

A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny

The Legume Phylogeny Working Group (LPWG)

Recommended citation: LPWG (2017)

This paper is a product of the Legume Phylogeny Working Group, who discussed, debated and agreed on the classification of the Leguminosae presented here, and are listed in alphabetical order. The text, keys and descriptions were written and compiled by a subset of authors indicated by §. Newly generated *matK* sequences were provided by a subset of authors indicated by *. All listed authors commented on and approved the final manuscript.

Nasim Azani,¹ Marielle Babineau,^{2*} C. Donovan Bailey,^{3*} Hannah Banks,⁴ Ariane R. Barbosa,^{5*} Rafael Barbosa Pinto,^{6*} James S. Boatwright,^{7*} Leonardo M. Borges,^{8*} Gillian K. Brown,^{9*} Anne Bruneau,^{2§*} Elisa Candido,^{6*} Domingos Cardoso,^{10§*} Kuo-Fang Chung,^{11*} Ruth P. Clark,⁴ Adilva de S. Conceição,^{12*} Michael Crisp,^{13*} Paloma Cubas,^{14*} Alfonso Delgado-Salinas,¹⁵ Kyle G. Dexter,^{16*} Jeff J. Doyle,¹⁷ Jérôme Duminiel,^{18*} Ashley N. Egan,^{19*} Manuel de la Estrella,^{4§*} Marcus J. Falcão,²⁰ Dmitry A. Filatov,^{21*} Ana Paula Fortuna-Perez,^{22*} Renée H. Fortunato,²³ Edeline Gagnon,^{2*} Peter Gasson,⁴ Juliana Gastaldello Rando,^{24*} Ana Maria Goulart de Azevedo Tozzi,⁶ Bee Gunn,^{13*} David Harris,²⁵ Elspeth Haston,²⁵ Julie A. Hawkins,^{26*} Patrick S. Herendeen,^{27§} Colin E. Hughes,^{28§*} João R.V. Iganci,^{29*} Firouzeh Javadi,^{30*} Sheku Alfred Kanu,³¹ Shahrokh Kazempour-Osaloo,^{32*} Geoffrey C. Kite,⁴ Bente B. Klitgaard,^{4§} Fábio J. Kochanowski,⁶ Erik J.M. Koenen,^{28§*} Lynsey Kovar,^{3*} Matt Lavin,^{33*} Marianne le Roux,³⁴ Gwilym P. Lewis,^{4§} Haroldo C. de Lima,²⁰ Maria Cristina López-Roberts,^{5*} Barbara Mackinder,^{25§*} Vitor Hugo Maia,^{35*} Valéry Malécot,³⁶ Vidal F. Mansano,^{20*} Brigitte Marazzi,^{37*} Sawai Mattapha,^{26*} Joseph T. Miller,³⁸ Chika Mitsuyuki,^{39*} Tania Moura,^{40*} Daniel J. Murphy,⁴¹ Madhugiri Nageswara-Rao,^{3*} Bruno Nevado,^{21*} Danilo Neves,^{4*} Dario I. Ojeda,¹⁸ R. Toby Pennington,^{25§*} Darién E. Prado,⁴² Gerhard Prenner,⁴ Luciano Paganucci de Queiroz,^{5§*} Gustavo Ramos,¹⁰ Fabiana L. Ranzato Filardi,^{20*} Pétala G. Ribeiro,⁵ Maria de Lourdes Rico-Arce,⁴ Michael J. Sanderson,⁴³ Juliana Santos-Silva,^{12*} Wallace M.B. São-Mateus,^{44*} Marcos J.S. Silva,^{45*} Marcelo F. Simon,^{46*} Carole Sinou,^{28§*} Cristiane Snak,^{5*} Élvia R. de Souza,^{12*} Janet Sprent,⁴⁷ Kelly P. Steele,^{48*} Julia E. Steier,^{49*} Royce Steeves,^{2*} Charles H. Stirton,⁵⁰ Shuichiro Tagane,^{39*} Benjamin M. Torke,^{51*} Hironori Toyama,^{39*} Daiane Trabuco da Cruz,^{5*} Mohammad Vatanparast,^{19*} Jan J. Wieringa,^{52§*} Michael Wink,^{53*} Martin F. Wojciechowski,^{49§*} Tetsukazu Yahara,^{39*} Tingshuang Yi⁵⁴ & Erin Zimmerman²

1 Department of Plant Science, University of Tehran, Iran

2 Institut de Recherche en Biologie Végétale and Département de Sciences Biologiques, Université de Montréal, Canada

3 Department of Biology, New Mexico State University, Las Cruces, U.S.A.

4 Royal Botanic Gardens, Kew, Richmond, Surrey, U.K.

5 Departamento Ciências Biológicas, Universidade Estadual de Feira de Santana, Brazil

6 Departamento de Biologia Vegetal, Universidade Estadual de Campinas, Brazil

7 Department of Biodiversity and Conservation Biology, University of the Western Cape, Cape Town, South Africa

8 Departamento de Botânica, Universidade Federal de São Carlos, Brazil

9 School of BioSciences, University of Melbourne, Australia and Queensland Herbarium, Toowong, Australia

10 Instituto de Biologia, Universidade Federal da Bahia, Salvador, Brazil

11 School of Forestry and Resource Conservation, National Taiwan University, Taiwan

12 Departamento de Educação, Campus VIII, Universidade do Estado da Bahia, Paulo Afonso, Brazil

13 Research School of Biology, The Australian National University, Acton, Australia

14 Departamento de Biología Vegetal II, Universidad Complutense, Madrid, Spain

15 Instituto de Biología – Botánica, Universidad Nacional Autónoma de México, México

16 School of GeoSciences, University of Edinburgh, U.K.

17 Plant Biology Department, Cornell University, Ithaca, New York, U.S.A.

18 Service Évolution Biologique et Écologie, Université Libre de Bruxelles, Belgium

19 Department of Botany, Smithsonian Institution, National Museum of Natural History, Washington D.C., U.S.A.

20 Instituto de Pesquisas, Jardim Botânico do Rio de Janeiro, Brazil

21 Department of Plant Sciences, University of Oxford, U.K.

Received: 6 Aug 2016 | returned for (first) revision: 2 Oct 2016 | (last) revision received: 19 Dec 2016 | accepted: 19 Dec 2016 || publication date(s): online fast track, n/a; in print and online issues, 23 Feb 2017 || Published online “open-access” under the terms of the Creative Commons Attribution 4.0 (CC BY 4.0) License || © International Association for Plant Taxonomy (IAPT) 2017

- 22 Departamento de Botânica, Universidade Estadual Paulista, Botucatu, São Paulo, Brazil
- 23 Instituto de Recursos Biológicos, CIRN-INTA, CONICET & University of Morón, Buenos Aires, Argentina
- 24 Ciências Ambientais, Universidade Federal do Oeste da Bahia, Barreiras, Brazil
- 25 Royal Botanic Gardens, Edinburgh, U.K.
- 26 School of Biological Sciences, University of Reading, U.K.
- 27 Chicago Botanic Garden, Glencoe, Illinois, U.S.A.
- 28 Department of Systematic and Evolutionary Botany, University of Zurich, Switzerland
- 29 Departamento de Botânica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
- 30 Institute of Decision Science for a Sustainable Society, Kyushu University, Fukuoka, Japan
- 31 Department of Agriculture and Animal Health, University of South Africa, South Africa
- 32 Department of Plant Biology, Tarbiat Modares University, Tehran, Iran
- 33 Plant Sciences & Plant Pathology, Montana State University, Bozeman, Montana, U.S.A.
- 34 South African National Biodiversity Institute, Silvertown, South Africa and Department of Botany and Plant Biotechnology, University of Johannesburg, South Africa
- 35 Departamento de Biologia, Pontifícia Universidade Católica do Rio de Janeiro, Brazil
- 36 Agrocampus-Ouest, INRA, Université d'Angers, France
- 37 Museo Cantonale di Storia Naturale, Lugano, Switzerland
- 38 Office of International Science and Engineering, National Science Foundation, Arlington, Virginia, U.S.A.
- 39 Department of Biology, Kyushu University, Fukuoka, Japan
- 40 Missouri Botanical Garden, Saint Louis, Missouri, U.S.A.
- 41 Plant Sciences and Biodiversity, Royal Botanic Gardens Victoria, Melbourne, Australia
- 42 Botânica, IICAR-CONICET, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Zavalla, Argentina
- 43 Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona, U.S.A.
- 44 Sistemática e Evolução, Universidade Federal do Rio Grande do Norte, Brazil
- 45 Departamento de Botânica, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Brazil
- 46 Embrapa Recursos Genéticos e Biotecnologia, Brasília, Brazil
- 47 Department of Plant Sciences, University of Dundee, U.K.
- 48 Faculty of Science and Mathematics, Arizona State University Polytechnic, Mesa, Arizona, U.S.A.
- 49 School of Life Sciences, Arizona State University, Tempe, Arizona, U.S.A.
- 50 Department of Biological Sciences, Bolus Herbarium, University of Cape Town, Rondebosch, South Africa
- 51 Institute of Systematic Botany, The New York Botanical Garden, Bronx, New York, U.S.A.
- 52 Naturalis Biodiversity Center, Leiden, The Netherlands
- 53 Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, Germany
- 54 Kunming Institute of Botany, Chinese Academy of Sciences, China
- Author for correspondence: Anne Bruneau, anne.bruneau@umontreal.ca

DOI <https://doi.org/10.12705/661.3>

Abstract The classification of the legume family proposed here addresses the long-known non-monophyly of the traditionally recognised subfamily Caesalpinioideae, by recognising six robustly supported monophyletic subfamilies. This new classification uses as its framework the most comprehensive phylogenetic analyses of legumes to date, based on plastid *matK* gene sequences, and including near-complete sampling of genera (698 of the currently recognised 765 genera) and ca. 20% (3696) of known species. The *matK* gene region has been the most widely sequenced across the legumes, and in most legume lineages, this gene region is sufficiently variable to yield well-supported clades. This analysis resolves the same major clades as in other phylogenies of whole plastid and nuclear gene sets (with much sparser taxon sampling). Our analysis improves upon previous studies that have used large phylogenies of the Leguminosae for addressing evolutionary questions, because it maximises generic sampling and provides a phylogenetic tree that is based on a fully curated set of sequences that are vouchered and taxonomically validated. The phylogenetic trees obtained and the underlying data are available to browse and download, facilitating subsequent analyses that require evolutionary trees. Here we propose a new community-endorsed classification of the family that reflects the phylogenetic structure that is consistently resolved and recognises six subfamilies in Leguminosae: a recircumscribed Caesalpinioideae DC., Cercidoideae Legume Phylogeny Working Group (stat. nov.), Detarioideae Burmeist., Dialioideae Legume Phylogeny Working Group (stat. nov.), Duparquetioideae Legume Phylogeny Working Group (stat. nov.), and Papilionoideae DC. The traditionally recognised subfamily Mimosoideae is a distinct clade nested within the recircumscribed Caesalpinioideae and is referred to informally as the mimosoid clade pending a forthcoming formal tribal and/or clade-based classification of the new Caesalpinioideae. We provide a key for subfamily identification, descriptions with diagnostic characteristics for the subfamilies, figures illustrating their floral and fruit diversity, and lists of genera by subfamily. This new classification of Leguminosae represents a consensus view of the international legume systematics community; it invokes both compromise and practicality of use.

Keywords Caesalpinioideae; Cercidoideae; Detarioideae; Dialioideae; Duparquetioideae; mimosoid clade; Papilionoideae; plastid *matK* phylogeny

Supplementary Material Electronic Supplement (Fig. S1), voucher information (Table S1), *matK* DNA sequence alignment (Data File A), phylogenetic tree files (Data Files B–F) and a poster illustrating the new classification of the Leguminosae (Figs. S2 & S3) are available in the Supplementary Data section of the online version of this article at <http://www.ingentaconnect.com/content/iapt/tax> and on Data Dryad (DOI: <https://doi.org/10.5061/dryad.61pd6>).

From arctic circle to tropics, desert to pergola, bacteria to plough, field to mouth, and legend to science, Leguminosae invest our lives, and a feeble backwash seeps through our universities. We wait that treatise which will quicken the herbarium into the living tree of phylogeny. Corner (1976: 162)

A general system of classification that is reasonably natural, mnemonic and traditional is likely to be most useful for most purposes. Polhill & al. (1981: 23–24)

Nutritious seeds for a sustainable future—The U.N. General Assembly declared 2016 the International Year of Pulses to raise awareness of the many benefits of legumes.

■ INTRODUCTION

The economically and ecologically important family Leguminosae (Lewis & al., 2005; Yahara & al., 2013), or Fabaceae (see Lewis & Schrire, 2003), has been the focus of numerous recent phylogenetic analyses at the subfamily, tribe and generic-group levels (see LPWG, 2013a and references therein). These, as well as phylogenies of the family as a whole (Käss & Wink, 1996; Doyle & al., 1997, 2000; Kajita & al., 2001; Wojciechowski & al., 2004; Lavin & al., 2005; McMahon & Sanderson, 2006; Bruneau & al., 2008; Simon & al., 2009; Cardoso & al., 2013b; LPWG, 2013a), all indicate that the currently accepted classification of the family into the three well-known, long-recognised and widely accepted subfamilies, Caesalpinioideae DC., Mimosoideae DC., and Papilionoideae DC., is outdated and does not reflect our current knowledge of phylogenetic relationships in the family.

With close to 770 genera and over 19,500 species (Lewis & al., 2005, 2013; LPWG, 2013a), the Leguminosae is the third-largest angiosperm family in terms of species numbers after Asteraceae and Orchidaceae. Economically, Leguminosae is second in importance only to Poaceae. It is estimated, for example, that total world exports of pulses (i.e., legume crops harvested for their dry seeds) have more than doubled between 1990 and 2012, expanding from 6.6 to 13.4 million tons, and in 2012 the value of pulse exports was estimated at US\$ 9.5 billion (Food and Agriculture Organisation [FAO]: <http://www.fao.org/pulses-2016/en/>). The United Nations General Assembly designated 2016 the International Year of Pulses to promote awareness of their nutritional benefits, importance in food security and sustainable agriculture, and in mitigating biodiversity loss and climate change. Legumes are important food crops providing highly nutritious sources of protein and micronutrients that can greatly benefit health and livelihoods,

particularly in developing countries (Yahara & al., 2013). Legumes have been domesticated alongside grasses in different areas of the world since the beginnings of agriculture and have played a key role in its early development (Gepts & al., 2005; Hancock, 2012). Legumes are also uniquely important as fodder and green manure in both temperate and tropical regions, and are used for their wood, tannins, oils and resins, in the manufacture of varnishes, paints, dyes and medicines, and in the horticultural trade.

Legumes are cosmopolitan in distribution, representing important ecological constituents in almost all biomes across the globe and occur in even the most extreme habitats (Schrire & al., 2005a, b). They constitute significant elements in terms of both species diversity and abundance, in lowland wet tropical forests in Africa, South America, and Asia (Yahara & al., 2013), and they dominate dry forests and savannas throughout the tropics (DRYFLOR, 2016), and also occur in Mediterranean, desert and temperate regions, up to high latitudes and at high elevations. They can be large emergent tropical trees with buttresses, small ephemeral annual herbs, climbing annuals or perennials with tendrils, desert shrubs, geoxylic subshrubs, woody lianas and, less commonly, aquatics. Flower symmetry spans the full range from radially symmetric (actinomorphic) to bilaterally symmetric (zygomorphic) and asymmetric flowers, which are in turn adapted to a wide range of pollinators such as insects, birds and bats. The ability of the majority of legume species to fix atmospheric nitrogen in symbiosis with soil rhizobia is perhaps the best-known ecological characteristic of the family; however, not all legumes form associations with nitrogen-fixing bacteria. Overall, the family is morphologically, physiologically and ecologically exceptionally diverse, representing one of the most spectacular examples of evolutionary diversification in plants. All of these characteristics have led to a continued fascination with the biology, diversity and evolution of the family, the evolution of functional traits, and the ecology and biogeography of the family by legume biologists (e.g., Stirton & Zarucchi, 1989; Lavin & al., 2004; Schrire & al., 2005a, b; Sprent, 2007, 2009; Champagne & al., 2007; Simon & al., 2009; Bouchenak-Khelladi & al., 2010; Cannon & al., 2010, 2015; Pennington & al., 2010; Doyle, 2011; Simon & Pennington, 2012; Koenen & al., 2013; Oliveira-Filho & al., 2013; Moncrieff & al., 2014; Werner & al., 2014, 2015; Dugas & al., 2015; BFG, 2015).

Here we propose a new subfamilial classification of the family Leguminosae that takes into account the phylogenetic pattern that is consistently resolved in numerous recent studies. This new classification is proposed and endorsed by the legume systematics community as reflected in the use of the Legume Phylogeny Working Group (LPWG) as the authority for all

new names proposed. The Legume Phylogeny Working Group was established explicitly to develop and foster collaborative research towards a comprehensive phylogeny and classification for Leguminosae (LPWG, 2013a).

The new classification proposed here follows a traditional Linnaean approach but is compatible with and complementary to emerging clade-based classifications of individual legume subfamilies (Wojciechowski, 2013). Rank-free naming of clades within (and across) subfamilies is already well-established and increasingly prevalent in the legume literature (e.g., Dalbergioid clade, Lavin & al., 2001; inverted repeat [IR]-lacking clade, Wojciechowski & al., 2000; Umtiza clade, Herendeen & al., 2003; *Acacia* s.l. clade, Miller & al., 2014), and additional important clades will continue to be named even after a fully fledged and stable subfamily and tribal classification is established. As noted by Wojciechowski (2013), use of Linnaean names does not preclude a system that also defines and names clades and their overall relationships outside of the traditional Linnaean framework. Instead, the two are considered complementary and necessary for developing a stable, flexible and useful classification of legumes.

■ THE NEW SUBFAMILY CLASSIFICATION

The monophyly of the family Leguminosae is strongly supported in all molecular phylogenetic analyses, regardless of taxon or gene sampling (see LPWG, 2013a and references therein). Indeed, despite uncertainty over their closest relatives (cf. Dickison, 1981; APG III, 2009; Bello & al., 2009), the monophyly and distinctiveness of the Leguminosae have never been questioned in terms of morphology since the family was first established (Adanson, 1763; Jussieu, 1789; Polhill & Raven, 1981; Polhill, 1994; Lewis & al., 2005; Bello & al., 2012). The most conspicuous characteristic of the family is, with only a few exceptions, a single superior carpel with one locule, marginal placentation, and usually two to many ovules, in two alternating rows on a single placenta (Lewis & al., 2005). However, legume systematists have been aware for a long time of the discrepancy between the current subfamily classification and emerging phylogenetic results (Irwin, 1981; Käss & Wink, 1996; Doyle & al., 1997), most notably the long known paraphyly of subfamily Caesalpinioideae, as well as many other problematic issues, such as lack of monophyly of many tribes and subtribes. This means that the phylogenetic structure of the family is not directly reflected in the current classification (Lewis & al., 2005). Thus, legume biologists studying particular clades have invented and used informal clade names that are biologically meaningful and appropriate for their study questions. This has resulted in a proliferation of informally named clades that can be inconsistent, ad hoc, and sometimes contradictory across studies, and which can lead to nomenclatural confusion unless they are properly defined (LPWG, 2013a, b; Wojciechowski, 2013). This is important not just within the legume taxonomic community but also for the legume biology, genomics, and indeed the wider evolutionary biology community as a whole (e.g., Cannon & al., 2015).

In contrast to some other large angiosperm families where the subfamily rank is perhaps not as widely recognised or used outside the immediate taxonomic community (e.g., Poaceae, Grass Phylogeny Working Group, 2001, 2012; Asteraceae, Panero & Funk, 2002, Funk & al., 2009), in legumes, the subfamily has always been a widely used and central rank. The three currently recognised subfamilies have long been considered as distinct groups and have often been recognised at the family rank (e.g., Hutchinson, 1964; Cronquist, 1981). In 1825, in his *Prodromus*, Candolle subdivided the Leguminosae into four suborders (= subfamilies), naming for the first time the three present-day subfamilies in addition to a fourth “sub-order”, Swartzieae, now included in subfamily Papilionoideae. This system was elaborated upon by Bentham (1865), who recognised three major groups within Leguminosae and whose classification formed the basis for all subsequent classifications of the family over the following 140 years (from, e.g., Taubert, 1891, to Polhill, 1994, and Lewis & al., 2005). In his *Families of flowering plants* (1926) and *Genera of flowering plants* (1964), Hutchinson raised the three subfamilies to the family level, but grouped them in the order Leguminales, a system that has been followed in a number of Floras (e.g., Hutchinson & Dalziel, 1928; Görts-van Rijn, 1989; Orchard & Wilson, 1998–2001; Mori & al., 2002; see also Lewis & Schrire, 2003). In the first volume of *Advances in legume systematics* (Polhill & Raven, 1981), the three groups were recognised at the subfamily rank. Regardless of rank, these three groups have been used as a division for identifying and classifying genera and species in Floras and herbaria throughout the world since the 19th century. These groupings are taught in botany, floristics and taxonomy courses, and are consistently used by agronomists, horticulturalists, and ecologists throughout the world. As remarked by Polhill & al. (1981: 24), “the basic classification of the family has remained remarkably stable and sensible. Users of classifications provide a strong selective force [...]”. Indeed, although the generic membership of the three subfamilies has changed somewhat over the centuries, these iconic groupings have remained useful concepts for identifying this diverse group of plants. Our objective here is to retain the utility of these well-known groups as far as possible while at the same time proposing a new classification that correctly reflects the evolutionary relationships and emphasises the distinctive features of each of the subfamilies.

