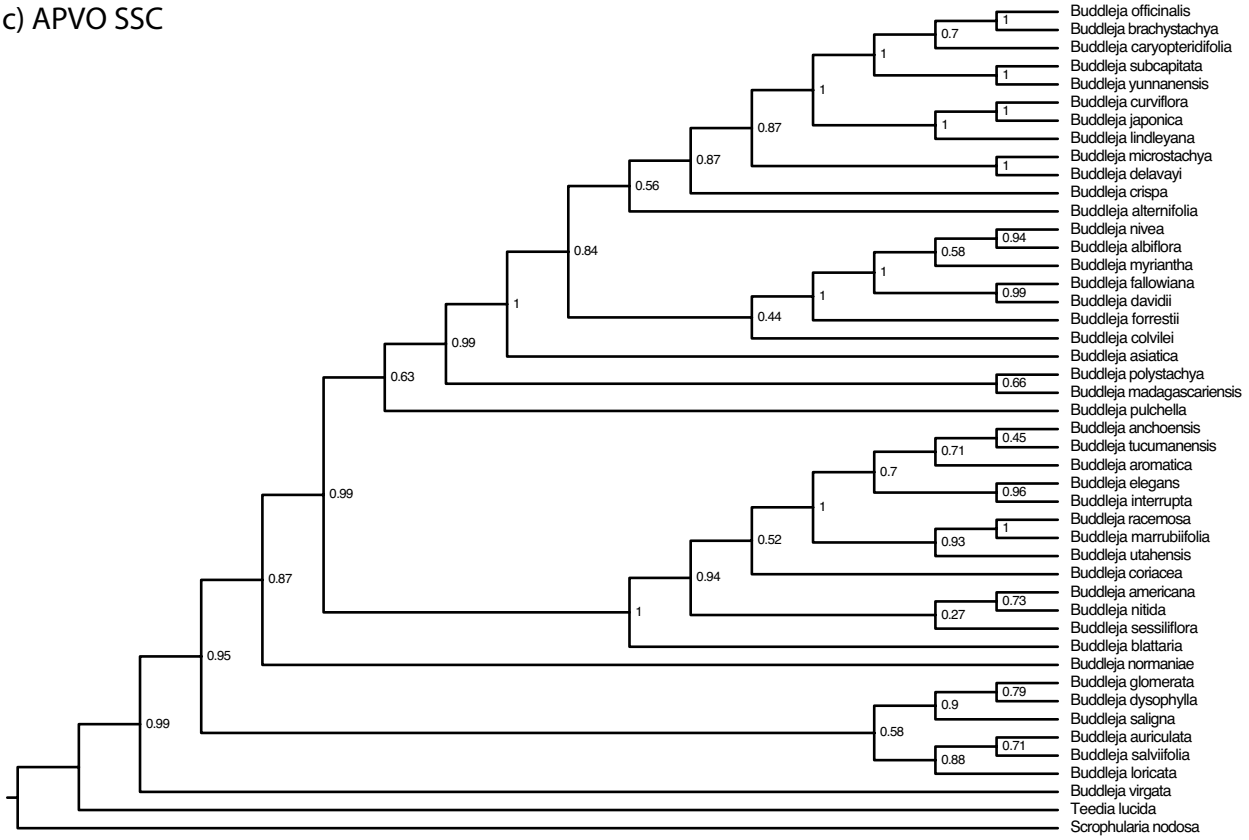
a) "taxon-specific"



b) COSII



c) APVO SSC



d) PPR

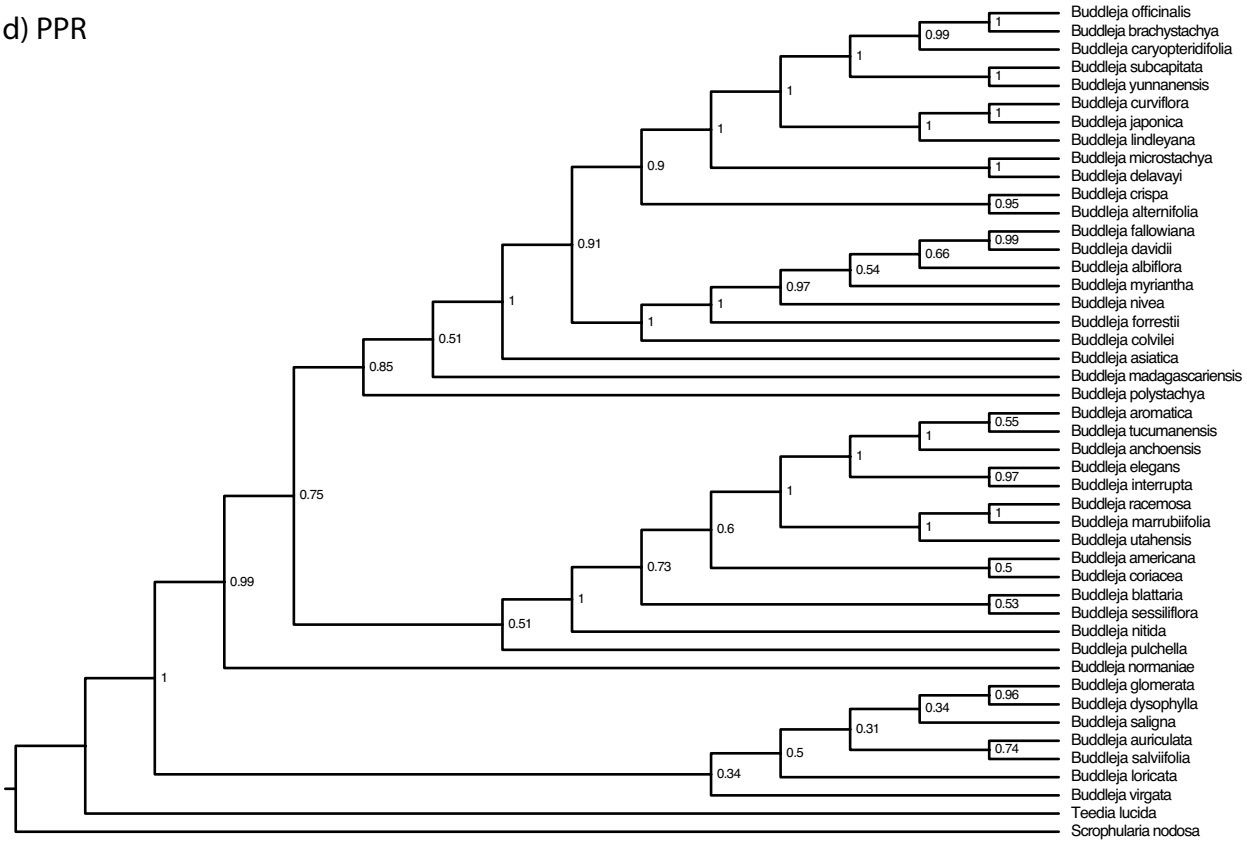