Despite tremendous progress in understanding phylogenetic relationships across the family (LPWG, 2013a), uncertainty remains regarding relationships amongst the six first branching lineages of legumes and within certain clades (Figs. 1 & 2) (Wojciechowski & al., 2004; Bruneau & al., 2008; LPWG, 2013a). For example, relationships among early-branching papilionoids (Cardoso & al., 2012a, 2013b), the large so-called Mimosoideae-Caesalpinieae-Cassieae clade, or MCC clade *sensu* Doyle (2011, 2012) (Bruneau & al., 2008; Manzanilla & Bruneau, 2012;), and the Ingeae-Acacia s.str. clade (Luckow & al., 2003; Simon & al., 2009) all lack resolution and support using conventional DNA sequence datasets (i.e., a few kilobases of plastid DNA sequence data). However, there is no uncertainty surrounding the paraphyly of subfamily

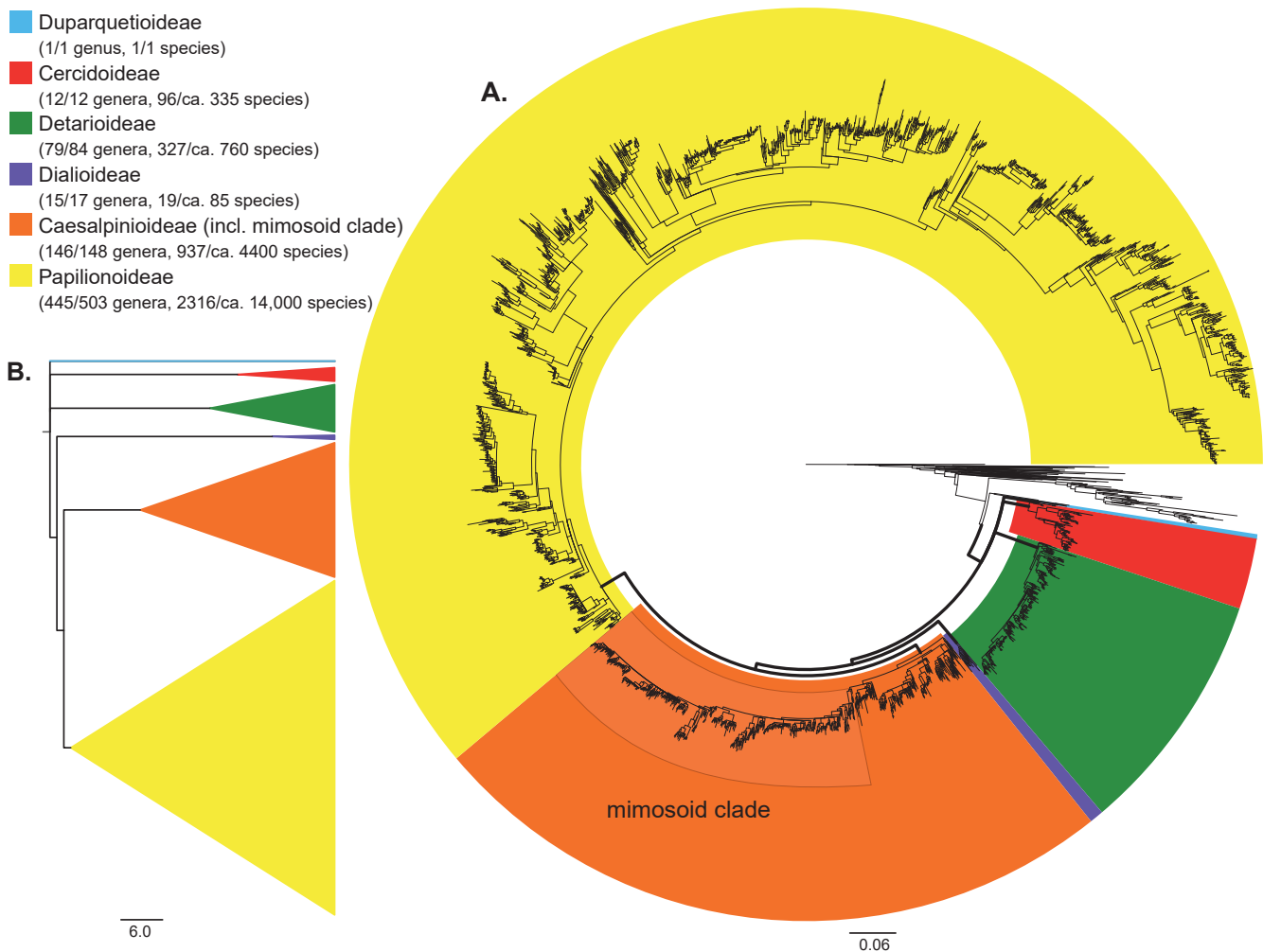
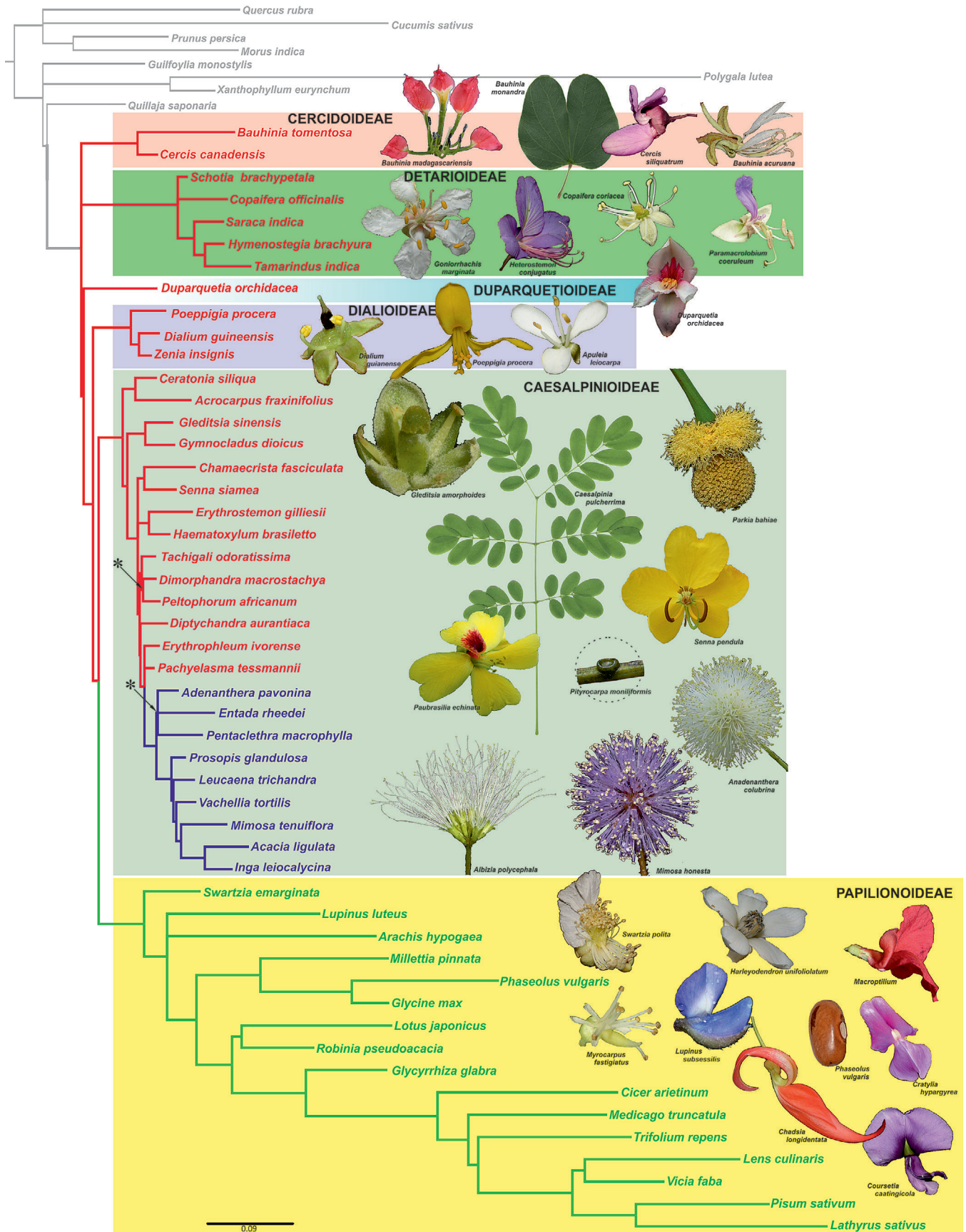


Fig. 1. A, Bayesian consensus phylogenetic tree of 3842 *matK* sequences representing 3696 of the ca. 19,500 species and 698 of the 765 genera (Table 2) of Leguminosae (for 30 species, multiple varieties or subspecies were included) and 100 outgroup taxa (uncoloured) spanning core Eudicots (see Appendix 1, Table S1). Branch lengths are proportional to numbers of *matK* substitutions. All subfamilies are supported with 1.0 posterior probability (indicated as thicker lines) and 100% maximum likelihood bootstrap values (Fig. S1). Support is weak across the backbone of the grade subtending the mimosoid clade, and this grade includes five or more lineages which would need to be recognised as additional small subfamilies if Mimosoideae had been retained at a subfamilial rank. Duparquetioideae forms a polytomy with Cercidoideae, Detarioideae and the clade that groups the other three subfamilies (but see Fig. 2, where Duparquetioideae is sister to the clade comprising Dialioideae, Caesalpinioideae and Papilionoideae based on analysis of a much larger plastid gene set). Numbers of genera and species (+intraspecific taxa) sampled/currently recognised are indicated for each subfamily. The phylogenetic tree can be visualised (e.g., with FigTree [<http://tree.bio.ed.ac.uk/software/figtree/>] or Dendroscope [<http://dendroscope.org/>]; Huson & Scornavacca, 2012), and downloaded from Supplementary Data: Data file B. **B**, Schematic phylogeny based on the *matK* Bayesian analysis showing the six subfamily classification of the Leguminosae, with clade sizes proportional to number of species. A schematic figure illustrating the diversity of the six subfamilies is available for download as a poster (Figs. S2, S3).

Fig. 2. Phylogeny and subfamily classification of the Leguminosae, depicted on a 95% majority-rule Bayesian consensus tree based on analysis of peptide sequences from 81 plastid encoded proteins, subsampling representative taxa from forthcoming phylogenomic analyses (E.J.M. Koenen & al., in prep.). This analysis resolves the relationships of Duparquetioideae (cf. Fig. 1 based on analysis of *matK* alone). The tree is unresolved in just a few places, including the root of the family and amongst clades in the Caesalpinioideae. All other nodes received 1.0 posterior probability, except the two nodes marked with an asterisk, which have 0.99 posterior probability. The tree was inferred using PhyloBayes v.1.6j (Lartillot & al., 2009) with the -CATGTR model selected and running two independent chains until they reached convergence. The six subfamilies are indicated by the coloured boxes to the right of the phylogeny. Coloured branches indicate the three traditionally recognised subfamilies of Leguminosae: red showing the paraphyletic old-sense Caesalpinioideae, blue the Mimosoideae and green the Papilionoideae.



Caesalpinioideae and hence the need for a new subfamilial classification (LPWG, 2013a, b). All adequately sampled phylogenetic analyses of the family indicate that the monophyletic Mimosoideae and Papilionoideae are nested within a paraphyletic assemblage of caesalpinoid lineages. This is perhaps no surprise. Already in 1981, in the preface to *Advances in legume systematics* volume 1, based on morphology alone, H.S. Irwin noted that Caesalpinioideae remained the most troublesome segment of the family and that, inevitably, a greater number of higher-level groups would need to be recognised.

The three traditional subfamilies were based essentially on a small set of conspicuous floral characters, particularly petal aestivation patterns (imbricate ascending in Caesalpinioideae vs. imbricate descending in Papilionoideae vs. valvate in Mimosoideae) and floral symmetry (variable in Caesalpinioideae [Figs. 3–5]; radially symmetric [i.e., actinomorphic] in Mimosoideae [Fig. 6]; bilaterally symmetric [i.e., zygomorphic] in Papilionoideae [Figs. 7–9]). While some of these floral characters may be useful for defining Papilionoideae and Mimosoideae, they are extremely variable across the traditional Caesalpinioideae (Tucker, 2003; Bruneau & al., 2014), which cannot be defined or diagnosed based on these characters. Furthermore, even for Papilionoideae and Mimosoideae, most of these floral traits are now known to be homoplasious (Pennington & al., 2000). For example, individual species or clades marked by radially symmetric flowers are independently derived multiple times across basal Papilionoideae, a large assemblage of florally heterogeneous lineages dominated by bilaterally symmetric flower morphology (Figs. 7–9) (Pennington & al., 2000; Cardoso & al., 2012b, 2013a; Ramos & al., 2016). Similarly, while Mimosoideae are the most conspicuously biodiverse clade with radially symmetric flowers, other closely related lineages scattered across the MCC clade also have radially symmetric, mimosoid-like flowers (Fig. 5). Thus, despite the central importance of floral characters in the traditional subfamilial classification, phylogenetic results over the past 20 years favour giving less weight to floral morphology because it is so prone to evolutionary modification and convergence, especially in the transition from radial to bilateral floral symmetry, which can be achieved in different ways.

There has been broad consensus about the need for a new classification within the legume systematics community since the first molecular phylogenies of the family were published (Käss & Wink, 1996; Doyle & al., 1997). However, the multi-lineage paraphyletic structure of subfamily Caesalpinioideae with respect to the monophyletic Mimosoideae and Papilionoideae poses significant questions about how many subfamilies should be recognised. Furthermore, until recently, incomplete sampling of many key genera in phylogenies suggested the need for caution before establishing a new subfamilial classification. More recent and densely sampled phylogenies (Luckow & al., 2003; Wojciechowski & al., 2004; Lavin & al., 2005; Bruneau & al., 2008; Simon & al., 2009; Cardoso & al., 2012a, 2013b), as well as the *matK* phylogeny with its near-complete sampling of genera that we present here (Figs. 1 & S1; Appendix 1), now provide adequate taxon sampling and phylogenetic support to

reveal in sufficient and definitive detail the overall phylogenetic structure of the family and allow us to properly evaluate the options and arrive at the best solution for translating the phylogenetic tree into a new classification. Furthermore, the main clades resolved in the *matK* phylogeny are also fully supported in whole plastid genome sequence analyses (Fig. 2) (E.J.M. Koenen & al., in prep.), and are corroborated by phylogenetic analyses of orthologous nuclear genes derived from representative sampling of multiple transcriptomes of all subfamilies, except Duparquetioideae (E.J.M. Koenen & al., in prep.).

A concerted effort to arrive at a new classification was initiated at the 6th International Legume Conference in Johannesburg, South Africa, in January 2013. Specifically, there was general consensus that sufficient data, in terms of taxon sampling and phylogenetic support, were available to propose a new subfamilial classification for Leguminosae, and there was universal agreement that the number of subfamilies needed to be increased (LPWG, 2013b). There was also broad agreement that several caesalpinoid clades (Cercideae, Detarieae, *Duparquetia*, Dialiineae s.l.) could be appropriately, uncontroversially and usefully recognised as new subfamilies, alongside Papilionoideae. The central problem for a new subfamilial classification, was how to deal with the large clade that includes the “Umtiza clade” or “grade”, “Caesalpinia Group clade”, “Cassia clade”, “Peltophorum clade”, scattered *Dimorphandra* Group genera, and which has Mimosoideae nested within it, i.e., the large MCC clade (sensu Doyle, 2011, 2012). Several participants suggested that the whole MCC clade should be recognized as a single subfamily (making a total of six subfamilies), but with the disadvantage that mimosoids, in the traditional sense, would no longer be recognised as a subfamily, which made some legume systematists uncomfortable. The alternative, whereby Mimosoideae is retained as a subfamily, entails recognition of six to eight (or more) additional small subfamilies to account for the multiple lineages that make up the large paraphyletic assemblage subtending mimosoids (Figs. 1, 2). However, many recognised that although resolution and support across this grade remains relatively weak in current phylogenies (Fig. 1; Bruneau & al., 2008; Manzanilla & Bruneau, 2012), improved resolution and support from larger datasets (e.g., Fig. 2; E.J.M. Koenen & al., in prep.) was not alone going to solve the problem of 6 vs. 11 or more subfamilies. These two main options for a new classification were summarised, the points of agreement noted, and the foundations for furthering the discussion presented in LPWG (2013b).

The advantages and disadvantages of these two main options for a new subfamily classification (6 vs. 11, or more subfamilies) were specifically discussed and evaluated at a subsequent legume systematics symposium, held during the Latin American Botanical Congress in October 2014, in Bahia, Brazil. A document was then drafted summarising the advantages and disadvantages and circulated to a LPWG electronic discussion group with wide, international membership for further discussion and opinion. The comments received from this draft were taken into account when developing the classification presented here, subfamily descriptions were discussed at a legume morphology workshop in Botucatu, Brazil (November



Fig. 3. A–F, Cercidoideae; G, Duparquetioideae; H–L, Dialioideae. **A**, *Cercis siliquastrum*; **B**, *Bauhinia galpinii*; **C**, *Bauhinia divaricata*; **D**, *Piliostigma thonningii*; **E**, *Griffonia physocarpa*; **F**, *Schnella cupreonitens*; **G**, *Duparquetia orchidacea*; **H**, *Zenia insignis*; **I**, *Apuleia leiocarpa*; **J**, *Poeppigia procera*; **K**, *Distemonanthus benthamianus*; **L**, *Kalappia celebica*. — Photos: A & B, Colin Hughes; C, Jonathan Amith; D, E & K, Xander van der Burgt; F & I, Domingos Cardoso; G, Paul Hoekstra; H, Shijin Li; J, Luciano P. de Queiroz; L, Liam Trethowan.

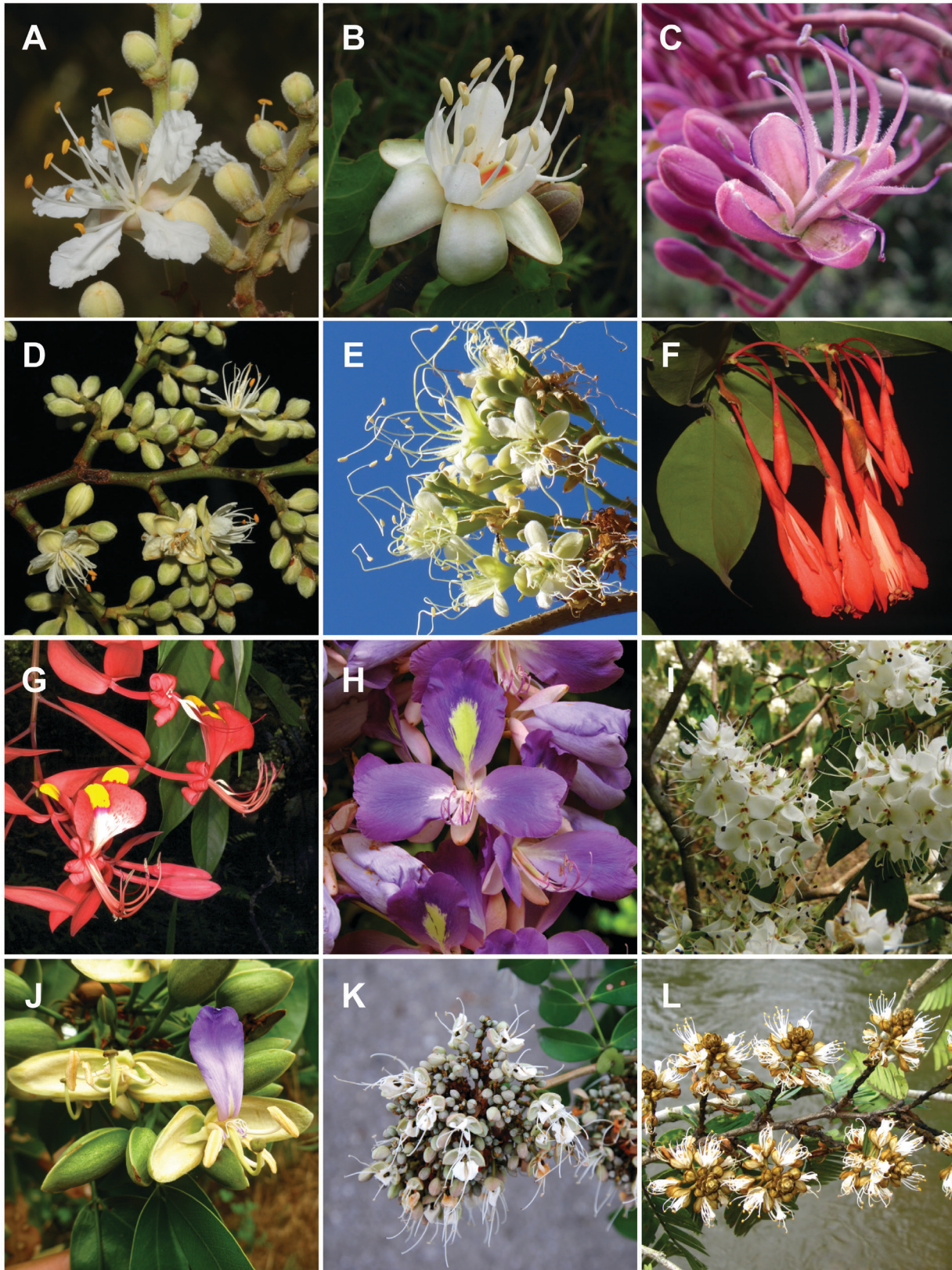


Fig. 4. Detarioideae. **A**, *Goniorrhachis marginata*; **B**, *Hymenaea stigonocarpa*; **C**, *Daniellia ogea*; **D**, *Peltogyne chrysoptis*; **E**, *Brodriguesia santosii*; **F**, *Brownea longipedicellata*; **G**, *Amherstia nobilis*; **H**, *Brachycylix vageleri*; **I**, *Cryptosepalum tetraphyllum*; **J**, *Paramacrolobium coeruleum*; **K**, *Gilbertiodendron quinquejugum*; **L**, *Aphanocalyx pteridophyllus*. — Photos: A, D & F, Domingos Cardoso; B, Luciano P. de Queiroz; C, I, J & L, Xander van der Burgt; E, Gwilym Lewis; G, Timothy Utteridge; H, Emilio Constantino; K, Jan Wieringa.