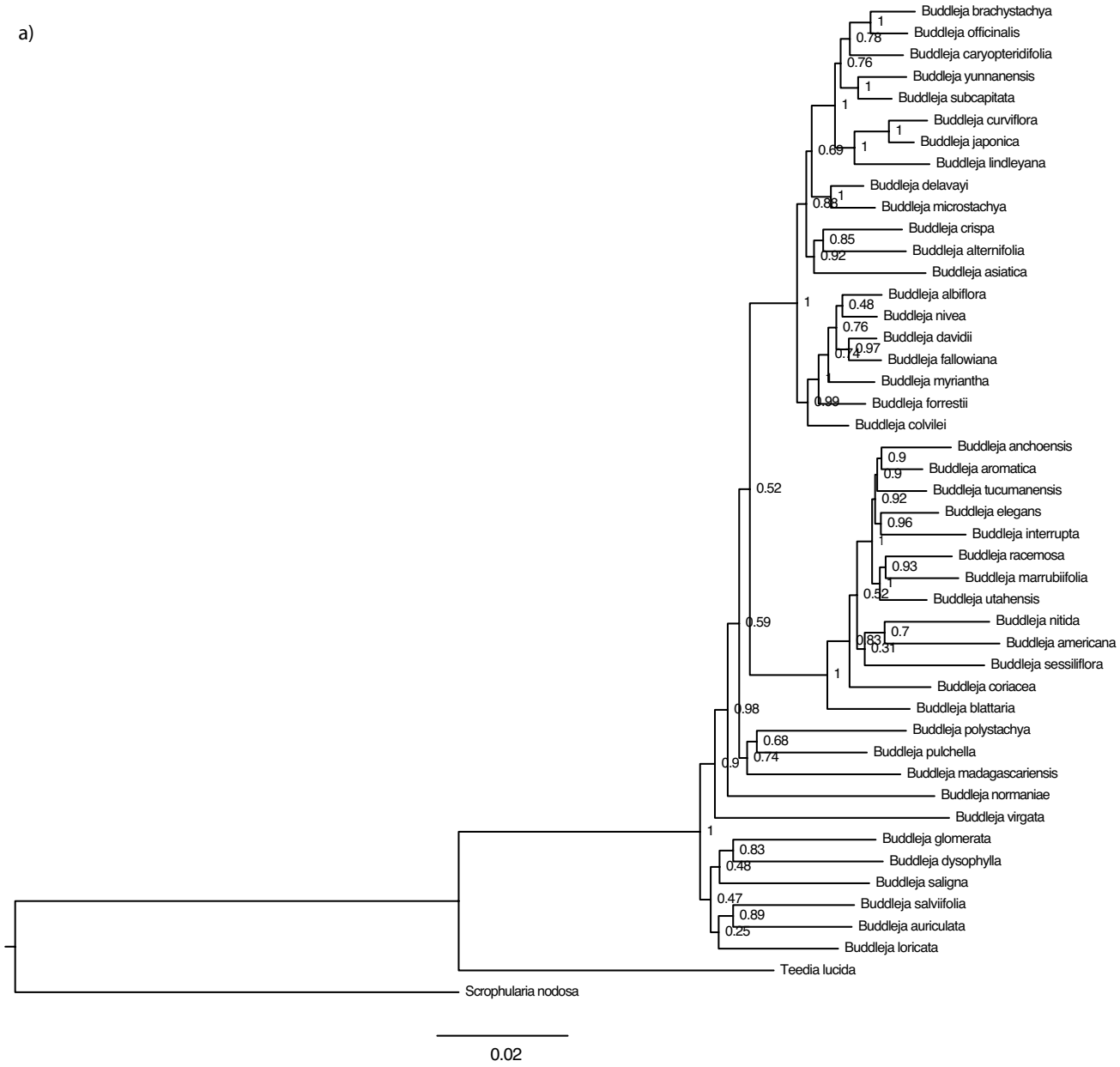
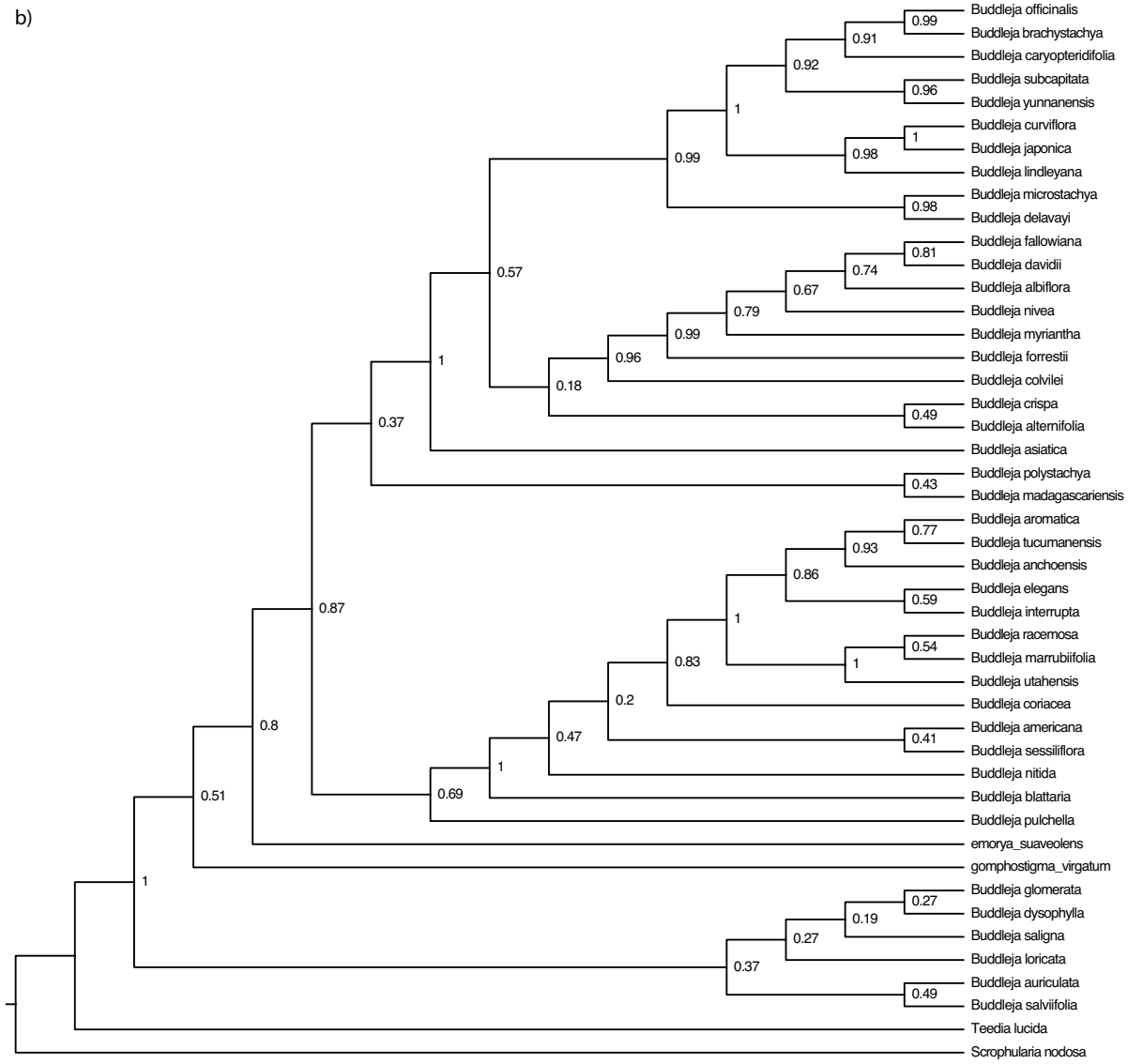


Figure 6. Phylogenetic trees from analyses of concatenated dataset from PPR locus set, reduced to length of COSII dataset: a) maximum likelihood phylogram from RAxML analysis, and b) species tree from SVD quartets analysis. Values at nodes indicate bootstrap support values.



b)


Buddleja officinalis
Buddleja brachystachya
Buddleja caryopteridifolia
Buddleja subcapitata
Buddleja yunnanensis
Buddleja curviflora
Buddleja japonica
Buddleja lindleyana
Buddleja microstachya
Buddleja delavayi
Buddleja fallowiana
Buddleja davidii
Buddleja albiflora
Buddleja nivea
Buddleja myriantha
Buddleja forrestii
Buddleja colvilei
Buddleja crispa
Buddleja alternifolia
Buddleja asiatica
Buddleja polystachya
Buddleja madagascariensis
Buddleja aromatica
Buddleja tucumanensis
Buddleja anchoensis
Buddleja elegans
Buddleja interrupta
Buddleja racemosa
Buddleja marrubiifolia
Buddleja utahensis
Buddleja coriacea
Buddleja americana
Buddleja sessiliflora
Buddleja nitida
Buddleja blattaria
Buddleja pulchella
emorya_suaveolens
gomphostigma_virgatum
Buddleja glomerata
Buddleja dysophylla
Buddleja saligna
Buddleja loricata
Buddleja auriculata
Buddleja salviifolia
Teedia lucida
Scrophularia nodosa
0.99
0.91
0.92
0.96
1
1
0.98
0.99
0.98
0.81
0.74
0.67
0.79
0.99
0.96
0.18
0.49
0.57
1
0.37
0.43
0.77
0.93
0.86
0.59
1
0.54
1
0.83
0.2
0.41
0.47
1
0.69
0.87
0.8
0.51
1
0.27
0.19
0.27
0.37
0.49

APPENDIX

Supplementary Table 1. Specimens included in study, with voucher information, infrageneric or familial classification, sample source, and ploidy level.