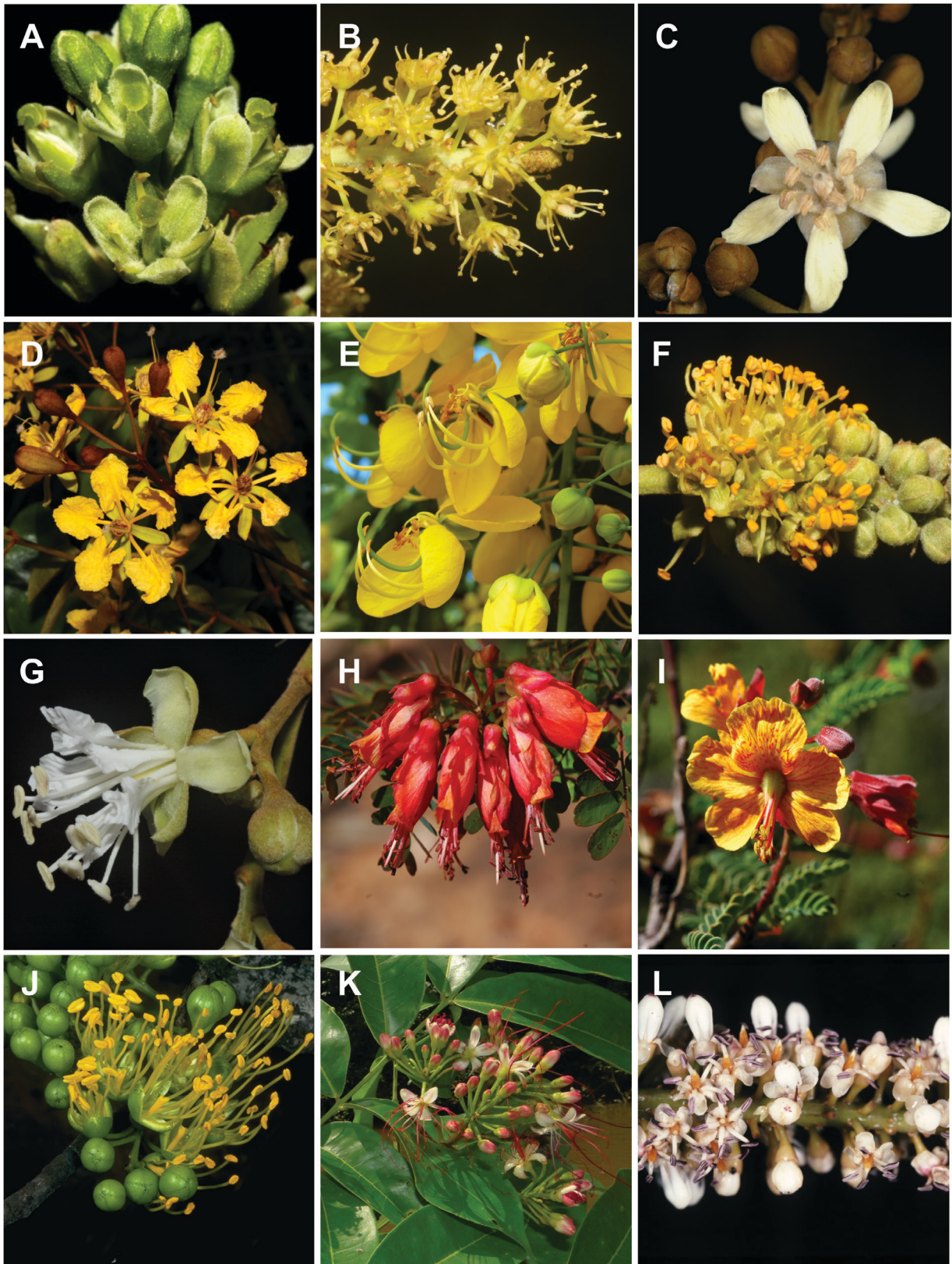


Fig. 5. Caesalpinioideae I. **A**, *Gleditsia amorphoides*; **B**, *Pterogyne nitens*; **C**, *Batesia floribunda*; **D**, *Moldenhawera blanchetiana*; **E**, *Cassia fistula*; **F**, *Tachigali rugosa*; **G**, *Arapatiella psilophylla*; **H**, *Caesalpinia cassioides*; **I**, *Arquita grandiflora*; **J**, *Delonix floribunda*; **K**, *Campsiandra comosa*; **L**, *Dimorphandra pennigera*. — Photos: A, B, D, F & G, Domingos Cardoso; C & L, Proyecto Flora Reserva Ducke, INPA/DFID, comm. Mike Hopkins; E, Gwilym Lewis; H & I, Colin Hughes; J, David Du Puy; K, Luciano P. de Queiroz.

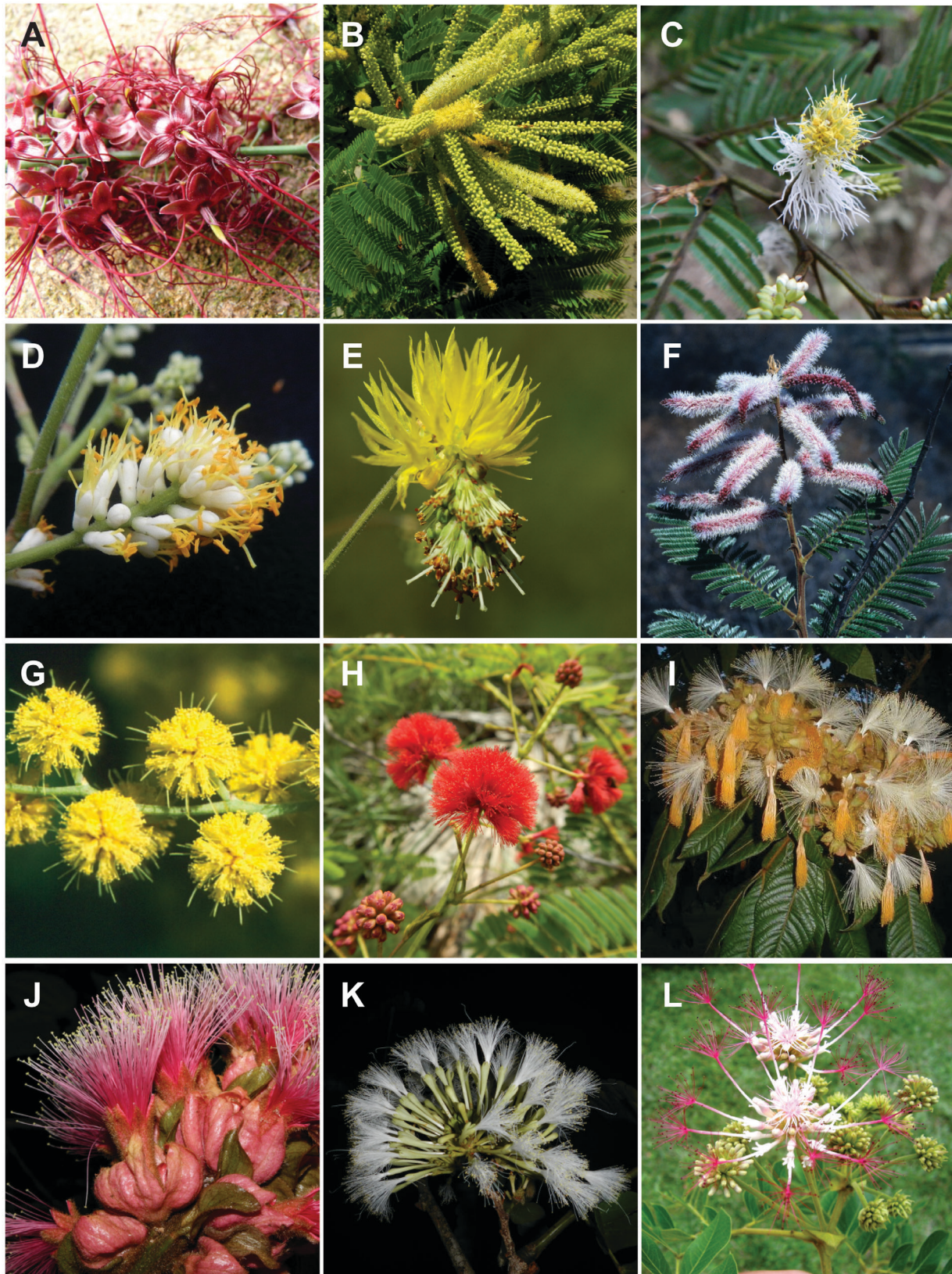


Fig. 6. Caesalpinioideae II. **A**, *Chidlowia sanguinea*; **B**, *Entada chrysostachys*; **C**, *Gagnebina commersoniana*; **D**, *Lemurodendron capuronii*; **E**, *Neptunia plena*; **F**, *Mimosa benthamii*; **G**, *Acacia dealbata*; **H**, *Senegalia sakalava*; **I**, *Inga calantha*; **J**, *Inga grazielae*; **K**, *Macrosamanea amplissima*; **L**, *Albizia grandibracteata*. — Photos: A, Xander van der Burgt; B–D, H, K & L, Erik Koenen; E–G, Colin Hughes; I, Flora do Acre, comm. Rosangela Melo; J, Domingos Cardoso.

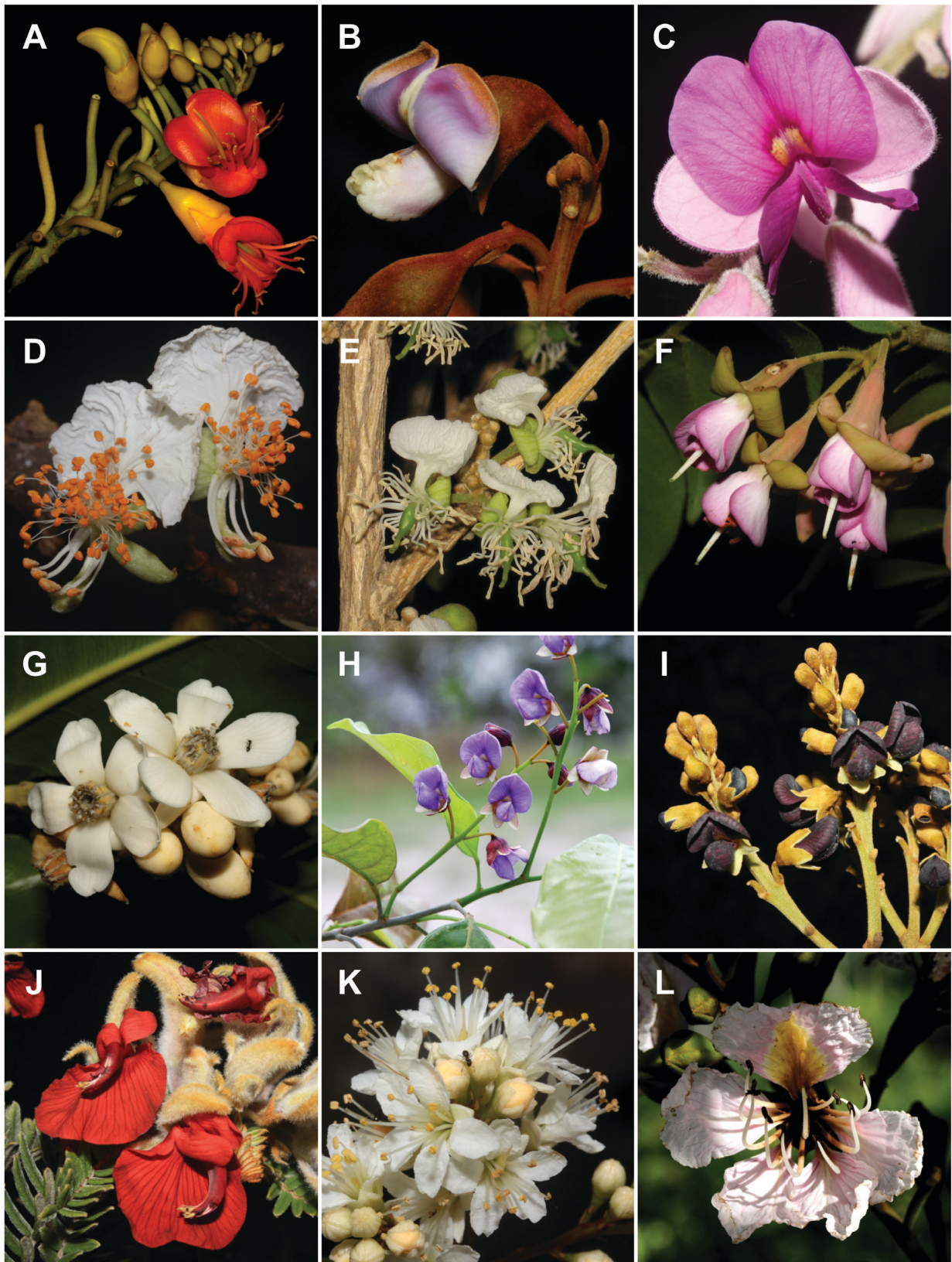


Fig. 7. Papilionoideae I. **A**, *Castanospermum australe*; **B**, *Petaladenium urceoliferum*; **C**, *Pterodon abruptus*; **D**, *Swartzia acutifolia*; **E**, *Trischidium molle*; **F**, *Exostyles venusta*; **G**, *Harleyodendron unifoliolatum*; **H**, *Haplormosia monophylla*; **I**, *Ormosia lewisii*; **J**, *Harpalyce lanata*; **K**, *Leptolobium brachystachyum*; **L**, *Camoensia brevicalyx*. — Photos: A–G & I–K, Domingos Cardoso; H, Jan Wieringa; L, André van Proosdij.

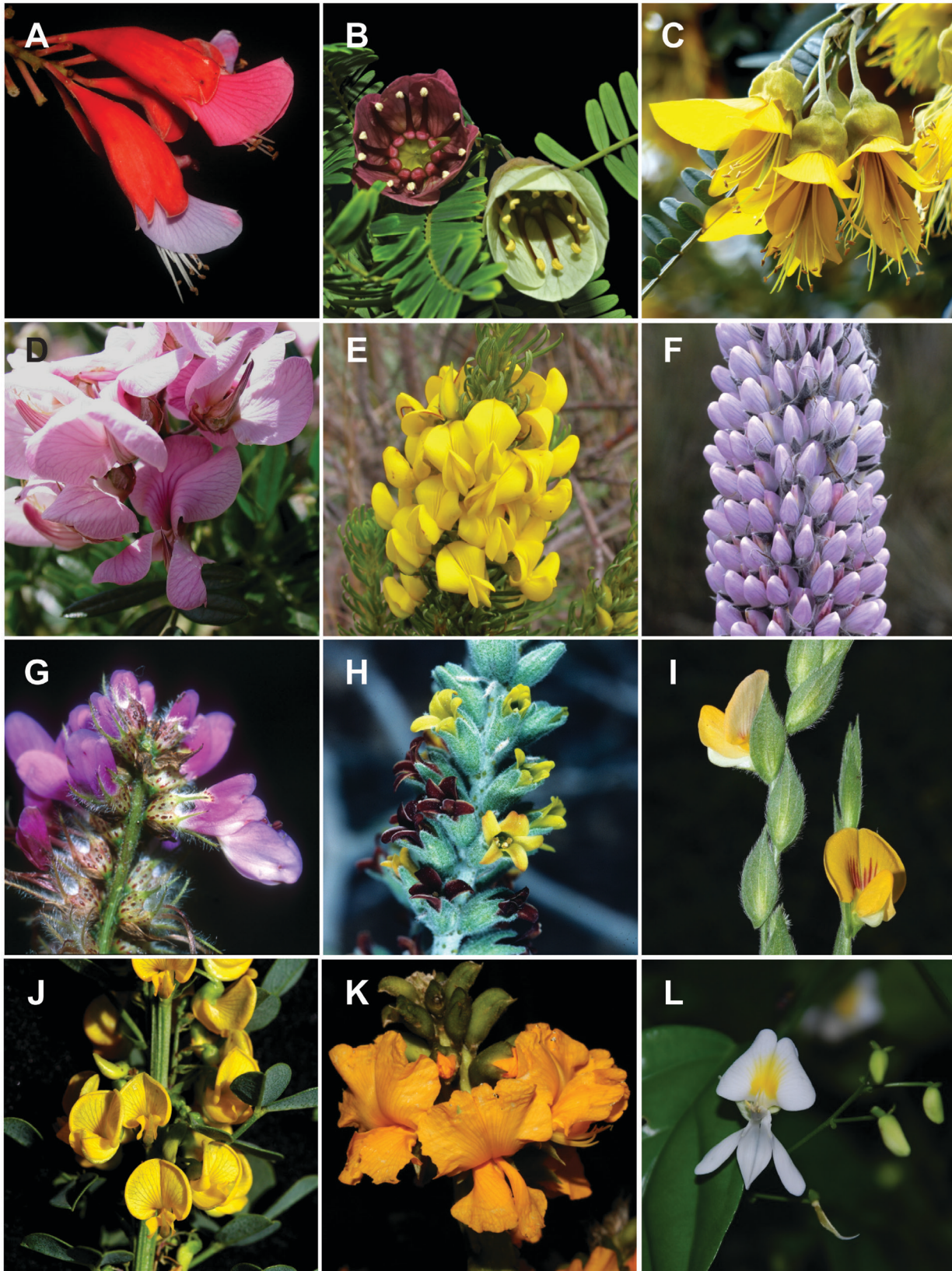


Fig. 8. Papilionoideae II. **A**, *Uleanthus erythrinoides*; **B**, *Cadia purpurea*; **C**, *Sophora* cf. *microphylla*; **D**, *Virgilia divaricata*; **E**, *Cyclopia pubescens*; **F**, *Lupinus weberbaueri*; **G**, *Dalea botterii*; **H**, *Errazurizia megacarpa*; **I**, *Zornia reticulata*; **J**, *Poirertia tetraphylla*; **K**, *Pterocarpus amazonum*; **L**, *Baphia leptobotrys*. — Photos: A, I & K, Domingos Cardoso; B, Wolfgang Stuppy; C, Gwilym Lewis; D & E, Stephen Boatwright; F, H & J, Colin Hughes; G, Donovan Bailey; L, Jan Wieringa.

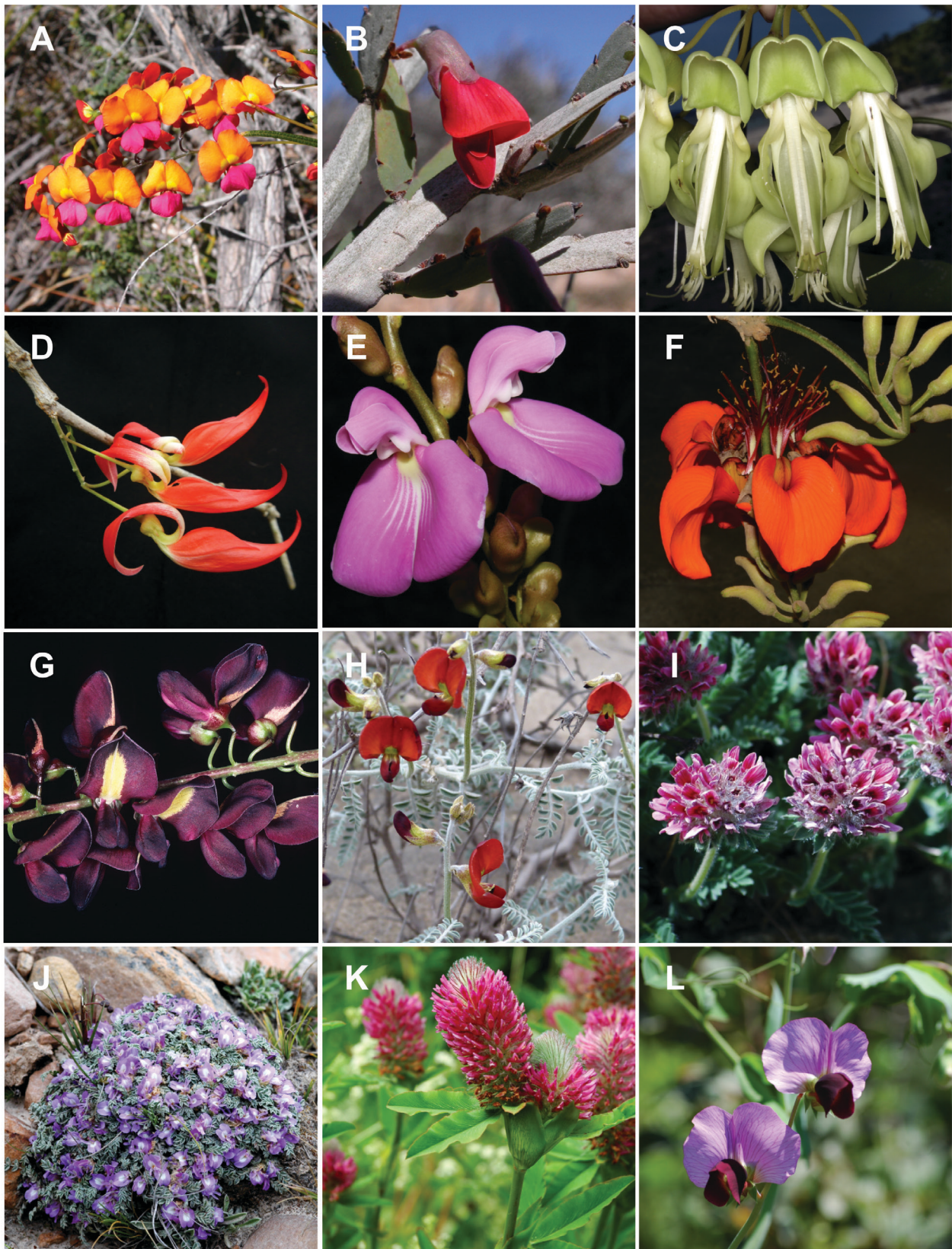


Fig. 9. Papilionoideae III. **A**, *Chorizema glycinifolium*; **B**, *Bossiaea walkeri*; **C**, *Mucuna gigantea*; **D**, *Chadsia longidentata*; **E**, *Canavalia brasiliensis*; **F**, *Erythrina velutina*; **G**, *Gliricidia robusta*; **H**, *Poissonia weberbaueri*; **I**, *Anthyllis montana*; **J**, *Astragalus uniflorus*; **K**, *Trifolium rubens*; **L**, *Pisum sativum* subsp. *biflorum*. — Photos: A & B, Michael Crisp; C, Timothy Utteridge; D, Erik Koenen; E, Domingos Cardoso; F, Luciano P. de Queiroz; G & I–L, Colin Hughes; H, Justin Moat.

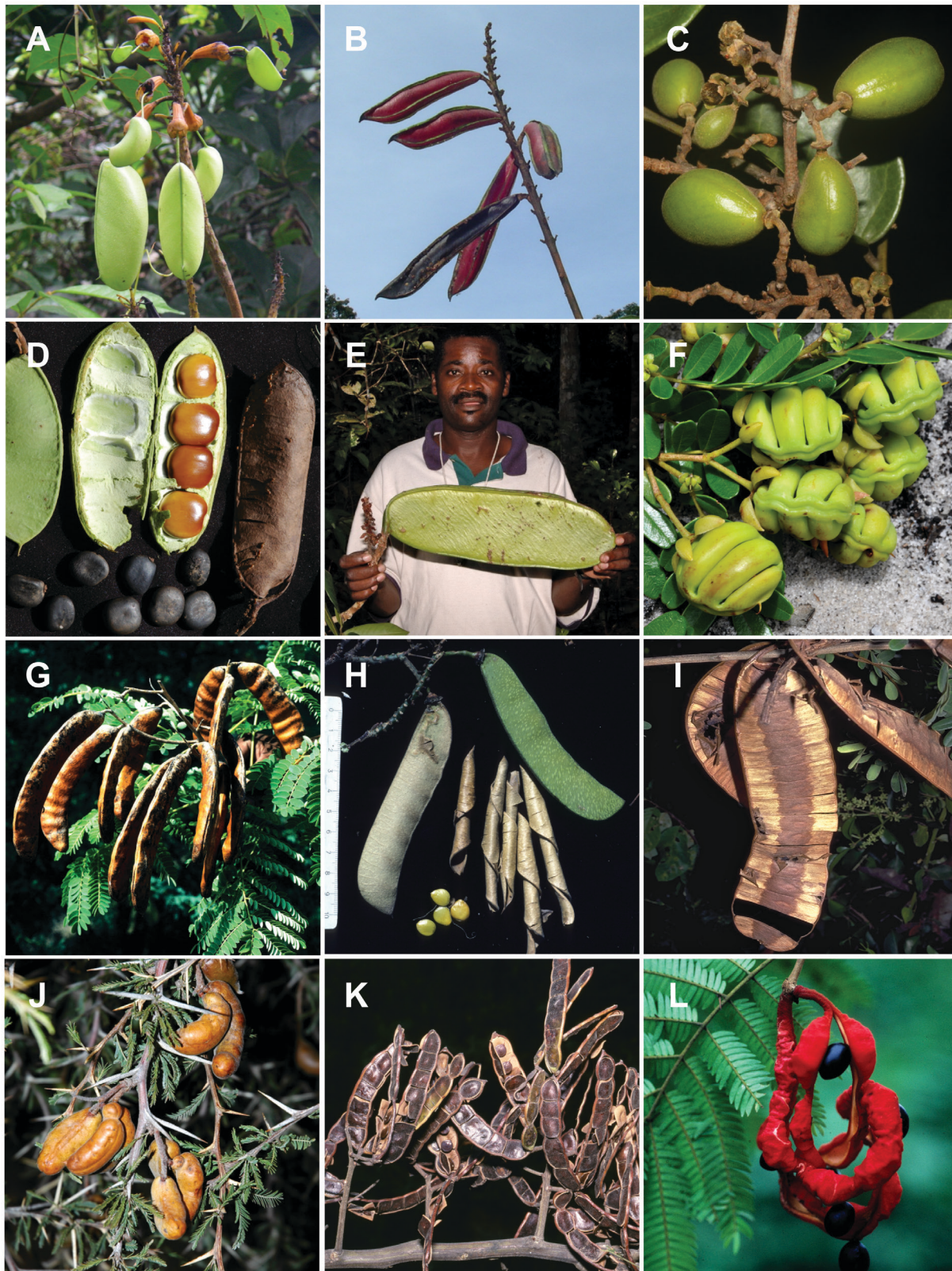


Fig. 10. Legume fruit diversity I. **A**, Cercidoideae; **B**, Duparquetioideae; **C**, Dialioideae; **D & E**, Detarioideae; **F–L**, Caesalpinoideae. **A**, *Griffonia physocarpa*; **B**, *Duparquetia orchidacea*; **C**, *Dialium guianense*; **D**, *Brodriguesia santosii*; **E**, *Berlinia razzifera*, held by Jean-Claude Mouzanda; **F**, *Eligmocarpus cynometroides*; **G**, *Heteroflorum sclerocarpum*; **H**, *Erythrostemon coccineus*; **I**, *Entada polystachya*; **J**, *Prosopis ferox*; **K**, *Mimosa townsendii*; **L**, *Cojoba arborea*. — Photos: A & B, Xander van der Burgt; C, Domingos Cardoso; D, G, H, J & L, Colin Hughes; E, David Harris; F, Felix Forest; I & K, Gwilym Lewis.

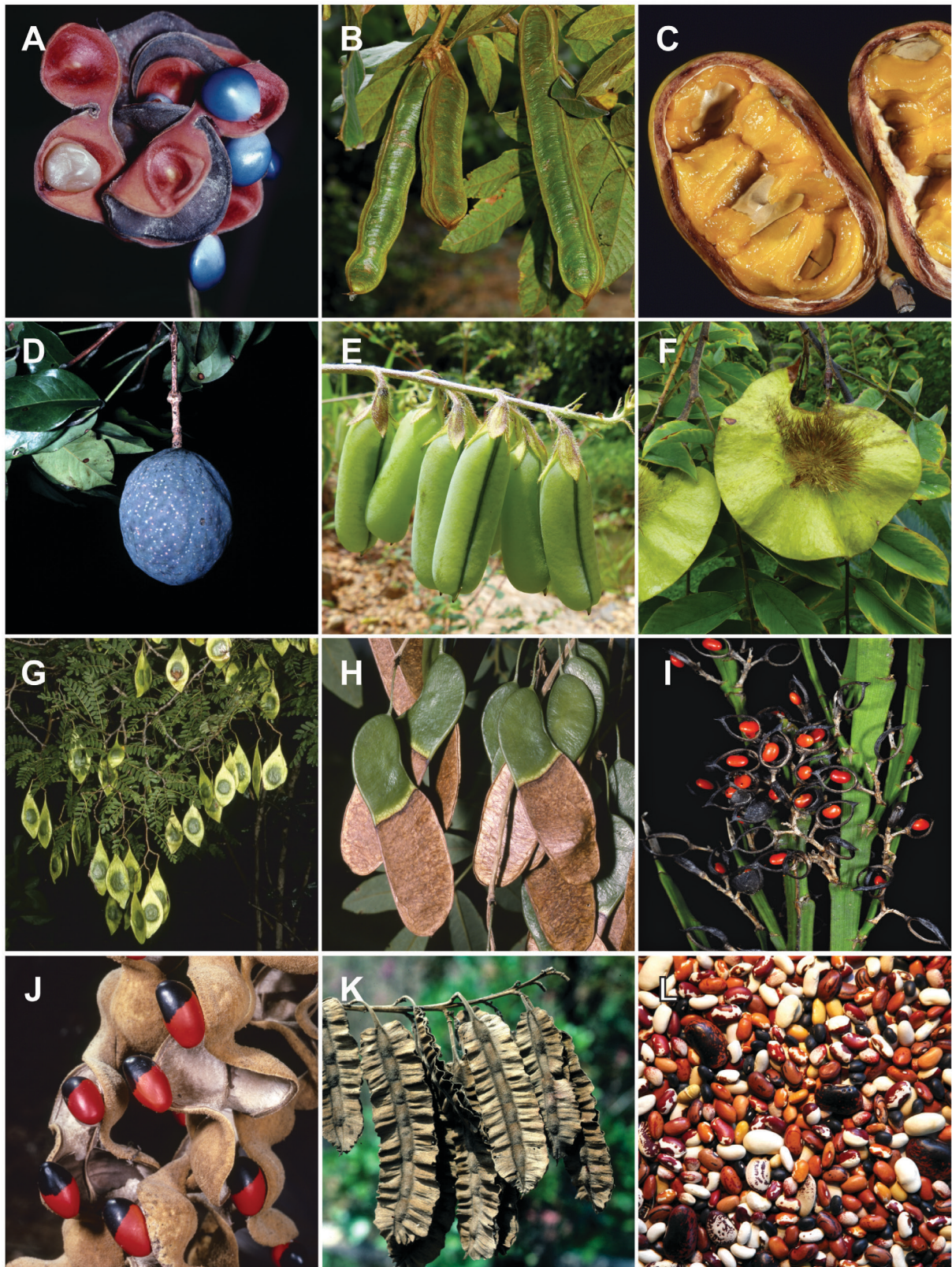


Fig. 11. Legume fruit diversity II. **A & B**, Caesalpinoideae; **C–L**, Papilionoideae. **A**, *Abarema jupunba*; **B**, *Inga feuillei*; **C**, *Swartzia parvipetala*; **D**, *Andira micrantha*; **E**, *Crotalaria* cf. *stipularia*; **F**, *Pterocarpus angolensis*; **G**, *Dalbergia lemurica*; **H**, *Machaerium millei*; **I**, *Carmichaelia* cf. *aligera*; **J**, *Erythrina madagascariensis*; **K**, *Piscidia grandifolia*; **L**, *Phaseolus* spp. — Photos: A & D, Projecto Flora Reserva Ducke, INPA/DFID, comm. Mike Hopkins; B & I, Wolfgang Stuppy; C, James Ratter; E, F & H, Gwilym Lewis; G & J, David Du Puy; K & L, Colin Hughes.

2015), and draft manuscripts circulated again to the LPWG membership for further comment prior to submission of this paper for publication.

After broad consultation within the legume systematics community, it was generally agreed that a six subfamily classification is the most appropriate option for naming subfamilies in a Linnaean system (Figs. 1, 2, S2 & S3). The six subfamily option is based on a set of clades with robust support (1.00 Bayesian posterior probabilities and 100% maximum likelihood bootstrap values in Figs. 1, 2 & S1) that are each subtended by long branches: Cercidoideae (Fig. 3A–F), Detarioideae (Fig. 4), Duparquetioideae (Fig. 3G), Dialioideae (Fig. 3H–L), Papilionoideae (Figs. 7–9), and the recircumscribed Caesalpinioideae (which equates to the MCC clade; Figs. 5 & 6). In addition to the molecular support all six subfamilies have

support from morphological data (Table 1). While morphological circumscription of the six subfamilies is not entirely straightforward given the complex and homoplasious nature of most morphological characters (Table 1; see Taxonomy below), it is certainly no more difficult or problematic than for the traditional three subfamilies, for which the supposed diagnostic morphological (mainly conspicuous floral) characters are beset by numerous exceptions, and where Caesalpinioideae, as traditionally circumscribed, lacks obvious diagnosability. Although Papilionoideae and the re-circumscribed Caesalpinioideae are still large and heterogeneous clades, the former retains its current definition and generic membership (Polhill, 1994; Lewis & al., 2005) (Table 2), while the latter is now more homogeneous, including, for example, all legumes with bipinnate leaves and most with extrafloral nectaries on the petiole and rachis (Fig. 2;

Table 1. Comparative morphology, chemistry and chromosome numbers of the six subfamilies of Leguminosae. The text in bold highlights characters and character states that are particularly valuable for identifying members of the subfamilies. See glossary in Appendix 2 and Figs. 12 & 13 for definitions and illustrations of technical terms.

	Cercidoideae	Detarioideae	Duparquetioideae	Dialioideae	Caesalpinioideae	Papilionoideae
Habit	Trees, shrubs or lianas, many with tendrils, mostly unarmed but frequently with prickles or infrastipular spines; branches rarely modified into cladodes	Usually unarmed trees, sometimes shrubs, rarely suffruticose	Unarmed scrambling liana	Unarmed trees or shrubs, rarely suffruticose	Trees, shrubs, lianas, suffruticose or functionally herbaceous, unarmed or commonly armed with prickles or spines	Usually unarmed trees, shrubs, lianas, herbs, or twining vines with tendrils
Specialised extrafloral nectaries	Mostly lacking	Often present on the underside, rarely on the margins of leaflets or on the leaf rachis	Lacking	Lacking	Often present on the petiole and/or on the primary and secondary rachises, usually between pinnae or leaflet pairs , sometimes on stipules or bracts	Lacking on petiole and leaf rachis; occasionally present on stipules, stipels, bracts, or swollen and nectar-secreting peduncles or sepals
Stipules	Lateral, free	Intrapetiolar (i.e., somewhere between the petiole and the axillary bud) and then free, valvate and connected by chaffy hairs, or fused, either partly (only at base) or entirely, rarely lateral and free	Lateral, free	Lateral, free or absent	Lateral, free or absent	Lateral, free or absent, very rarely interpetiolar
Leaves	Unifoliolate or bifoliolate	Usually paripinnate or bifoliolate, rarely unifoliolate	Imparipinnate	Usually imparipinnate , rarely paripinnate, 1-foliolate or palmately compound	Commonly bi-pinnate , otherwise pinnate, and then mostly paripinnate , rarely imparipinnate or bifoliolate, modified into phyllodes or lacking	Mostly pari- or imparipinnate or palmately compound , commonly unifoliolate, trifoliolate , rarely bifoliolate or tetrafoliolate

Table 1) (e.g., Marazzi & al., 2012). The six subfamilies have similar stem ages, all having apparently diverged soon after the first appearance of the family (Lavin & al., 2005; Bruneau & al., 2008; Simon & al., 2009).

The major disadvantage of adopting a six subfamily classification, namely abandoning the well-known Mimosoideae, is mitigated by continuing to recognise this lineage as a distinct clade, informally referred to as the mimosoid clade at this point, but with scope to be formally named as a tribe within a new Linnaean tribal classification, and/or in a rank-free clade-based phylogenetic classification of new sense Caesalpinioideae, once relationships within this subfamily are better resolved. It is also worth noting that options recognising fewer than six subfamilies would both reduce morphological diagnosability and result in subfamilies with even more unwieldy morphological

heterogeneity. The six subfamily option minimises the number of new Linnaean names, which is likely to be more easily accepted by a wider user community, and we considered this option as more likely to remain stable through time. With a six subfamily system, we are ensuring greater nomenclatural stability than a system that would describe 11 or more new subfamilies, particularly as several of the additional subfamilies that would need to be recognised lack robust support in current phylogenies being subtended by short branches (Figs. 1, 2 & S1) (Bruneau & al., 2008; Manzanilla & Bruneau, 2012; E.J.M. Koenen & al., in prep.) and might later need to be changed.

Although Caesalpinioideae DC. and Mimosoideae DC. have equal priority under the *International Code of Nomenclature for algae, fungi and plants (Melbourne Code)* (McNeill & al., 2012) because they were published in the same

Table 1. Continued.

	Cercidoideae	Detarioideae	Duparquetioideae	Dalioideae	Caesalpinioideae	Papilionoideae
Leaflets and pinnae	Opposite (when bifoliolate); blade (when unifoliolate) entire or bilobed	Opposite or alternate; translucent glands sometimes present	Opposite; blade entire	Alternate, rarely opposite	Mostly opposite, rarely alternate	Opposite or alternate, sometimes modified into tendrils, rarely in phyllodes
Inflorescence	Raceme or pseudoraceme	Raceme or panicle	Terminal raceme	Branched, thyrsoid inflorescences , less commonly racemes with distichous flower arrangement, or flowers solitary	Globose, spikes, panicles, racemes or flowers in fascicles	Mostly racemes, pseudoracemes or panicles, less often cymes, spicate or capitate, or flowers solitary
Bracteoles	Large or minute	Large or small, frequently petaloid, valvate, imbricate or partly fused with the hypanthium, partially or completely enclosing the bud	Small	Small or absent	Small or absent	Mostly small, rarely large, valvate, enveloping the bud
Flowers	Bisexual, rarely unisexual, slightly to strongly bilaterally symmetrical, sometimes papilionate	Bisexual or with both bisexual and male flowers, radially or slightly to strongly bilaterally symmetrical, but never papilionate	Bisexual, strongly bilaterally symmetrical, never papilionate	Bisexual, radially or slightly to strongly bilaterally symmetrical, sometimes papilionate	Usually bisexual, rarely unisexual, or bisexual flowers combined with unisexual and/or sterile flowers in heteromorphic inflorescences; radially, less frequently bilaterally symmetrical, sometimes papilionate or asymmetric	Bisexual, rarely unisexual, usually bilaterally symmetrical, usually papilionate , rarely asymmetrical, radially symmetrical or nearly so
Hypanthium	Present, greatly elongated to almost absent	Present, elongated to almost absent	Absent	Usually absent, rarely present, receptacle may be broad and flattened, bearing nectary-like bodies	Lacking or cupular , rarely tubular	Present or absent

Table 1. Continued.

	Cercidoideae	Detarioideae	Duparquetioideae	Dialioideae	Caesalpinioideae	Papilionoideae
Sepals	United in a spatheous or 2–5-lobed calyx or sepals free	Commonly 5 or 4 (two adaxial sepals often fused), rarely some or all absent or more (–7)	4, unequal, free, the abaxial and adaxial sepals cucullate and sepaloid, the laterals petaloid	(3 or 4)–5–(6), free, equal to sub-equal	(3–)5(–6), free or fused, or sepal whorl lacking	(3–)5, united at least at the base, sometimes entire and splitting into irregular lobes or lobes dimorphic and some petaloid
Petals	5, rarely 2, 6 or absent, free, when present imbricate, the adaxial petal innermost and frequently differentiated	0–5(–7), free, when present imbricate, the adaxial petal generally outermost , all equal or the adaxial large and either the other 4 or only the abaxial ones smaller to rudimentary	5, free, imbricate, the adaxial petal outermost , adaxial and two lateral petals ovate, two abaxial petals strap-like, oblong, all 5 petals with stalked glands along their margins	5 or fewer (0, 1, 3, 4), rarely 6, free, imbricate, the adaxial petal innermost , all equal to sub-equal	(3–)5(–6), free or fused, or petal whorl lacking, valvate or imbricate, then adaxial petal innermost	Usually (0–)5(–6), rarely 1 (standard) petal and 4 absent, imbricate, the adaxial petal outermost , in radially symmetrical flowered species, corolla with 5 small or undifferentiated petals, less often only one (standard) petal is present or all petals absent
Stamens	Usually 10 (sometimes fewer) in two whorls of alternate length	Usually 10, sometimes 2–numerous	4	5 or fewer, rarely 6–10, uniform, rarely dimorphic	Diplostemonous or haplostemonous, sometimes reduced to 3, 4 or 5, frequently many (100+), sometimes heteromorphic, some or all sometimes modified or staminodial	Usually 10, rarely 9 or many
Stamen fusion	Filaments partly connate or free	Filaments partly connate or free	Filaments free	Filaments free	Filaments free or connate	Filaments usually connate into a sheath or tube , uppermost filament wholly or partly free, sometimes all filaments free
Anthers	Mostly uniform, dorsifixed, dehiscing via longitudinal slits or central pores; reduced stamens or staminodes sometimes present	Mostly uniform, dorsifixed or basifixed, dehiscing via longitudinal slits	Uniform, basifixed, with pointed appendages, dehiscing via short apical, poricidal slits; anthers post-genitally fused into a curving synandrium	Uniform, rarely dimorphic, basifixed, rarely dorsifixed, dehiscing via longitudinal slits, often reduced to short apical, poricidal slits; staminodes present or absent	Uniform or heteromorphic, basifixed or dorsifixed, often with a stipitate or sessile apical gland, dehiscing via longitudinal slits or apical or basal poricidal slits or pores	Uniform or dimorphic, basifixed or dorsifixed, dehiscing via longitudinal slits
Pollen	Monads, 3-colporate, 3–6-colpate, 3-porate, 3-porate, 3–4-colporoidate or inaperturate, rarely in tetrads	Monads, mostly 3-colporate with a vast array of sculptures	Monads, asymmetrical, one equatorial-encircling ectoaperture with two equatorial endoapertures	Monads tricolporate, with punctate or finely reticulate, rarely striate sculpture patterns	Monads, tricolporate or porate tetrads, bitetrads or polyads , sculpture pattern never striate	Monads, mostly 3-colporate, 3-colpate or 3-porate
Gynoecium	1-carpellate, stipitate, stipe free or adnate to the wall of the hypanthium	1-carpellate, stipitate, stipe free or adnate to the wall of the hypanthium	1-carpellate, stipitate, stipe free	1-carpellate, sometimes 2-carpellate, stipitate or sessile, stipe free	Usually 1-carpellate, rarely polycarpellate , stipitate or sessile, stipe free	Usually 1-carpellate, rarely 2-carpellate, stipitate or sessile, stipe free

Table 1. Continued.