Species	Voucher	Section (Family, if outside *Buddleja*)	Herbarium sample?	Expected ploidy
Buddleja albiflora	J. Chau 260 (WTU, A)	*Alternifoliae*	no	hexaploid
Buddleja alternifolia	J. Chau 262 (WTU, A)	*Alternifoliae*	no	diploid
Buddleja americana	L. Frost 148 (WTU)	*Buddleja*	no	tetraploid
Buddleja anchoensis	J. Chau 224 (WTU, LPB)	*Buddleja*	no	diploid
Buddleja aromatica	J. Chau 206 (WTU, LPB)	*Buddleja*	no	diploid
Buddleja asiatica	J. Chau 157 (WTU)	*Alternifoliae*	no	diploid
Buddleja auriculata	J. Chau 246 (WTU)	*Chilianthus*	no	diploid
Buddleja blattaria	J. Chau 101 (WTU)	*Buddleja*	no	tetraploid?
Buddleja brachystachya	(KUN 22547)	*Alternifoliae*	yes	diploid
Buddleja caryopteridifolia	J. Chau 171 (WTU)	*Alternifoliae*	no	diploid
Buddleja colvilei	J. Chau 42 (WTU)	*Alternifoliae*	no	dodecaploid+
Buddleja coriacea	J. Chau 194 (WTU, LPB)	*Buddleja*	no	tetraploid
Buddleja crispa	J. Chau 170 (WTU)	*Alternifoliae*	no	diploid
Buddleja curviflora	R. Olmstead 2010-49 (WTU)	*Alternifoliae*	no	diploid
Buddleja davidii	J. Chau 177 (WTU)	*Alternifoliae*	no	tetraploid
Buddleja delavayi	J. Chau 165 (WTU)	*Alternifoliae*	no	hexaploid
Buddleja dysophylla	J. Chau 233 (WTU)	*Chilianthus*	no	diploid
Buddleja elegans	R. Olmstead 2010-214 (ICN)	*Buddleja*	no	diploid
Buddleja fallowiana	J. Chau 166 (WTU)	*Alternifoliae*	no	tetraploid
Buddleja forrestii	J. Chau 161 (WTU)	*Alternifoliae*	no	hexaploid
Buddleja glomerata	J. Chau 254 (WTU)	incertae sedis	no	diploid
Buddleja interrupta	J. Chau 123 (WTU)	*Buddleja*	no	diploid
Buddleja japonica	J. Wood 124-2014 (A)	*Alternifoliae*	no	diploid
Buddleja lindleyana	J. Wood & K. Richardson 125-2014 (A)	*Alternifoliae*	no	diploid
Buddleja loricata	J. Chau 253 (WTU)	*Chilianthus*	no	diploid
Buddleja macrostachya	J. Chau 159 (WTU)	*Alternifoliae*	no	hexaploid
Buddleja madagascariensis	J. Chau 256 (WTU)	*Nicodemia*	no	diploid
Buddleja marrubiifolia	M. Moore 1567 (WTU, MEXU)	*Buddleja*	no	diploid
Buddleja microstachya	E. Liu 925 (KUN)	*Alternifoliae*	yes	?
Buddleja myriantha	J. Chau 158 (WTU)	*Alternifoliae*	no	tetraploid
Buddleja nitida	J. Chau 150 (WTU)	*Buddleja*	no	tetraploid
Buddleja nivea	R. Olmstead 2010-47 (WTU)	*Alternifoliae*	no	hexaploid or dodecaploid
Buddleja normaniae	D. Riskind 23860 (TEX)	*Buddleja*	yes	diploid
Buddleja officinalis	J. Chau 179 (WTU)	*Alternifoliae*	no	diploid
Buddleja polystachya	G. Simon 308 (MO)	*Nicodemia*	yes	diploid
Buddleja pulchella	I. Nanni 319 (NBG)	*Pulchellae*	yes	diploid
Buddleja racemosa	J. Chau 324 (WTU)	*Buddleja*	no	diploid
Buddleja rinconensis	S. Aguilar Ruiz 164 (TEX)	*Buddleja*	yes	diploid
Buddleja saligna	J. Chau 231 (WTU)	*Chilianthus*	no	diploid
Buddleja salviifolia	J. Chau 240 (WTU)	*Salviifoliae*	no	diploid
Buddleja sessiliflora	G. Webster 31455 (DAV)	*Buddleja*	yes	tetraploid
Buddleja subcapitata	H. Peng 5153 (KUN)	*Alternifoliae*	yes	?
Buddleja tucumanensis	J. Chau 212 (WTU, LPB)	*Buddleja*	no	diploid
Buddleja utahensis	J. Chau 322 (WTU)	*Buddleja*	no	diploid
Buddleja virgata	J. Chau 180 (WTU)	*Gomphostigma*	no	diploid
Buddleja yunnanensis	J. Chau 178 (WTU)	*Alternifoliae*	no	diploid
Teedia lucida	J. Chau 318 (WTU)	Scophulariaceae	no	diploid
Scrophularia nodosa	J. Chau 228 (WTU)	Scophulariaceae	no	diploid
Parmentiera aculeata	S. Grose 93 (WTU)	Bignoniaceae	no	diploid
Lantana leonariorum	P. Lu-Irving 2012-105 (WTU)	Verbenaceae	no	diploid