	Cercidoideae	Detarioideae	Duparquetioideae	Dialioideae	Caesalpinioideae	Papilionoideae
Ovules	Ovary 1–many-ovulate	Ovary 1–many-ovulate	Ovary 2–5-ovulate	Ovary frequently 2-ovulate, rarely 1–many-ovulate	Ovary 1–many-ovulate	Ovary 1–many-ovulate
Fruits	Dehiscent pods, often explosive with twisted valves, or indehiscent, then generally samaroid	Mostly woody, de- hiscent pods, some- times indehiscent and woody or thin valved samaroid, rarely filled with pulpy mesocarp or endocarp	Woody dehiscent pods, 4-angled, valves spirally coiled	Commonly inde- hiscent drupaceous or samaroid , rarely dehiscent or the dru- paceous fruit with indurating endocarp into one-seeded segments	Commonly thin- valved, 1–many- seeded pod, dehis- cent along one or both sutures, also often a lomentum, a craspedium, or thick and woody and then indehiscent or explosively dehiscent, often curved or spirally coiled	Dehiscent pods along one or both sutures, or indehis- cent, or lomenta, samaras or drupes
Seeds	With apical cre- scent-shaped hilum , rarely circular; lens inconspicuous, lack- ing pleurograms , pseudopleuro- grams , wing or aril	Often overgrown, sometimes hard and then occasionally with pseudopleuro- grams ; occasionally arillate	2–5, oblong to ovoid, the testa thick, lacking pleurograms	1–2, rarely more, lacking pleurograms	Usually with an open or closed pleurogram on both faces , sometimes with a fleshy aril or sarcotesta, some- times winged; hilum usually api- cal; lens usually inconspicuous	Usually with hard testa, rarely over- grown, sometimes with a fleshy aril or sarcotesta; com- plex hilar valve, elongate hilum and lens usually pres- ent, pleurogram lacking
Embryo	Straight, very rarely curved	Straight	Straight	Straight	Straight	Usually curved, rarely straight
Vestured pits in 2° xylem	Lacking	Present	Lacking	Usually lacking, rarely present	Present	Present
Root nodules	Absent	Absent	Absent	Absent	Variably present and indeterminate	Usually present, either indeterminate or determinate
Chromosome counts	2n = 14, 24, 26, 28 (42, 56)	2n mostly 24 (occa- sionally 16, 20, 22, 36, 68)	Unknown	2n = 28 (most gen- era not surveyed)	2n mostly 24, 26, 28 (but 14, 16, 52, 54, 56 also reported)	2n mostly 16, 18, 20, 22 (but 12, 14, 24, 26, 28, 30, 32, 38, 40, 48, 64, 84 also reported)
Chemistry	Coumarins and cyanogenic glu- cosides reported; non-protein amino acids common (5-hydroxy-L-tryp- tophan only reported to this subfamily)	Coumarins reported, frequently with terpenes (resins) and non-protein amino acids	Chemical character- istics unknown	Chemical character- istics unknown	Non-protein amino acids frequently reported; coumarins, cyanogenic gluco- sides, phenylethyl- amine, tryptamines, and β-carboline alkaloids also reported	Isoflavonoids, prenylated flavonoids, indol- izidine or quino- lizidine alkaloids reported. Non- protein amino acids widespread, some exclusively found in the subfamily (e.g., canavanine)

volume by Candolle (1825), Caesalpinioideae was chosen as the preferred name for the MCC clade. Because of the broader concept associated with Caesalpinioideae, it corresponds more closely to the more inclusive MCC clade. Furthermore, this leaves open the option in future classifications of naming the morphologically distinct mimosoid clade at the tribal level and/or under the International Code of Phylogenetic Nomenclature (ICPN) (Cantino & de Queiroz, 2010).

In our new classification, three subfamily names are new at this rank. We ascribe these names to the collective known as the “Legume Phylogeny Working Group”. This uncommon practice in botanical nomenclature does not prevent valid publication of the names under the botanical code as stipulated in Chapter VI, Section 1 (Author Citations). Although we could have adopted a modification of Recommendation 46C.2, which suggests citing the first author followed by “et al.” (and at first appearance of that authority, listing all 97 authors), we considered that ascribing authorship to the Legume Phylogeny Working Group is more straightforward, more clearly gives due credit to the legume systematics community and reflects much better the collaborative approach used to arrive at this new classification. At a time when systematics papers may have increasing numbers of authors, for example, as genomic datasets become routine, we feel that a desire for authorship ascribed to research groups and communities rather than individuals will become more commonplace.

■ INTEGRATING TRIBAL AND CLADE-BASED CLASSIFICATIONS

In addition to the need for a new Linnaean-based subfamily classification, there are important questions about the best approach to naming clades within subfamilies. New phylogenies of many legume groups have unequivocally demonstrated the inadequacies of the tribal classifications of Polhill & Raven (1981), Polhill (1994), and Lewis & al. (2005) because of the non-monophyly of most of the traditionally recognised tribes (LPWG, 2013a). In addition, questions remain about the monophyly and placement of several genera, with considerable ongoing uncertainty surrounding generic delimitation and relationships (LPWG, 2013a; Lewis & al., 2013). However, numerous phylogenetic studies are ongoing and revised tribal classifications of subfamilies will be forthcoming in the near future. The emergence of clade-based phylogenetic classification systems provides an additional option to facilitate rank-free naming of robustly supported legume clades under the draft ICPN. Such clade-based classifications can be easily integrated with traditional Linnaean rank-based classification to name additional clades coinciding with the evolution of key biological traits that are hypothesised as synapomorphies. For example, several important legume clades corresponding to biologically important apomorphies (sometimes in the form of deep homologies), including nodulation, bipinnate leaves (here corresponding to the redefined Caesalpinioideae), extrafloral petiolar or leaf rachis nectaries, pollen in tetrads/polyads, and valvate petal aestivation (mimosoid clade) could

be named in this way, as pursued by Wojciechowksi (2013) for Papilionoideae using many of the recommendations of the ICPN. We believe this approach, integrating Linnaean ranks alongside clade-based ICPN classification, will greatly enhance the biological meaning and utility of future classifications with significant benefits for effective communication across a wide spectrum of biological disciplines.

A new classification is clearly needed for the recircumscribed subfamily Caesalpinioideae, which has been the most difficult and controversial to delimit in the new subfamilial classification because of the inclusion of the formerly recognised and morphologically distinctive subfamily Mimosoideae. Because relationships amongst major groups within the recircumscribed Caesalpinioideae remain poorly resolved (Figs. 1, 2 & S1), we refrain from establishing a new tribal and/or clade-based classification for this subfamily here. Although most mimosoids are morphologically distinct (Fig. 6), the morphological distinctions between some members of the mimosoid clade and subtending caesalpinoid lineages are not always clearcut. For example, *Dinizia* Ducke, once considered to be in Mimosoideae, is placed outside the mimosoid clade in molecular phylogenetic analyses (Luckow & al., 2003; Bruneau & al., 2008), and *Chidlowia* Hoyle (Fig. 6A), which has always been considered a caesalpinoid legume (Lewis & al., 2005), is placed within the mimosoid clade in recent molecular phylogenetic analyses (Manzanilla & Bruneau, 2012; E.J.M. Koenen & al., in prep.). For these reasons, we refrain from formally naming this clade until relationships amongst lineages within Caesalpinioideae can be better resolved, and refer to the former subfamily Mimosoideae DC. simply as the “mimosoid clade” for the time being.

In Cercidoideae and Dialioideae, both of which have relatively few genera (Lewis & Forest, 2005; Sinou & al., 2009; E. Zimmerman, unpub. data), infra-subfamilial classifications may not be needed, and Duparquetioideae is monospecific. In Detarioideae, phylogenetic relationships amongst basal lineages have been too poorly resolved until now to permit their classification (Bruneau & al., 2001, 2008; Fougère-Danezan & al., 2007), but ongoing studies are leading to better resolution with the possibility for recognising clades as tribes and/or formally named clades (M. de la Estrella & al., unpub. data). Similarly, ongoing studies in Papilionoideae and in the recircumscribed Caesalpinioideae should help resolve key relationships, with the ultimate outcome that names of strongly supported and biologically meaningful clades will be proposed in forthcoming publications.

■ REFERENCE PHYLOGENY

The classification proposed here uses as its framework the most comprehensively sampled phylogenetic analysis of legumes to date (Figs. 1, S1; Table S1; Methods described in Appendix 1). This new phylogeny is based on plastid *matK* gene sequences because this gene region is the most widely sequenced across the legumes (cf. LPWG, 2013a) and it is sufficiently variable to resolve generic membership of many

strongly supported higher-level clades as demonstrated by a large number of studies such as those referenced herein. Although this analysis is based on a single plastid locus, the topology observed and the groups that are supported have been consistently resolved in numerous previous phylogenetic analyses of the entire family or of particular clades within the family using diverse plastid (*trnL-F*, *trnD-T*, *rbcL*, *rps16*, *rpl16*) and nuclear loci (e.g., rDNA ITS, *SucS*) (see LPWG, 2013a and references therein). In recent analyses of all 81 plastid genes (Fig. 2) and of a large nuclear gene dataset derived from transcriptome sequences (E.J.M. Koenen & al., in prep.), all five of the non-monospecific subfamilies are strongly supported, and the relationships amongst them do not conflict with the *matK* analyses (see below), although the nuclear gene dataset does not include Duparquetioideae.

The analysis presented here includes 3696 legume species (with an additional 48 infraspecific taxa) representing 698 of the currently recognised 765 legume genera (Figs. 1 & S1; Tables 2 & S1; Appendix 1). Subfamilies Cercidoideae and Duparquetioideae are fully sampled at the generic level. In the Detarioideae, five genera are not sampled, all of them monospecific, in Dialioideae two monospecific genera are missing, and in Caesalpinioideae, two genera are not sampled (Table 2, missing genera identified with *). Papilionoideae are represented by 445 genera, with most of the missing 48 genera belonging to the tribe Loteae and phaseoloid clades. The phylogenetic trees and the underlying alignment and voucher data are available to browse and download from the online Supplementary Data (Table S1; Data Files A–F) and on Data Dryad (DOI: <https://doi.org/10.5061/dryad.61pd6>).

Bayesian analyses (Fig. 1) and maximum likelihood (Fig. S1) of the *matK* sequence data resolve the Leguminosae as monophyletic with 1.0 posterior probability and 100% bootstrap support. Each of the five non-monospecific subfamilies of Leguminosae is also supported with 1.0 posterior probability and 100% bootstrap support. Relationships amongst subfamilies Cercidoideae, Detarioideae, Duparquetioideae and the clade that groups the remaining legumes (i.e., the other three subfamilies) are unresolved, forming a basal polytomy (Fig. 1). Dialioideae is resolved as sister to Caesalpinioideae+Papilionoideae (1.0 posterior probability, Fig. 1; 100% bootstrap support, Fig. S1), which are sister to each other. In the full plastid analyses of E.J.M. Koenen & al. (in prep.), Duparquetioideae is robustly supported as sister to the Dialioideae+Caesalpinioideae+Papilionoideae clade, but the relationship of this clade to the Cercidoideae and Detarioideae remains unresolved (Fig. 2). Many genera of Leguminosae are supported as monophyletic in the *matK* analysis, with notable exceptions for certain large genera that are the focus of ongoing taxonomic and phylogenetic studies (e.g., *Bauhinia* s.l. in Cercidoideae, several genera of Detarioideae, of the mimosoid clade, and of tribe Millettieae in Papilionoideae). In the mimosoid clade, and in other parts of Caesalpinioideae and Detarioideae, genera are often not supported as monophyletic, and generic-level relationships are often poorly resolved. This can likely be attributed in part to striking substitution rate heterogeneity in plastid genes, and hence variable phylogenetic

resolution across legumes, as previously noted by Lavin & al. (2005) and Dugas & al. (2015) (see also Figs. 1 & 2).

Several recent large-scale angiosperm/rosid phylogenetic analyses (Zanne & al., 2014; Li & al., 2016; Sun & al., 2016) included thousands of legume nuclear and plastid and, in some cases, mitochondrial sequences. These analyses contain many taxa that were mis-identified or labelled using outdated taxon names, or are based on apparent sequence contaminants that have been deposited in GenBank without being properly checked and annotated. These inaccuracies, compounded by large amounts of missing data (e.g., 80% in Zanne & al., 2014), together interact to cause unpredictable and chaotic problems in phylogenetic analyses, a phenomenon highlighted several years ago by McMahon & Sanderson (2006) in their supermatrix phylogenetic analysis of papilionoid legumes. Unfortunately, such potentially flawed topologies have been used as the basis for several recent large-scale evolutionary studies focused, for example, on key characteristics of legumes such as the origins of nodulation and nitrogen fixation (e.g., Werner & al., 2014, 2015; Li & al., 2016). A cursory examination of many of these large-scale phylogenies has revealed a number of unusual and demonstrably inaccurate relationships. Using such badly flawed phylogenies can obviously lead to weak or even erroneous conclusions regarding the evolution of particular traits (cf. Doyle, 2016). In contrast, the phylogeny presented here uses a fully curated set of sequences that are vouchered and taxonomically validated by the legume systematics community.

The phylogenetics of legumes, like that of any major clade, is of course a work in progress. New densely sampled phylogenies at the species, generic and higher levels based on full plastome sequences, as well as transcriptomes and hundreds of nuclear loci are ongoing, and will in due course supersede the phylogeny presented here. Regardless, the taxonomically validated tree presented here can be used for downstream analyses that require an accurate and densely sampled phylogenetic framework of the Leguminosae.

■ TAXONOMY

Based on the phylogenetic structure of the family Leguminosae presented here, we recognise six subfamilies. We provide a key, taxonomic descriptions for each of the subfamilies, and illustrate the diversity of flowers and fruits across these subfamilies (Figs. 3–11). Comparative morphology, chemistry and chromosome numbers of the six subfamilies (Table 1) and a full list of genera by subfamily, noting recent synonyms (Table 2) are presented. Technical terms are defined and illustrated in Appendix 2 and Figs. 12 & 13.

Key to the subfamilies of Leguminosae

1. Petals with marginal glandular structures; flowers with 4 stamens, anthers fused in a synandrium with poricidal dehiscence; leaves once pinnate; endemic to Central and West Africa **Duparquetioideae**
1. Petals not glandular (except in the Amazonian, papilionoid

- genus *Petaladenium*); flowers with 4 stamens uncommon (but then anthers never fused in a synandrium); anther dehiscence longitudinal (except poricidal in some genera of Caesalpinioideae); leaves various; widely distributed ... **2**
2. Flowers mostly papilionate (“pea-flowered”) and bilaterally symmetrical, less commonly radially symmetrical; median (standard) petal outermost, enclosing the wing and keel petals (especially in bud) or the wing and keel petals lacking; sepals united, at least at the base, into a calyx tube or completely enclosing the floral bud; seeds with a complex hilar valve, pleurogram absent; embryo radicle usually curved **Papilionoideae**
2. Flowers not papilionate (if rarely appearing papilionate then the median petal innermost), either bilaterally or radially symmetrical, median (standard) petal innermost, or petals valvate (in the mimosoid clade of the Caesalpinioideae); sepals free or fused; seeds lacking complex hilar valve, with or without a pleurogram; embryo radicle usually straight **3**
3. Leaves bipinnate; seeds commonly with an open or closed pleurogram on each side **Caesalpinioideae**
3. Leaves never bipinnate; seeds without an open or closed pleurogram on either side **4**
4. Leaves unifoliolate, bilobed or entire, or compound and bifoliolate; seed hilum circular or crescent-shaped **Cercidoideae**
4. Leaves various; if simple or bifoliolate, then the seed hilum not crescent-shaped, and rarely circular **5**
5. Extra-floral nectaries and other glandular structures (when present) on the lower surface or margin of leaflets; stipules usually intrapetiolar (free or united), rarely lateral **Detarioideae**
5. Extra-floral nectaries absent or present on the petiole or on the leaf rachis; stipules lateral and free or absent **6**
6. Inflorescences highly branched and thyrsoid or racemes with distichous anthotaxy; leaves mostly imparipinnate with alternate leaflets (rarely paripinnate with opposite leaflets in *Eligmocarpus* and *Poeppegia*), extra-floral nectaries on the petiole or leaf rachis absent **Dialioideae**
6. Inflorescences mostly racemes with spiral anthotaxy, commonly compounded into branched panicles or contracted in spikes or fascicles; leaves mostly paripinnate with opposite leaflets, rarely bifoliolate or with alternate leaflets; extra-floral nectaries (when present) on the petiole or on the leaf rachis between the leaflet pairs **Caesalpinioideae**

Descriptions of the six subfamilies

A short description is presented for each subfamily, highlighting (in **bold**) the diagnostic features of each.

Subfam. **Cercidoideae** Legume Phylogeny Working Group, **stat. nov.** ≡ Cercideae Bronn, Form. Pl. Legumin.: 134, 131. 1822 (“Cerceae”) – Type: *Cercis* L.

Trees, shrubs or tendriled lianas (Figs. 3A–F), mostly unarmed but frequently with prickles or infrastipular spines,

branches rarely modified into flattened cladodes (*Brenierea* Humbert); specialised extrafloral nectaries stipular when present (*Bauhinia* L.), never on petiole and leaf rachis. Stipules in lateral position, free. **Leaves uni- or bifoliolate (bipinnate, pinnate, palmate and trifoliolate leaves totally lacking), pulvinate, leaflet blade (when unifoliolate) entire or bilobed with a small mucro at the apex or between the lobes**, exstipellate. Inflorescence racemose or pseudoracemose; bracteoles minute or large. Flowers bisexual, rarely unisexual (plants polygamous or dioecious), slightly to strongly bilaterally symmetrical, sometimes papilionate (*Cercis*), hypanthium greatly elongated to almost absent; sepals united in a spathaceous or 2–5-lobed calyx or free; petals free, 5, rarely 2, 6 (some *Bauhinia*) or absent (*Brenierea*), imbricate, the adaxial petal innermost and frequently differentiated; stamens usually 10 (sometimes fewer) in two whorls of alternate length, the filaments partly connate or free, anthers mostly uniform and dorsifixed, opening by a longitudinal slit or central pore in each theca, reduced stamens or staminodes sometimes present; pollen 3-colporate, 3–6-colpate, 3-porate, 3-pororate, 3–4-colporoidate or inaperturate monads, rarely in tetrads (only in some *Bauhinia*); gynoecium 1-carpellate, stipe of ovary free or adnate to abaxial wall of the hypanthium, ovary 1–many-ovulate. Fruits dehiscent (often explosively with twisting valves) or indehiscent and then generally samaroid. Seeds with apical **crescent-shaped hilum**, rarely circular (*Cercis*), lens inconspicuous, lacking pleurograms, pseudopleurograms, or wing or aril (in *Brenierea* two funicular aril-like lobes adnate to the testa leaving a short crescent-shaped scar or a long scar running nearly around the seed circumference); embryo straight, very rarely curved (*Barklya* F.Muell.). Vestured pits lacking in secondary xylem; silica bodies absent; septate fibres and storeyed rays sometimes present. Root nodules absent. $2n = 14, 24, 26, 28 (42, 56)$. Coumarins and cyanogenic glucosides reported; non-protein amino acids common (5-hydroxy-L-tryptophan only reported in this subfamily, *Griffonia* Baill., *Brenierea*).

Currently 12 genera and ca. 335 species, mainly tropical, *Cercis* in the warm temperate Northern Hemisphere.

Clade-based definition (included taxa): The most inclusive crown clade containing *Cercis canadensis* L. and *Bauhinia divaricata* L. but not *Poeppegia procera* C.Presl, *Duparquetia orchidacea* Baill., or *Bobgunnia fistuloides* (Harms) J.H.Kirkbr. & Wiersema.

Subfam. **Detarioideae** Burmeist., Handb. Naturgesch.: 319. 1837 (“Detarieae”) – Type: *Detarium* Juss.

Unarmed trees, sometimes shrubs, rarely suffruticose (*Cryptosepalum* Benth.) (Fig. 4); specialised extrafloral nectaries often present abaxially, rarely on the margins of leaflets or on leaf rachis, and never on the petiole. **Stipules in intrapetiolar position (i.e., somewhere between the petiole and the axillary bud) and then free, valvate and connected by chaffy hairs, or fused, either partly (only at base) or entirely**, rarely lateral and free. **Leaves paripinnate** (ending in a pair of leaflets or, if leaflets alternate and appearing imparipinnate, the terminal leaflet exceeded by a more or less caducous

rachis-extension) with 1 (bifoliolate) to numerous pairs of leaflets, or rarely unifoliolate (*Paloue* Aubl., *Paloveopsis* R.S.Cowan, *Zenkerella* Taub., some *Cryptosepalum*, *Didelotia* Baill. and *Guibourtia* Benn.), bipinnate leaves totally lacking, leaves pulvinate, leaflets opposite or alternate, exstipellate, translucent glands sometimes present. Inflorescence a raceme or panicle; bracteoles small to **large, frequently petaloid, valvate or imbricate, free or partially fused or partly fused with the hypanthium, partially or completely enclosing the bud.** Flowers bisexual or with both bisexual and male flowers radially or slightly to strongly bilaterally symmetrical (but never papilionate), hypanthium elongated to almost absent; sepals commonly 5 or 4 (two adaxial sepals often fused), rarely some or all absent or more (–7); petals free, 0–5(–7), when present imbricate, the adaxial petal generally innermost (outermost in some flowers of *Hymenaea* L. and allies), all equal or the adaxial large and either the other 4 or only the abaxial ones smaller to rudimentary; stamens 2–numerous but usually 10, the filaments partly connate or free, staminodes occasionally present; anthers dorsifixed or basifixed; pollen mostly 3-colporate monads with a vast array of sculptures; gynoecium 1-carpellate, 1–many ovulate, stipe of ovary free or adnate to the wall of the hypanthium. Fruits mostly woody, dehiscent pods, sometimes indehiscent and woody or thin-valved, samaroid (*Brandzeia* Baill., *Barnebydendron* J.H.Kirkbr., *Gossweilerodendron* Harms, *Hardwickia* Roxb., *Neoapaloxylon* Rauschert), rarely filled with pulpy mesocarp (*Tamarindus* L.) or endocarp (*Hymenaea*). Seeds often overgrown, sometimes hard and then occasionally with pseudopleurograms (*Lysidice* Hance, *Paramacrolobium* J.Léonard, *Peltogyne* Vogel, *Tamarindus*), occasionally arillate; embryo straight. Vestured pits present in secondary xylem; axial (resin) canals sometimes present; silica bodies rarely present (*Hymenostegia* Harms, *Loesenera* Harms); septate fibres and storeyed rays sometimes present. Root nodules absent. $2n$ mostly 24 but occasionally 16, 20, 22, 36, 68. Coumarins reported; frequently with terpenes (resins) and non-protein amino acids.

Currently 84 genera and ca. 760 species, almost exclusively tropical, *Schotia* Jacq. in sub-tropical South Africa.

Clade-based definition (included taxa): The most inclusive crown clade containing *Goniorrhachis marginata* Taub. and *Aphanocalyx cynometroides* Oliv., but not *Cercis canadensis*, *Duparquetia orchidacea* or *Bobgunnia fistuloides*.

Subfam. **Duparquetioideae** Legume Phylogeny Working Group, **stat. nov.** ≡ *Duparquetiinae* H.S.Irwin & Barneby in Polhill & Raven, Adv. Legume Syst. 1: 102. 1981 – Type: *Duparquetia* Baill.

Unarmed, scrambling liana (Fig. 3G), often climbing to the forest canopy; specialised extrafloral nectaries lacking on petiole and leaf rachis. Stipules in lateral position, free, narrowly triangular. Leaves imparipinnate, pulvinate, leaflets opposite, exstipellate. Inflorescence a terminal, erect, 10–30-flowered raceme; bracteoles 2, small. Flowers bisexual, strongly bilaterally symmetrical, hypanthium lacking; **sepals 4, unequal, the abaxial and adaxial sepals cucullate, sepaloid, the lateral sepals petaloid;** petals 5, free, dimorphic, the adaxial and the

two lateral petals ovate, two abaxial petals strap-like, oblong, **all 5 petals with stalked gland-like extrusions along their margins**, imbricate, the adaxial petal outermost; stamens 4, the anthers basifixed, oblong, with pointed appendages, the thecae dehisce by a short, apical, poricidal slit, **the anthers postgenitally fused into a curving synandrium**, the appendages remain free; pollen in monads, **asymmetrical, one equatorial-encircling ectoaperture with two equatorial endoapertures;** gynoecium 1-carpellate, stipitate, 2–5-ovuled, with four ridges running along the length of the ovary. Fruit an oblong four-angled, woody pod, dehiscent, valves spirally coiled. Seeds 2–5 per fruit, oblong to ovoid, the testa thick; embryo straight. Vestured pits lacking in secondary xylem; silica bodies, septate fibres and storeyed rays absent. Root nodules absent. Chromosome number unknown.

Monospecific: *Duparquetia orchidacea* Baill. Distributed in humid tropical forests of West and Central Africa.

Subfam. **Dialioideae** Legume Phylogeny Working Group, **stat. nov.** ≡ *Dialiinae* H.S.Irwin & Barneby in Polhill & Raven, Adv. Legume Syst. 1: 100. 1981 – Type: *Dialium* L.

Unarmed trees or shrubs, rarely suffruticose (*Labichea* Gaudich. ex DC., *Petalostylis* R.Br.) (Fig. 3H–L); specialised extrafloral nectaries lacking on petiole and leaf rachis and on leaflet surface. Stipules in lateral position, free or absent. **Leaves imparipinnate**, rarely paripinnate (*Eligmocarpus* Capuron, *Poeppigia* C.Presl), 1-foliolate (*Baudouinia* Baill., *Labichea*, *Mendoravia* Capuron, *Uittienia* Steenis) or palmately compound (*Labichea*), **leaflets alternate**, rarely opposite (*Eligmocarpus*, *Poeppigia*), exstipellate. **Inflorescences highly branched, thyrsoid**, less commonly racemes with distichous anthotaxy (*Labichea*, *Petalostylis*), borne in both terminal and axillary positions, or reduced to one axillary flower (*Petalostylis*); bracteoles small or absent. Flowers bisexual (polygamous in *Apuleia* Mart.), radially or slightly to strongly bilaterally symmetrical, **hypanthium rarely present**, receptacle may be broad and flattened, bearing nectary-like bodies; sepals commonly 5, reduced to 4 (*Labichea*, *Storckiella* Seem.) or 3 (*Apuleia*, *Dialium*), rarely 6 (*Mendoravia*), free, equal to sub-equal; petals 5 or fewer (0, 1, 3, 4), rarely 6 (petal number often equivalent to sepal number), free, equal to sub-equal, imbricate, the adaxial petal innermost; **fertile stamens 5 or fewer**, rarely 6–10 (some *Dialium* spp., *Poeppigia*), **usually only antesealous whorl present**, free, uniform, rarely dimorphic (*Eligmocarpus*), **anthers basifixed**, rarely dorsifixed (*Poeppigia*), dehiscing via longitudinal slits, **often reduced to a short apical, poricidal slit**, staminodes present or absent; pollen in tricolporate monads with punctate or finely reticulate, rarely striate (some *Dialium*) sculpture patterns; gynoecium 1-carpellate (sometimes bicarpellate in scattered flowers of *Dialium*), ovary stipitate or sessile, ovules frequently 2 (1–many). **Fruits commonly indehiscent drupaceous or samaroid**, rarely dehiscent (*Eligmocarpus*, *Labichea*, *Mendoravia*, *Petalostylis*) or the drupaceous fruit with indurating endocarp breaking up in one seeded segments (*Baudouinia*). Seeds 1–2, rarely more; embryo straight. Vestured pits absent in the secondary xylem, rarely present

(*Poeppegia*, *Mendoravia*); silica bodies sometimes present (*Apuleia*, *Dialium*, *Dicorynia* Benth., *Distemonanthus* Benth.); septate fibres rarely present (*Apuleia*, *Distemonanthus*, *Poeppegia*); storeyed rays often present. Root nodules absent. $2n = 28$ (most genera unsurveyed).

Currently 17 genera and ca. 85 species. Widespread throughout the tropics, with taxa occurring in South, Central and North America, Africa, Madagascar, South and Southeast Asia, south China, Australia, New Guinea and some Pacific islands.

Clade-based definition (included taxa): The most inclusive crown clade containing *Poeppegia procera* and *Dialium guianense* (Aubl.) Sandwith, but not *Cercis canadensis*, *Duparquetia orchidacea*, or *Bobgunnia fistuloides*.

Subfam. **Caesalpinioideae** DC., Prodr. 2: 473. 1825 – Type: *Caesalpinia* L.

= Mimosoideae DC., Prodr. 2: 424. 1825 – Type: *Mimosa* L.

= Cassioideae Burmeist., Handb. Naturgesch.: 319. 1837 (“Cassieae”) – Type: *Cassia* L., nom. cons.

Trees, shrubs, lianas, suffruticose or functionally herbaceous, occasionally aquatic (Figs. 5 & 6), either unarmed or commonly armed with prickles or nodal or infranodal spines; **specialised extrafloral nectaries often present on the petiole and/or on the primary and secondary rachises, usually between pinnae or leaflet pairs**, more rarely stipular or bracteal (*Senna* Mill., *Macrosamanea* Britton & Rose ex Britton & Killip and some *Archidendron* F.Muell.). Stipules in lateral position and free or absent. Leaves usually pulvinate, **commonly bipinnate**, otherwise pinnate (sometimes both types on the same plant in *Arcoa* Urb., *Cenostigma* Tul., *Gleditsia* L., *Stuhlmannia* Taub., rarely in *Ceratonia* L. and *Moldenhawera* Schrad.) and then mostly paripinnate, rarely imparipinnate, less often bifoliolate, modified into phyllodes or lacking, arrangement of the pinnae and leaflets mostly opposite, rarely alternate; stipels rare and not to be confused with the more commonly present paraphyllidia. **Inflorescences globose, spicate**, paniculate, racemose or in fascicles; bracteoles commonly absent or small. Flowers usually bisexual, rarely unisexual (*Ceratonia*, *Gleditsia* and *Gymnocladus* Lam., species dioecious or monoecious), or bisexual flowers combined with unisexual and/or sterile flowers in heteromorphic inflorescences (mimosoid clade), radially, less frequently bilaterally symmetrical or asymmetric, hypanthium lacking or cupular, rarely tubular; sepals (3–)5(–6), free or fused; petals (3–)5(–6), free or fused (the sepal or petal or both whorls sometimes lacking), **aestivation valvate** (mimosoid clade) or imbricate and then the adaxial petal innermost; stamens commonly diplostemonous or haplostemonous, sometimes reduced to 3, 4 or 5 (in some *Mimosa* spp.), frequently many (100+ in some mimosoids), free or fused, sometimes heteromorphic, some or all sometimes modified or staminodial, anthers basifixed or dorsifixed, **often with a stipitate or sessile apical gland**, dehiscing via longitudinal slits or apical or basal poricidal slits or pores; pollen in tricolporate monads, or **commonly in tetrads, bitetrads or polyads** (most mimosoids); gynoeceum uni- or rarely polycarpellate, 1–many-ovulate. Fruit a thin-valved,

1–many-seeded pod, dehiscent along one or both sutures, also often a lomentum, a craspedium, or thick and woody and then indehiscent or explosively dehiscent, often curved or spirally coiled. **Seeds usually with an open or closed pleurogram on both faces**, sometimes with a fleshy aril (*Pithecellobium* Mart. and some *Acacia* Mill.) or sarcotesta (*Inga* Mill.), sometimes winged; hilum usually apical, lens usually inconspicuous; embryo straight. Vestured pits present in secondary xylem; silica bodies sometimes present (*Tachigali* Aubl., *Diptychandra* Tul.); septate fibres and storeyed rays sometimes present. **Root nodules variably present and indeterminate** (prevalent in the mimosoid clade). $2n$ mostly 24, 26, 28, but also reported $2n = 14, 16, 52, 54, 56$. Non-protein amino acids frequently reported, for example mimosine, albizine (mimosoids), djenkolic acid, pipelicolic acid and its derivatives; coumarins, cyanogenic glucosides, phenylethylamines, tryptamines, and β -carboline alkaloids also reported.

Caesalpinioideae in its emended circumscription contains 148 genera and ca. 4400 species. Pantropical, common in both wet and dry regions, with a handful of species extending to the temperate zone, less frequently frost-tolerant (*Gleditsia*, *Gymnocladus* and some species of *Desmanthus* Willd. and *Senna*).

This clade was referred to as the MCC clade (Doyle, 2011, 2012) or GCM-clade (Marazzi & al., 2012).

Clade-based definition (included taxa): The most inclusive crown clade containing *Arcoa gonavensis* Urb. and *Mimosa pudica* L., but not *Bobgunnia fistuloides*, *Duparquetia orchidacea*, or *Poeppegia procera*.

Subfam. **Papilionoideae** DC., Prodr. 2: 94. 1825 \equiv Faboideae Rudd in *Rhodora* 70: 496. 1968 – Type: *Faba* Mill., (\equiv *Vicia* L.).

= Lotoideae Burnett, *Outlines Bot.*: 643. 1835 (“Lotidae”) – Type: *Lotus* L.

Mostly unarmed trees, shrubs, lianas, herbs, and twining or tendriled vines (Figs. 7–9); specialised extrafloral nectaries lacking on petiole and leaf rachis, occasionally stipular, stipellar or bracteal nectaries, or swollen and nectar-secreting peduncles, rarely on sepals (*Erythrina* L.). Stipules in lateral position (very rarely interpetiolar, in all species of *Platymiscium* Vogel), free or absent. Leaves mostly pari- or imparipinnate to palmately compound, also commonly uni- or **trifoliolate**, rarely bi- or tetrafoliolate, never bipinnate (palmately-pinnate in *Rhynchosia ferulifolia* Benth. ex Harv.), either pulvinate or not, leaflets opposite or alternate, sometimes modified into tendrils, rarely phyllodinous, stipels present or absent. Inflorescence mostly racemose, pseudoracemose or paniculate, less often cymose, spicate or capitate, axillary or terminal, or flowers solitary; bracteoles usually present, rarely enlarged, valvate, enveloping bud. Flowers bisexual, rarely unisexual, usually bilaterally symmetrical, rarely asymmetrical, radially symmetrical or nearly so, rarely cleistogamous flowers also present; hypanthium present or absent; sepals (3–)5, united at least at the base, sometimes the calyx entire and splitting into irregular lobes or the calyx lobes dimorphic and some petaloid; petals (0–)5(–6) and then imbricate, **corolla mostly**

papilionate, with the adaxial petal (= standard) outermost and largest, usually overlapping lateral wing petals which in turn overlap the abaxial keel petals or, in radially symmetrical flowered species, corolla with 5 small or undifferentiated petals, less often only one (standard) petal is present or all petals absent; stamens typically 10, rarely 9 or many, **filaments most commonly connate into a sheath or tube, or uppermost filament wholly or partly free**, sometimes all filaments free, anthers uniform or dimorphic, basifixed or dorsifixed, dehiscing longitudinally; pollen in monads, mostly 3-colpate, 3-colpate or 3-porate; gynoecium 1-carpellate, very rarely 2-carpellate, 1–many-ovuled. Fruit a 1–many seeded pod, dehiscing along one or both sutures, or indehiscent, or a loment, samara or drupe. Seeds usually with a hard testa, rarely overgrown, sometimes with a fleshy aril or sarcotesta, a **complex hilar valve, elongate hilum and lens usually present**, pleurogram absent; **embryo usually curved**, rarely straight. Vestured pits present in secondary xylem; silica bodies absent; septate fibres sometimes present; all elements (vessels, parenchyma, strands rays) usually in storeyed structure. **Root nodules generally present, either indeterminate or determinate**. $2n =$ more commonly 16, 18, 20, 22 but other numbers also reported ($2n = 12, 14, 24, 26, 28, 30, 32, 38, 40, 48, 64, 84$). Isoflavonoids, prenylated flavonoids, indolizidine or quinolizidine alkaloids reported. Non-protein amino acids widespread, some exclusively found in the subfamily (e.g., canavanine).

Currently 503 genera and ca. 14,000 species, nearly cosmopolitan.

Clade-based definition (included taxa): The most inclusive crown clade containing *Castanospermum australe* A.Cunn. ex Mudie and *Vicia faba* L., but not *Erythrostemon gilliesii* (Hook.) Klotzsch., *Gleditsia triacanthos* L., or *Dialium guianense*. For ICPN classification of particular Papilionoideae clades see Wojciechowski (2013).

The mimosoid clade

Although the mimosoid clade (Fig. 6) is not formally named here, it is morphologically distinct and can be defined as the most inclusive crown clade containing all Leguminosae with radially symmetrical flowers having valvate petal aestivation, homologous to those found in *Pentaclethra macrophylla* Benth. and *Inga edulis* Mart.

The mimosoid clade contains all genera previously assigned to subfamily Mimosoideae plus *Chidlowia*, previously considered to be a member of the former Caesalpinioideae, but now shown to belong to the mimosoid clade (Manzanilla & Bruneau, 2012; E.J.M. Koenen & al., in prep.). This clade of 3300+ species is morphologically highly distinctive with radially symmetrical flowers with valvate aestivation of the calyx and corolla (except in *Parkia*, which has partially imbricate calyx lobes). Typically, flowers are combined in spicate or capitate inflorescences, often these are in turn combined into compound inflorescences (e.g., a panicle of globose heads). Pantropical, common in both wet and dry regions, with a handful of species extending to the temperate zone, and less frequently into frost-prone regions.

ACKNOWLEDGEMENTS

This project was made possible thanks to ongoing collaboration with researchers worldwide who have collected samples of legumes for DNA studies and who have shared material with members of the legume systematics community for over three decades. In particular, we thank the following people for their help with collection, preparation and curation of specimens, and/or with sequencing of the *matK* locus: Alexandra Clark, Michelle Hart (Royal Botanic Gardens, Edinburgh); Adilson M. Pintor, Marcelo T. Nascimento, Pablo Prieto (Rio de Janeiro); Aécio A. Santos (Goiás, Tocantins); Antônio S.L. da Silva, Camilo Barbosa, Catarina S. Carvalho, Leandro V. Ferreira, Lisandra A. Teixeira, Nara Mota, Pedro L. Viana, Rafael Salomão (Pará); Caio V. Vivas, José Lima Paixão, Tim Baker (Bahia); Claudio Nicoletti, Geovane S. Siqueira (Espírito Santo); Eric Hattori, Fernanda S. Freitas, Flávia Pezzini, Pedro Taucce (Minas Gerais); Marcella Baroni (Mato Grosso do Sul); Flávia Costa (Manaus); Luzmilla Arroyo, Daniel Villaroel, Alexander Germaine Parada (Bolivia); Aniceto Daza, Jose Luis Marcelo-Peña, Reynaldo Linares Palomino, Carlos Reynel, Isau Huamantupa (Peru); Tanja Schuster, Mansi Trivedi, Gabe Johnson, William Cagle, Ailsa Holland, and Xin-Fen Gao (Smithsonian Institution). We thank the ScienceCloud at the University of Zurich for computational resources. Funding for this project was provided by the Natural Sciences and Engineering Research Council of Canada, the U.K. Natural Environment Research Council (grant NE/I028122/1), the Swiss National Science Foundation (grant 31003A_13552), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/Program POS CSF # 1951/13-0), Conselho Nacional de Desenvolvimento Científico e Tecnológico (Project Sisbiota 563084/2010-3 and Project Casadinho/Procad # 5525892011-0), Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB PES 0053/2011) and the Fundação de Amparo à Pesquisa do Estado de São Paulo FAPESP; Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq of Brazil, the Smithsonian Institution, the U.S. National Science Foundation (grant DEB-1352217), and the Environment Research and Technology Development Fund (S9) of the Ministry of the Environment of Japan, and Arizona State University. Anne Bruneau acknowledges the Royal Botanic Gardens, Edinburgh, and the Department of Systematic and Evolutionary Botany, University of Zurich for logistical support during a sabbatical in 2015. Finally, we thank Jonathan Amith, Xander van der Burgt, Emilio Constantino, David Du Puy, Flora do Acre, Felix Forest, Paul Hoekstra, Mike Hopkins, Rosangela Melo, Justin Moat, Projecto Flora Reserva Ducke INPA/DFID, Shijin Li, Alex Popovkin, James Ratter, Wolfgang Stuppy, Liam Trethowan, Timothy Utteridge, and André van Proosdij for contributing excellent legume images to Figs. 3–11. The authors thank Michael Pirie, Lars Chatrou, Jefferson Prado, John McNeill, and two anonymous reviewers for advice on nomenclatural issues and comments on the manuscript.

LITERATURE CITED

- Adanson, M. 1763. *Familles des plantes*, vol. 2. Paris: chez Vincent. <https://doi.org/10.5962/bhl.title.271>
- APG (Angiosperm Phylogeny Group) III 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161: 105–121. <https://doi.org/10.1111/j.1095-8339.2009.00996.x>

- Bello, M.A., Bruneau, A., Forest, F. & Hawkins, J.A.** 2009. Elusive relationships within order Fabales: Phylogenetic analyses using *matK* and *rbcL* sequence data. *Syst. Bot.* 34: 102–114. <https://doi.org/10.1600/036364409787602348>
- Bello, M.A., Rudall, P.J. & Hawkins, J.A.** 2012. Combined phylogenetic analyses reveal interfamilial relationships and patterns of floral evolution in the eudicot order Fabales. *Cladistics* 28: 393–421. <https://doi.org/10.1111/j.1096-0031.2012.00392.x>
- Bentham, G.** 1865. Leguminosae. Pp. 434–600 in: Bentham, G. & Hooker, J.D. (eds.), *Genera plantarum*, vol. 1(2). Londini [London]: venit apud Lovell Reeve. <https://doi.org/10.5962/bhl.title.747>
- Bouchenak-Khelladi, Y., Maurin, O., Hurter, J. & Van der Bank, M.** 2010. The evolutionary history and biogeography of Mimosoideae (Leguminosae): An emphasis on African acacias. *Molec. Phylogen. Evol.* 57: 495–508. <https://doi.org/10.1016/j.ympev.2010.07.019>
- BFG – The Brazil Flora Group** 2015. Growing knowledge: An overview of seed plant diversity in Brazil. *Rodriguésia* 66: 1085–1113. <https://doi.org/10.1590/2175-7860201566411>
- Bruneau, A., Forest, F., Herendeen, P.S., Klitgaard, B.B. & Lewis, G.P.** 2001. Phylogenetic relationships in the Caesalpinioideae (Leguminosae) as inferred from chloroplast *trnL* intron sequences. *Syst. Bot.* 26: 487–514. <http://www.jstor.org/stable/3093978>
- Bruneau, A., Mercure, M., Lewis, G.P. & Herendeen, P.S.** 2008. Phylogenetic patterns and diversification in the caesalpinioideae legumes. *Botany* 86: 697–718. <https://doi.org/10.1139/B08-058>
- Bruneau, A., Klitgaard, B.B., Prenner, G., Fougère-Danezan, M. & Tucker, S.C.** 2014. Floral evolution in the Detarioideae (Leguminosae): Phylogenetic evidence for labile floral development in an early diverging legume lineage. *Int. J. Pl. Sci.* 175: 392–417. <https://doi.org/10.1086/675574>
- Candolle, A. de** 1825. *Prodromus systematis naturalis regni vegetabilis*, vol. 2. Parisiis [Paris]: sumptibus sociorum Treuttel et Würtz. <https://doi.org/10.5962/bhl.title.286>
- Cannon, S.B., Hut, D., Farmer, A.D., Maki, S.L., May, G.D., Singer, S.R. & Doyle, J.J.** 2010. Polyploidy did not predate the evolution of nodulation in all legumes. *PLoS ONE* 5(7): e11630. <https://doi.org/10.1371/journal.pone.0011630>
- Cannon, S.B., McKain, M.R., Harkess, A., Nelson, M.N., Dash, S., Deyholos, M.K., Peng, Y., Joyce, B., Stewart, C.N., Rolf, M., Kutchan, T., Tan, X., Chen, C., Zhang, Y., Carpenter, E., Wong, G.K.S., Doyle, J.J. & Leebens-Mack, J.** 2015. Multiple polyploidy events in the early radiation of nodulating and non-nodulating legumes. *Molec. Biol. Evol.* 32: 193–210. <https://doi.org/10.1093/molbev/msu296>
- Cantino, P.D. & Queiroz, K. de** 2010. PhyloCode: International Code of Phylogenetic Nomenclature, version 4c. <http://www.ohio.edu/phylocode/>
- Cardoso, D., Queiroz, L.P. de, Pennington, R.T., Lima, H.C. de, Fonty, E., Wojciechowski, M.F. & Lavin, M.** 2012a. Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively sampled early-branching lineages. *Amer. J. Bot.* 99: 1991–2013. <https://doi.org/10.3732/ajb.1200380>
- Cardoso, D., Lima, H.C. de, Rodrigues, R.S., Queiroz, L.P. de, Pennington, R.T. & Lavin, M.** 2012b. The realignment of *Acosmium* sensu stricto with the Dalbergioid clade (Leguminosae, Papilionoideae) reveals a proneness for independent evolution of radial floral symmetry among early branching papilionoid legumes. *Taxon* 61: 1057–1073.
- Cardoso, D., Queiroz, L.P. de, Lima, H.C. de, Suganuma, E., Van den Berg, C. & Lavin, M.** 2013a. A molecular phylogeny of the vataireoid legumes underscores floral evolvability that is general to many early-branching papilionoid lineages. *Amer. J. Bot.* 100: 403–421. <https://doi.org/10.3732/ajb.1200276>
- Cardoso, D., Pennington, R.T., Queiroz, L.P. de, Boatwright, J.S., Van Wyk, B.-E., Wojciechowski, M.F. & Lavin, M.** 2013b. Reconstructing the deep-branching relationships of the papilionoid legumes. *S. African J. Bot.* 89: 58–75. <https://doi.org/10.1016/j.sajb.2013.05.001>
- Champagne, C.E.M., Goliber, T.E., Wojciechowski, M.F., Mei, R.W., Townsley, B.T., Wang, K., Paz, M.M., Geeta, R. & Sinha, N.R.** 2007. Compound leaf development and evolution in the legumes. *Pl. Cell* 19: 3369–3378. <https://doi.org/10.1105/tpc.107.052886>
- Corner, E.J.H.** 1976. *The seeds of dicotyledons*. Cambridge: Cambridge University Press.
- Cronquist, A.** 1981. *An integrated system of classification of flowering plants*. New York: Columbia University Press.
- Dickson, W.C.** 1981. The evolutionary relationships of the Leguminosae. Pp. 35–54 in: Polhill, R.M. & Raven, P.H. (eds.), *Advances in legume systematics*, part 1. Richmond, U.K.: Royal Botanic Gardens, Kew.
- Doyle, J.J.** 2011. Phylogenetic perspectives on the origins of nodulation. *Molec. Pl.-Microbe Interact.* 24: 1289–1295. <https://doi.org/10.1094/MPMI-05-11-0114>
- Doyle, J.J.** 2012. Polyploidy in legumes. Pp 147–180 in: Soltis, P.S. & Soltis, D.E. (eds.), *Polyploidy and genome evolution*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-31442-1_9
- Doyle, J.J.** 2016. Chasing unicorns: Nodulation origins and the paradox of novelty. *Amer. J. Bot.* 103: 1865–1868. <https://doi.org/10.3732/ajb.1600260>
- Doyle, J.J., Doyle, J.L., Ballenger, J.A., Dickson, E.E., Kajita, T. & Ohashi, H.** 1997. A phylogeny of the chloroplast gene *rbcL* in the Leguminosae: Taxonomic correlations and insights into the evolution of nodulation. *Amer. J. Bot.* 84: 541–554. <https://doi.org/10.2307/2446030>
- Doyle, J.J., Chappill, J.A., Bailey, C.D. & Kajita, T.** 2000. Towards a comprehensive phylogeny of legumes: Evidence from *rbcL* sequences and non-molecular data. Pp. 1–20, in: Herendeen, P.S. & Bruneau, A. (eds.), *Advances in legume systematics*, part 9. Richmond, U.K.: Royal Botanic Gardens, Kew.
- DRYFLOR** 2016. Plant diversity patterns in neotropical dry forests and their conservation implications. *Science* 353: 1383–1387. <https://doi.org/10.1126/science.aaf5080>
- Dugas, D.V., Hernandez, D., Koenen, E.J.M., Schwarz, E., Straub, S., Hughes, C.E., Jansen, R.K., Nageswara-Rao, M., Staats, M., Trujillo, J.T., Hajrah, N.H., Alharbi, N.S., Al-Malki, A.L., Sabir, J.S.M. & Bailey, C.D.** 2015. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. *Sci. Rep.* 5: 16958. <https://doi.org/10.1038/srep16958>
- Fougère-Danezan, M., Maumont, S. & Bruneau, A.** 2007. Relationships among resin producing Detarioideae s.l. (Leguminosae) as inferred by molecular data. *Syst. Bot.* 32: 748–761. <https://doi.org/10.1600/036364407783390755>
- Funk, V.A., Susanna, A., Stuessy, T.F. & Bayer, R.J.** 2009. *Systematics, evolution, and biogeography of Compositae*. Vienna: International Association for Plant Taxonomy.
- Gepts, P., Beavis, W.D., Brummer, E.C., Shoemaker, R.C., Stalker, H.T., Weeden, N.F. & Young, N.D.** 2005. Legumes as a model plant family: Genomics for food and feed report of the cross-legume advances through genomics conference. *Pl. Physiol.* 137: 1228–1235. <https://doi.org/10.1104/pp.105.060871>
- Görts-van Rijn, A.R.A. (ed.)** 1989. *Flora of the Guianas*. Königstein: Koeltz Scientific Books.
- Grass Phylogeny Working Group** 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Ann. Missouri Bot. Gard.* 88: 373–457. <https://doi.org/10.2307/3298585>
- Grass Phylogeny Working Group II** 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytol.* 193: 304–312. <https://doi.org/10.1111/j.1469-8137.2011.03972.x>
- Hancock, J.F.** 2012. *Plant evolution and the origin of crop species*, ed. 3. Wallingford, U.K. & Cambridge, U.S.A.: CAB International. <https://doi.org/10.1079/9781845938017.0000>

- Herendeen, P.S., Lewis, G.P. & Bruneau, A. 2003. Floral morphology in caesalpinoid legumes: Testing the monophyly of the “*Umtiza*” clade. *Int. J. Pl. Sci.* 164: S393–S407. <https://doi.org/10.1086/376881>
- Huson, D.H. & Scornavacca, C. 2012. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* 61: 1061–1067. <https://doi.org/10.1093/sysbio/sys062>
- Hutchinson, J. 1926. *The families of flowering plants*, vol. 1, *Dicotyledons*. London: MacMillan.
- Hutchinson, J. 1964. *The genera of flowering plants (Angiospermae)*, vol. 1. Oxford: Oxford University Press.
- Hutchinson, J. & Dalziel, J.M. 1928. *Flora of tropical West Africa*, vol. 1(2). London: Crown Agents for the Colonies.
- Irwin, H.S. 1981. Preface. Pp. vii–xi in: Polhill, R.M. & Raven, P.H. (eds.), *Advances in legume systematics*, part 1. Richmond, U.K.: Royal Botanic Gardens, Kew.
- Jussieu, A.L. de 1789. *Genera plantarum*. Parisii [Paris]: apud viduam Herissant et Theophilum Barrois. <https://doi.org/10.5962/bhl.title.7762>
- Kajita, T., Ohashi, H., Tateishi, Y., Bailey, C.D. & Doyle, J.J. 2001. *rbcl* and legume phylogeny with particular reference to Phaseoleae, Millettieae and allies. *Syst. Bot.* 26: 515–536.
- Käss, E. & Wink, M. 1996. Molecular evolution of the Leguminosae: Phylogeny of the three subfamilies based on *rbcl*-sequences. *Biochem. Syst. Ecol.* 24: 365–378. [https://doi.org/10.1016/0305-1978\(96\)00032-4](https://doi.org/10.1016/0305-1978(96)00032-4)
- Katoh, K. & Standley, D.M. 2013. MAFFT: Multiple sequence alignment software version 7: Improvements in performance and usability. *Molec. Biol. Evol.* 30: 772–780. <https://doi.org/10.1093/molbev/mst010>
- Koenen, E.J.M., De Vos, J.M., Atchison, G.W., Simon, M.F., Schrire, B.D., Souza, E.R. de, Queiroz, L.P. de & Hughes, C.E. 2013. Exploring the tempo of species diversification in legumes. *S. African J. Bot.* 89: 19–30. <https://doi.org/10.1016/j.sajb.2013.07.005>
- Lanfear, R., Calcott, B., Ho, S.Y.W. & Guindon, S. 2012. Partition-Finder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molec. Biol. Evol.* 29: 1695–1701. <https://doi.org/10.1093/molbev/mss020>
- Lartillot, N., Lepage, T. & Blanquart, S. 2009. PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25: 2286–2288. <https://doi.org/10.1093/bioinformatics/btp368>
- Lavin, M., Pennington, R.T., Klitgaard, B.B., Sprent, J.I., Lima, H.C. de & Gasson, P.E. 2001. The Dalbergioid legumes (Fabaceae): Delimitation of a pantropical monophyletic clade. *Amer. J. Bot.* 88: 503–533. <https://doi.org/10.2307/2657116>
- Lavin, M., Schrire, B., Lewis, G., Pennington, R.T., Delgado-Salinas, A., Thulin, M., Hughes, C., Beyra-Matos, A. & Wojciechowski, M.F. 2004. Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. *Philos. Trans., Ser. B* 359: 1509–1522.
- Lavin, M., Herendeen, P.S. & Wojciechowski, M.F. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Syst. Biol.* 54: 575–594. <https://doi.org/10.1080/10635150590947131>
- Lewis, G.P. & Forest, F. 2005. Cercideae. Pp. 57–67 in: Lewis, G., Schrire, B., Mackinder, B. & Lock, M. (eds.), *Legumes of the World*. Richmond, U.K.: Royal Botanic Gardens, Kew.
- Lewis, G.P. & Schrire, B.D. 2003. Leguminosae or Fabaceae? Pp. 1–3 in: Klitgaard, B. & Bruneau, A. (eds.), *Advances in legume systematics*, part 10, *Higher level systematics*. Richmond, U.K.: Royal Botanic Gardens, Kew.
- Lewis, G., Schrire B., Mackinder, B. & Lock, M. (eds.) 2005. *Legumes of the World*. Richmond, U.K.: Royal Botanic Gardens, Kew.
- Lewis, G.P., Schrire, B.D., Mackinder, B.A., Rico, L. & Clark, R. 2013. A 2013 linear sequence of legume genera set in a phylogenetic context: A tool for collections management and taxon sampling. *S. African J. Bot.* 89: 76–84. <https://doi.org/10.1016/j.sajb.2013.06.005>
- Li, H.L., Wang, W., Li, R.Q., Zhang, J.B., Sun, M., Naem, R., Su, J.X., Xiang, X.G., Mortimer, P.E., Li, D.Z., Hyde, K.D., Xu, J.C., Soltis, D.E., Soltis, P.S., Li, J.H., Zhang, S.Z., Wu, H., Chen, Z.D. & Lu, A.M. 2016. Global versus Chinese perspectives on the phylogeny of the N-fixing clade. *J. Syst. Evol.* 54: 392–399. <https://doi.org/10.1111/jse.12201>
- LPWG, Legume Phylogeny Working Group 2013a. Legume phylogeny and classification in the 21st century: Progress, prospects and lessons for other species-rich clades. *Taxon* 62: 217–248. <https://doi.org/10.12705/622.8>
- LPWG, Legume Phylogeny Working Group 2013b. Towards a new classification system for legumes: Progress report from the 6th International Legume Conference. *S. African J. Bot.* 89: 3–9.
- Luckow, M., Miller, J.T., Murphy, D.J. & Livshultz, T. 2003. A phylogenetic analysis of the Mimosoideae (Leguminosae) based on chloroplast DNA sequence data. Pp. 197–220 in: Klitgaard, B.B. & Bruneau, A. (eds.), *Advances in legume systematics*, part 10, *Higher level systematics*. Richmond, U.K.: Royal Botanic Gardens, Kew.
- Manzanilla, V. & Bruneau, A. 2012. Phylogeny reconstruction in the Caesalpinieae grade (Leguminosae) based on duplicated copies of the sucrose synthase gene and plastid markers. *Molec. Phylog. Evol.* 65: 149–162. <https://doi.org/10.1016/j.jmpev.2012.05.035>
- Marazzi, B., Ané, C., Simon, M.F., Delgado-Salinas, A., Luckow, M. & Sanderson, M.J. 2012. Locating evolutionary precursors on a phylogenetic tree. *Evolution* 66: 3918–3930. <https://doi.org/10.1111/j.1558-5646.2012.01720.x>
- McMahon, M.M. & Sanderson, M.J. 2006. Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. *Syst. Biol.* 55: 818–836. <https://doi.org/10.1080/10635150600999150>
- McNeill, J., Barrie, F.R., Buck, W.R., Demoulin, V., Greuter, W., Hawksworth, D.L., Herendeen, P.S., Knapp, S., Marhold, K., Prado, J., Prud'homme van Reine, W.F., Smith, G.F., Wiersema, J.H. & Turland, N.J. (eds.) 2012. *International Code of Nomenclature for algae, fungi, and plants (Melbourne Code)*. Regnum Vegetabile 154. Königstein: Koeltz Scientific Books.
- Miller, J.T., Seigler, D. & Mishler, B.D. 2014. A phylogenetic solution to the *Acacia* problem. *Taxon* 63: 653–658. <https://doi.org/10.12705/633.2>
- Miller, M.A., Pfeiffer, W. & Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Pp. 45–52 in: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, Louisiana, 14 Nov 2010. Piscataway: IEEE. <https://doi.org/10.1109/GCE.2010.5676129>
- Moncrieff, G.R., Lehmann, C.E.R., Schnitzler, J., Gambiza, J., Hiernaux, P., Ryan, C.M., Shackleton, C.M., Williams, R.J. & Higgins, S.I. 2014. Contrasting architecture of key African and Australian savanna tree taxa drives intercontinental structural divergence. *Global Ecol. Biogeogr.* 23: 1235–1244. <https://doi.org/10.1111/geb.12205>
- Mori, S.A., Cremers, G., Gracic, C.A., Granville, J.-J. de, Heald, S.V., Hoff, M. & Mitchell, J.D. 2002. *Guide to the vascular plants of central French Guiana*, part 2, *Dicotyledons*. Memoirs of the New York Botanical Garden 76. Bronx: New York Botanical Garden Press.
- Oliveira-Filho, A.T., Cardoso, D., Schrire, B.D., Lewis, G.P., Pennington, R.T., Brummer, T.J., Rotella, J. & Lavin, M. 2013. Stability structures tropical woody plant diversity more than seasonality: Insights into the ecology of high legume-succulent-plant diversity. *S. African J. Bot.* 89: 42–57. <https://doi.org/10.1016/j.sajb.2013.06.010>
- Orchard, A.E. & Wilson, A.J.G. (eds.) 1998–2001. *Flora of Australia*, vols. 11–15. Melbourne: CSIRO.
- Panero, J.L. & Funk, V.A. 2002. Toward a phylogenetic subfamilial classification for the Compositae (Asteraceae). *Proc. Biol. Soc. Wash.* 115: 909–922.
- Pennington, R.T., Klitgaard, B.B., Ireland, H. & Lavin, M. 2000.

- New insights into floral evolution of basal Papilionoideae from molecular phylogenies. Pp. 233–248 in: Herendeen, P.S. & Bruneau, A. (eds.), *Advances in legume systematics*, part 9. Richmond, U.K.: Royal Botanic Gardens, Kew.
- Pennington, R.T., Lavin, M., Särkinen, T., Lewis, G.P., Klitgaard, B.B. & Hughes, C.E.** 2010. Contrasting plant diversification histories within the Andean biodiversity hotspot. *Proc. Natl. Acad. Sci. U.S.A.* 107: 13783–13787. <https://doi.org/10.1073/pnas.1001317107>
- Polhill, R.M.** 1994. Pp. xxxv–lviii in: Bisby, F.A., Buckingham, J. & Harborne, J.B. (eds.), *Phytochemical dictionary of the Leguminosae*, vol. 1, *Plants and their constituents*. London: Chapman and Hall.
- Polhill, R.M. & Raven, P.H. (eds.)** 1981. *Advances in legume systematics*, part 1. Richmond, U.K.: Royal Botanic Gardens, Kew.
- Polhill, R.M., Raven, P.H. & Stirton, C.H.** 1981. Evolution and systematics of the Leguminosae. Pp. 1–26 in: Polhill, R.M. & Raven, P.H. (eds.), *Advances in legume systematics*, part 1. Richmond, U.K.: Royal Botanic Gardens, Kew.
- Rambaut, A., Suchard, M.A., Xie, D. & Drummond, A.J.** 2014. Tracer, version 1.6. Available from <http://beast.bio.ed.ac.uk/Tracer>
- Ramos, G., Lima, H.C. de, Prenner, G., Queiroz, L.P. de, Zartman, C.E. & Cardoso, D.** 2016. Molecular systematics of the Amazonian genus *Aldina*, a phylogenetically enigmatic ectomycorrhizal lineage of papilionoid legumes. *Molec. Phylogen. Evol.* 97: 11–18. <https://doi.org/10.1016/j.ympev.2015.12.017>
- Ranwez, V., Harispe, S., Delsuc, F. & Douzery, E.J.P.** 2011. MACSE: Multiple alignment of coding sequences accounting for frameshifts and stop codons. *PLoS ONE* 6(9): e22594. <https://doi.org/10.1371/journal.pone.0022594>
- Schrire, B.D., Lavin, M. & Lewis, G.P.** 2005a. Global distribution patterns of the Leguminosae: Insights from recent phylogenies. *Biol. Skr.* 55: 375–422.
- Schrire, B.D., Lewis, G.P. & Lavin, M.** 2005b. Biogeography of the Leguminosae. Pp. 21–54 in: Lewis, G.P., Schrire, B., MacKinder, B. & Lock, M. (eds.), *Legumes of the World*. Richmond, U.K.: Royal Botanic Gardens, Kew.
- Simon, M.F. & Pennington, R.T.** 2012. Evidence for adaptation to fire regimes in the tropical savannas of the Brazilian Cerrado. *Int. J. Pl. Sci.* 173: 711–723. <https://doi.org/10.1086/665973>
- Simon, M.F., Grether, R., Queiroz, L.P. de, Skema, C., Pennington, R.T. & Hughes, C.E.** 2009. Recent assembly of the Cerrado, a Neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proc. Natl. Acad. Sci. U.S.A.* 106: 20359–20364. <https://doi.org/10.1073/pnas.0903410106>
- Sinou, C., Forest, F., Lewis, G.P. & Bruneau, A.** 2009. The genus *Bauhinia* s.l. (Leguminosae): A phylogeny based on the plastid *trnL-trnF* region. *Botany* 87: 947–960. <https://doi.org/10.1139/B09-065>
- Sprent, J.I.** 2007. Evolving ideas of legume evolution and diversity: A taxonomic perspective on the occurrence of nodulation. *New Phytol.* 174: 11–25. <https://doi.org/10.1111/j.1469-8137.2007.02015.x>
- Sprent, J.I.** 2009. *Legume nodulation: A global perspective*. West Sussex: Wiley-Blackwell. <https://doi.org/10.1002/9781444316384>
- Stamatakis, A.** 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stirton, C.H. & Zarucchi, J.L. (eds.)** 1989. *Advances in legume biology*. Monographs in Systematic Botany of the Missouri Botanical Garden 29. St. Louis: Missouri Botanical Garden Press.
- Sun, M., Naeem, R., Su, J.-X., Cao, Z.-Y., Burleigh, J.G., Soltis, P.S., Soltis, D.E. & Chen, Z.-D.** 2016. Phylogeny of Rosidae: A dense taxon sampling analysis. *J. Syst. Evol.* 54: 363–391. <https://doi.org/10.1111/jse.12211>
- Taubert, P.** 1891. Leguminosae. Pp. 70–396 in: Engler, A. & Prantl, K. (eds.), *Die natürlichen Pflanzenfamilien*, vol. 3(3). Leipzig: Engelmann.
- Tucker, S.C.** 2003. Floral development in legumes. *Pl. Physiol.* 131: 911–926. <https://doi.org/10.1104/pp.102.017459>
- Werner, G.D.A., Cornwell, W.K., Sprent, J.I., Kattge, J. & Kiers, E.T.** 2014. A single evolutionary innovation drives the deep evolution of symbiotic N₂-fixation in angiosperms. *Nature, Commun.* 5: 4087. <https://doi.org/10.1038/ncomms5087>
- Werner, G.D.A., Cornwell, W.K., Cornelissen, J.H.C. & Kiers, E.T.** 2015. Evolutionary signals of symbiotic persistence in the legume–rhizobia mutualism. *Proc. Natl. Acad. Sci. U.S.A.* 102: 10262–10269. <https://doi.org/10.1073/pnas.1424030112>
- Wojciechowski, M.F.** 2013. Towards a new classification of Leguminosae: Naming clades using non-Linnaean phylogenetic nomenclature. *S. African J. Bot.* 89: 85–93. <https://doi.org/10.1016/j.sajb.2013.06.017>
- Wojciechowski, M.F., Sanderson, M.J., Steele, K.P. & Liston, A.** 2000. Molecular phylogeny of the “temperate herbaceous tribes” of papilionoid legumes: A supertree approach. Pp. 277–298 in: Herendeen, P.S. & Bruneau, A. (eds.), *Advances in legume systematics*, part 9. Richmond, U.K.: Royal Botanic Gardens, Kew.
- Wojciechowski, M.F., Lavin, M. & Sanderson, M.J.** 2004. A phylogeny of the legumes (Leguminosae) based on analysis of the plastid *matK* gene sequences resolves many well-supported subclades within the family. *Amer. J. Bot.* 91: 1846–1862. <https://doi.org/10.3732/ajb.91.11.1846>
- Yahara, T., Javadi, F., Onoda, Y., Queiroz, L.P. de, Faith, D., Prado, D.E., Akasaka, M., Kadoya, T., Ishihama, F., Davies, S., Slik, J.W.F., Yi, T., Ma, K., Bin, C., Darnaedi, D., Pennington, R.T., Tuda, M., Shimada, M., Ito, M., Egan, A.N., Buerki, S., Raes, N., Kajita, T., Vatanparast, M., Mimura, M., Tachida, H., Iwasa, Y., Smith, G.F., Victor, J.E. & Nkonki, T.** 2013. Global legume diversity assessment: Concepts, key indicators, and strategies. *Taxon* 62: 249–266. <https://doi.org/10.12705/622.12>
- Zanne, A.E., Tank, D.C., Cornwell, W.K., Eastman, J.M., Smith, S.A., FitzJohn, R.G., McGlenn, D.J., O’Meara, B.C., Moles, A.T., Reich, P.B., Royer, D.L., Soltis, D.E., Stevens, P.F., Westoby, M., Wright, I.J., Aarssen, L., Bertin, R.I., Calaminus, A., Govaerts, R., Hemmings, F., Leishman, M.R., Oleksyn, J., Soltis, P.S., Swenson, N.G., Warman, L. & Beaulieu, J.M.** 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* 506: 89–92. <https://doi.org/10.1038/nature12872>

Appendix 1. Materials and Methods: A densely sampled phylogeny of the Leguminosae based on analyses of *matK* gene sequences.

Sampling. — Previously published and 637 newly generated *matK* gene sequences were obtained from multiple laboratories. Only fully vouchered samples, authoritatively identified by taxonomic specialists are included, and all sequences have been submitted to GenBank (Table S1). Most sequences comprise the full *matK* coding sequence, but for a subset of species only 620–780 nucleotides of the central gene region, the “barcode” *matK* region (from ca. 600 to 1450 in the aligned sequence matrix), were available. Our objective was to include as many legume genera and species as possible,

while at the same time ensuring sequence quality and taxonomic accuracy.

Multiple accessions per species were included in initial phylogenetic analyses (all accessions listed in Table S1) in order to verify sequence accuracy and try to eliminate problems of sequence contamination or specimen taxonomic identity. A total of 5560 legume sequences were verified and analysed. Most sequences were also subjected to a BLAST search (<http://blast.ncbi.nlm.nih.gov/>) to verify sequence accuracy. Subsequently a single sequence per species (or infraspecific taxon) was chosen for the full phylogenetic analyses

Appendix 1. Continued.

(specimens selected in the single-taxon analyses are indicated by * in Table S1). Full-length gene sequences were preferentially selected. The aligned matrix, accession list with voucher information and GenBank numbers, and tree files (Bayesian and bootstrap majority-rule consensus trees, best-scoring ML tree, 1000 Bayesian posterior trees and 1000 bootstrapped trees, all in newick format) are available as Supplementary Data and in Data Dryad (DOI: <https://doi.org/10.5061/dryad.61pd6>).

The final matrix includes 3842 sequences representing 3696 legume species (identified with * in Table S1) and covering 698 of the 765 currently recognised legume genera (Table 2). The sampling for subfamily Cercidoideae includes 96 species representing all 12 genera; for Detarioideae, 327 species (plus 3 infraspecific taxa) are included from 79 of the 84 genera; for Dialioideae, 19 species are included, representing 15 of the 17 genera; for Caesalpinioideae, we include 937 species (plus 5 infraspecific taxa), representing 146 of the 148 genera; and for Papilionoideae, 2316 species are included (plus 38 infraspecific taxa), representing 445 of the 503 genera. This represents the most comprehensive generic and species sampling of Leguminosae in a phylogenetic analysis of the family to date.

We also included 100 outgroup sequences, sampled across Eudicots, including relatively dense sampling of the three other families of Fabales (Table S1). Broad sampling of outgroup taxa was included to facilitate downstream analyses requiring branch lengths and wider interfamilial relationships. Sequences for outgroup taxa were obtained from vouchered GenBank sequences (including published complete plastome sequences) and the 1000 Plants Project (OneKP or IKP), as indicated in Table S1.

Phylogenetic analyses. — We initially built four separate matrices for Papilionoideae, the mimosoid clade, lineages of the former Caesalpinioideae, and the outgroup taxa. For each, an initial alignment for a subset of taxa was made using MACSE v1.01b (Ranwez & al., 2011) using default settings, in order to obtain an alignment that respects the open reading frame (ORF) and does not allow indels within codons. Running the complete alignment on MACSE was not possible because it is too computationally intensive. The four

initial alignments were then merged with the MERGE function and additional sequences added using the --add function in MAFFT v.7 (Katoh & Standley, 2013). The complete matrix was checked by eye and alignments were corrected to ensure that all sequences were aligned with respect to the ORF. An exception was made for sequences belonging to new sense Caesalpinioideae, which share a frameshift mutation near the end of the ORF. Two ambiguity symbols (“?”s) were inserted, disrupting the ORF but ensuring assumed homology at the nucleotide level.

Aligned matrices were analysed using maximum likelihood and Bayesian inference. Initial analyses were implemented with RAxML v.8.0 (Stamatakis, 2014) using the GTRGAMMA model with 100 bootstrap replicates to check for problematic sequences and to ensure that shorter incomplete gene region sequences did not lead to spurious phylogenetic relationships (e.g., grouping together of shorter sequences). We used PartitionFinder v.2 (Lanfear & al., 2012) to determine whether to partition codons separately or not. The program favoured a single partition for all three codon positions together. Complete analyses of the final matrix were implemented using a maximum likelihood approach analysed with RAxML, using the GTRGAMMA model, and support was assessed through 1000 rapid bootstrap replicates. The Bayesian analyses were implemented in PhyloBayes-MPI v.1.5a (Lartillot & al., 2009), with the GTR model, running two chains until they reached convergence, as determined with Tracer v.1.6 (Rambaut & al., 2014). The two chains were run for a total of 25,891 and 25,512 cycles, and the majority-rule consensus tree produced by the program bpcmp (included in the PhyloBayes package) was based on 1075 posterior trees sampled from both chains. A total of 20 accessions were pruned from all trees post-analysis. These were duplicate accessions or problematically vouchered accessions that were not discovered until after running the final analyses. We decided to prune these to ensure that the phylogenetic trees are as clean as possible for potential downstream comparative analyses. The final RAxML and PhyloBayes analyses were conducted on the Cipres Portal (Miller & al., 2010) and on the ScienceCloud of the University of Zurich, respectively.

Appendix 2. Glossary of some morphological terms used in Table 1, the key and subfamily descriptions. Illustrations of some key traits are provided in Figs. 12 & 13.

Anthotaxy (inflorescence) – the arrangement of flowers along the inflorescence axis.

Bipinnate (leaves) – a twice pinnately compound leaf, in which leaflets are arranged in pinnae along the main leaf axis (rachis) (Fig. 12C).

Crescent shaped (hilum) – a U- or V-shaped hilum.

Craspedium (fruit) – an indehiscent fruit that breaks apart, either with the valves separating as a single unit, or into one-seeded segments (articles), but leaves the sutures as a persistent margin (the replum) (Fig. 13A).

Drupaceous (fruit) – here used to refer to true drupes and similar fruits. A drupe is an indehiscent fruit with an outer fleshy part surrounding the pyrene (“stone”) of hardened endocarp (Fig. 13C).

Exstipellate (leaves) – a leaf with no stipels at the leaflet bases.

Hilum – a scar left on the seed coat from its attachment by a funicle to the placenta. In subfam. Papilionoideae, the hilum is elongate and split lengthwise by a hilar groove and the hilar region is usually provided with a lens (Fig. 13J & K). In the other subfamilies, the hilum is circular, elliptic, punctiform or crescent shaped and can occur apically, subapically or laterally.

Imparipinnate (leaves) – a pinnately compound leaf (with a rachis) with a single terminal leaflet (= odd pinnate) (Fig. 12B).

Lens (seed) – a mound situated near the hilum, usually located opposite the micropyle with the hilum between both structures; an area of weakness where water initially penetrates the seed prior to germination (Fig. 13J & K).

Loment (fruit) – a jointed indehiscent fruit (common in legumes) that breaks apart in one-seeded segments (articles) (Fig. 13B).

Overgrown (seed) – a seed that enlarges and fills the seed-cavity of the pod without differentiation of the testa and thus the growth is limited by the size of the pod (Fig. 13F).

Palmate (leaves) – a leaf in which leaflets arise from the apex of the petiole (i.e., there is no leaf rachis), as fingers originate from the palm of a hand; in legumes used for such leaves with 4 or more leaflets (i.e., not for digitately trifoliolate leaves) (Fig. 12D).

Paraphyllidium, plural paraphyllidia (leaves) – reduced leaflets situated at the base of a pinna-rachis, immediately contiguous to its pulvinus (Fig. 12F).

Paripinnate (leaves) – a pinnately compound leaf (with a rachis) with a pair of opposite terminal leaflets (= even pinnate) (Fig. 12A).

Pleurogram (seed) – a fracture line in a seed exotestal palisade leaving a U- or O-shape on both seed faces (Fig. 13G & H).

Prickles (mechanical defense) – extensions of the plant surface (cortex and epidermis) with sharp, stiff ends; the prickles detachable without tearing the organ which they protect.

Pseudopleurogram (seed) – a coloured line on the seed surface but this not resulting from a break in exotestal palisade (i.e., not a fracture line) (Fig. 13I).

Pseudoraceme (inflorescence) – a compound raceme in which each bract subtends two or more flowers in highly condensed lateral axes (Fig. 12G & H).

Samaroid (fruit) – here used to refer to true samaras and similar fruits. A true samara is a dry, indehiscent, winged fruit, the flattened wing derived from the ovary wall and usually longer than the seed-bearing part; in samaroids the wing can encircle the seed chamber (Fig. 13D & E).

Spathaceous (calyx) – a bilaterally symmetrical calyx in which all sepals are unilaterally joined, usually splitting along one line of weakness at flower anthesis.

Spines (mechanical defense) – modified leaves, stipules, branches, or parts of leaves with sharp, stiff ends; always with a vascular origin.

Stipel (leaves) – a stipule-like appendage at the base of a leaflet (Fig. 12E).

Stipellate (leaves) – a leaf with leaflets provided with stipels (Fig. 12E).

Synandrium (androecium) – an androecium in which the stamens are fused both by the filaments and anthers.

Thyrse (inflorescence) – a panicle composed of cymose lateral units (Fig. 12I & J).

Thyrsoid (inflorescence) – like a thyrse.

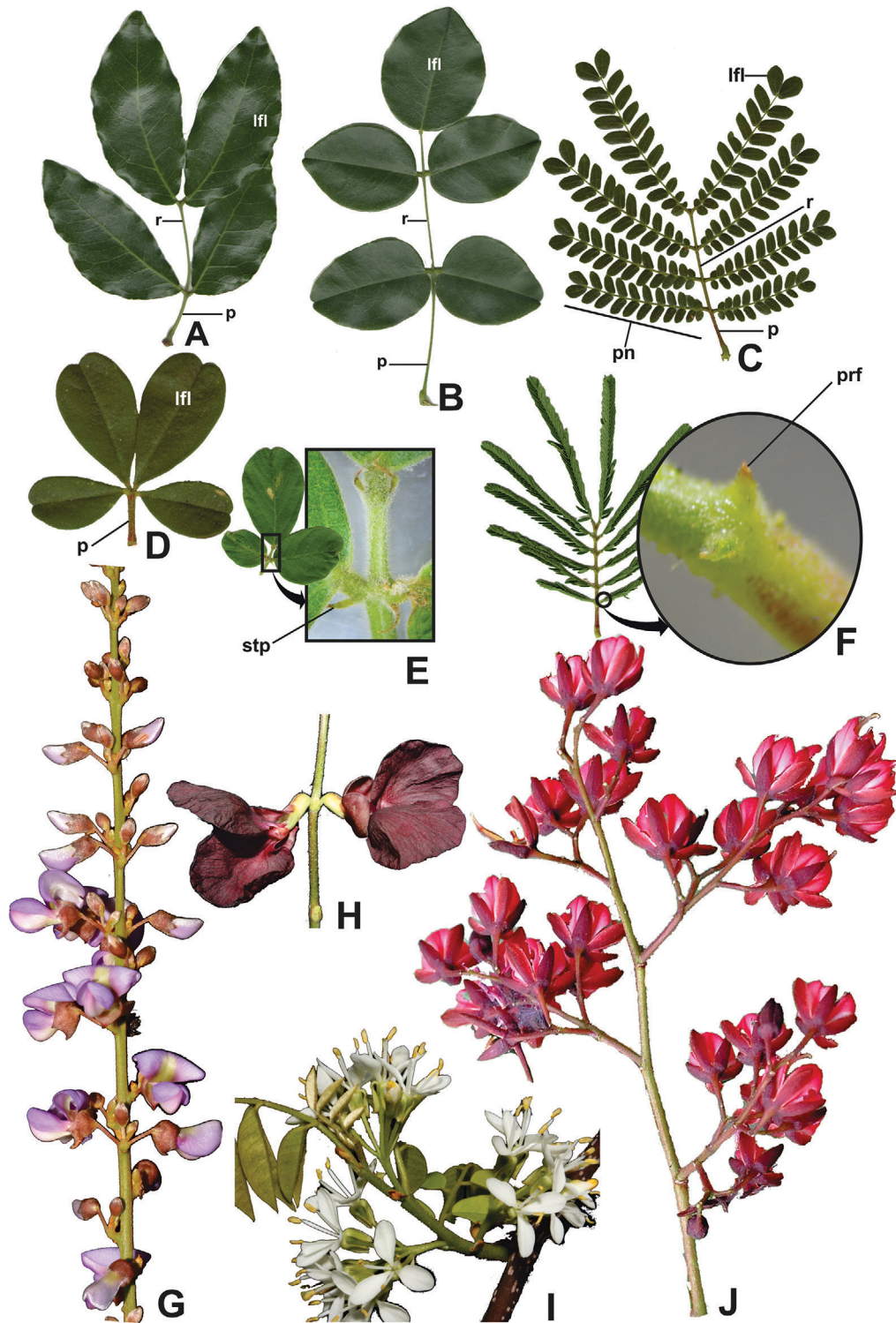


Fig. 12 (Appendix 2). Leaves and inflorescences. **A**, A paripinnate leaf of *Goniorrhachis marginata* Taub. (Detarioideae); **B**, An imparipinnate leaf of *Luetzelburgia bahiensis* Yakovlev (Papilionoideae); **C**, A bipinnate leaf of *Pityrocarpa moniliformis* (Benth.) Luckow & R.W.Jobson (Caesalpinioideae, mimosoid clade); **D**, A palmately compound leaf of *Zornia myriadena* Benth. (Papilionoideae); **E**, A pinnately trifoliolate leaf of *Centrosema arenarium* Benth. (Papilionoideae) highlighting the stipels at the base of leaflets (inset); **F**, A bipinnate leaf of *Mimosa tenuiflora* (Willd.) Poir. (Caesalpinioideae, mimosoid clade) showing a pair of paraphyllidia near the base of the pinna (inset); **G**, A pseudoraceme of *Deguelia nitidula* (Benth.) A.M.G.Azevedo & R.A.Camargo (Papilionoideae) with condensed multiflorous lateral axes; **H**, Part of a pseudoraceme of *Macroptilium bracteatum* (Nees & Mart.) Maréchal & Baudet (Papilionoideae) with biflorous lateral axis; **I & J**, Thyrsoid inflorescences of *Apuleia leiocarpa* (Vogel) J.F.Macbr. (**I**, Dialioideae) and *Zenia insignis* Chun (**J**, Dialioideae). — lfl, leaflet; p, petiole; pn, pinna; prf, paraphyllidium; r, leaf rachis; stp, stipel. — Photos: G, Luciano P. de Queiroz; H & I, Domingos Cardoso; J, Shijin Li.

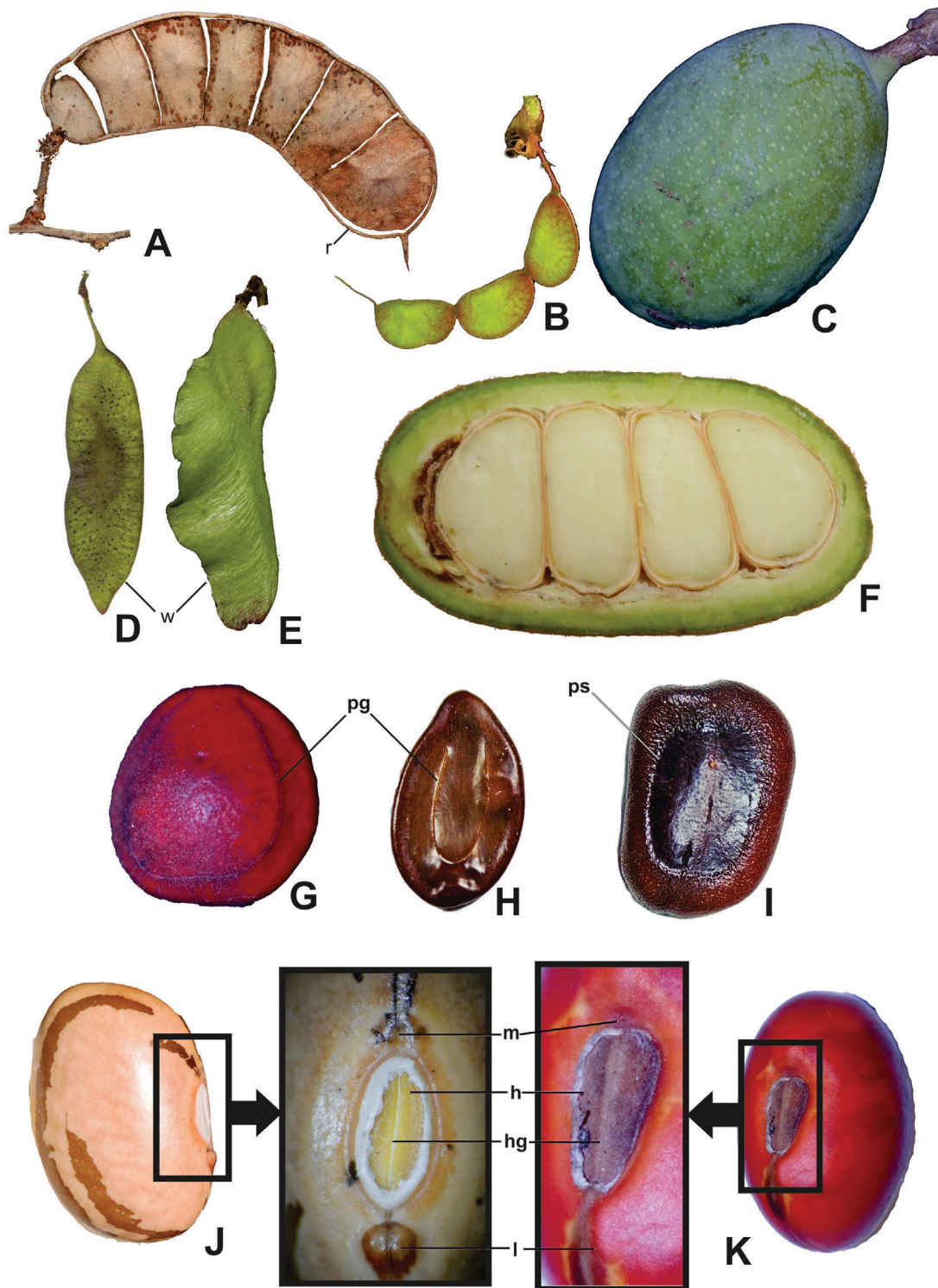


Fig. 13 (Appendix 2). Fruits and seeds. **A**, A craspedium of *Mimosa irrigua* Barneby (Caesalpinioideae, mimosoid clade) showing the one-seeded segments (articles) and the persistent marginal replum (r); **B**, A loment of *Aeschynomene martii* Benth. showing the one-seeded segments but no persistent replum; **C**, A drupe of *Andira humilis* Mart. ex Benth.; **D & E**, Two kinds of samaroid fruits (wings, w); **D**, *Dalbergia nigra* (Vell.) Allemão ex Benth. (Papilionoideae) and **E**, *Luetzelburgia andrade-limae* H.C.Lima (Papilionoideae); **F**, Indehiscent fruit of *Dioclea edulis* Kuhl. split lengthwise to show the overgrown seeds; **G & H**, Seeds of Caesalpinioideae legumes (mimosoid clade): **G**, *Adenanthera pavonina* L. and **H**, *Leucaena leucocephala* (Lam.) De Wit showing the pleurogram (pg); **I**, Seed of *Tamarindus indica* L. (Detarioideae) showing the pseudopleurogram (ps); **J–K**, Seeds of the common bean: **J**, *Phaseolus vulgaris* L. and **K**, *Erythrina velutina* Willd. highlighting the major features of Papilionoideae seeds (insets), with an elongate hilum (h) split lengthwise by a hilar groove (hg) and bearing the micropyle (m) and the lens (l) at the opposite poles of the hilar region. — Photos: A & E, Domingos Cardoso; B–D & G–K, Luciano P. de Queiroz; F, Alex Popovkin.