

The Rest of the Iceberg. Legume Diversity and Evolution in a Phylogenetic Context¹

Jeff J. Doyle* and Melissa A. Luckow

L.H. Bailey Hortorium, Department of Plant Biology, 462 Mann Library, Cornell University, Ithaca, New York 14853

Most readers of a special issue of *Plant Physiology* on legumes will be familiar with only a handful of species, primarily pea (*Pisum sativum*) and the various economically important “beans” such as soybean (*Glycine max*), and of course, the model legumes *Medicago truncatula* and *Lotus japonicus*. That leaves around 700 other legume genera and 20,000 species left to consider—legumes are the third largest flowering plant family, behind only orchids (Orchidaceae) and asters (Asteraceae). And the numbers tell only part of the story. Neither these two larger families, nor the grasses, another large group, even begin to approach the legumes in their overall range of variation. The legumes are incredibly diverse in every way imaginable and defy generalization about almost any attribute. Even the characteristic fruit type that gives legumes their name is highly variable and ranges from tiny single-seeded forms to meter-long woody pods and from typical dehiscent legumes to indehiscent wind-dispersed winged fruits and articulated loments with “stick-tight” dispersal strategies.

Ecologically the family ranges from rain forests to deserts, and from lowland to alpine habitats; there are even aquatic species. They include giant forest trees that are prominent sources of lumber and expensive woods (e.g. Brazilian rosewood [*Dalbergia nigra*]), shrubs of all sizes and habits, lianas from annual twiners to woody behemoths such as wisteria or kudzu, and tiny annual herbs. Nodulation, that trademark symbiosis of legumes, is conspicuously absent in several major lineages. If you think you could recognize all legume genera by the butterfly-shaped (“papilionoid”) flowers so familiar from pea and its relatives, think again (Fig. 1). The early-diverging lineages have flowers that look more like those of a wild rose, and many modifications exist, from the reduced flowers of the aptly named *Amorpha* (not too far in the family tree from soybean and pea) to spectacular orchid-like flowers of species that share only a more distant ancestry with these models. There are flowers with one stamen, and whole

groups of genera with flowers that compensate for their nondescript petals with their many, showy stamens. There is pollination by bees, moths, butterflies, perching birds, hummingbirds, and bats, not to mention the do-it-yourselfers. Many species shed their pollen in tetrads or larger masses rather than as individual grains (might some geneticist be interested in developing a *Neurospora*-like legume system of tetrad analysis?). There is a tremendous diversity of secondary compounds, particularly alkaloids, many of them biologically active, such as the fish poisons of several woody tropical groups.

Current model legumes cover only a fraction of this diversity—the tip of the iceberg of variation in the family. New models would be useful to provide the kind of phylogenetic coverage that exists in grasses, for example, where the primary model, rice (*Oryza sativa*), is about as distantly related from the other well-studied taxa, maize (*Zea mays*) and the triticoid grasses (wheat [*Triticum aestivum*], barley [*Hordeum vulgare*], and oats [*Avena sativa*]), as is possible. As we will show, this is not the case in legumes. However, among the distant relatives of *Medicago*, *Glycine*, and *Lotus* are species with short life cycles and small genomes just waiting to be exploited.

One thing that has been lacking has been a road map of legume diversity. But great progress in understanding the phylogeny and evolution of the legume family has been made in the last decade. We will concentrate here on sketching out the legume family tree, and we will highlight some features of legume diversity in this phylogenetic context. Some of the biologically interesting questions that could be addressed in the family will be obvious, and some are mentioned here. But legumes are such a treasure trove of diversity that we can only scratch the surface. The map is now in hand, and the model systems provide jumping off points for further exploration.

BUILDING THE LEGUME FAMILY TREE

The last decade has seen a tremendous explosion of phylogenetic studies in plants as a whole and particularly in flowering plants. This has been due to parallel revolutions in theory, computing, and molecular technology, whose synergies have resulted in the focusing (some would say “narrowing”) of the discipline of systematics—the scientific study of the

¹ This work was supported by the National Science Foundation Systematic Biology Program.

* Corresponding author; e-mail jjd5@cornell.edu; fax 607-255-7979.

www.plantphysiol.org/cgi/doi/10.1104/pp.102.018150.



Figure 1. The diversity of flowers in the Leguminosae. Top row, right to left, *Delonix regia* (Caesalpinioideae, *Peltophorum* group) is known in the wild from only several small populations in northern Madagascar, but is now cultivated throughout the tropics; *Senna alata* (Caesalpinioideae, Cassieae) has been used extensively in folk medicine as a cure for skin diseases; and *Bauhinia xerophyta* (Caesalpinioideae, Cercideae) is another Malagasy plant and is a member of the first-diverging legume lineage. Second row, left to right, *Colvillea racemosa* (Caesalpinioideae, *Peltophorum* group) is closely related to
(Legend continues on following page.)

kinds and diversity of organisms and of any and all relationships among them" (Simpson, 1961)—on the production of phylogenetic trees. A look inside journals dealing with systematics and evolution will reveal paper after paper presenting phylogenetic trees constructed from gene sequence data and hypothesizing relationships among diverse groups of organisms. Any gene can be used, theoretically, for reconstructing phylogenies; function is not important. Choice of gene is based largely on characteristics such as copy number (gene families other than ribosomal genes are notoriously difficult) and rate of evolution relative to the taxonomic group being studied (there must be enough changes to mark all of the interesting species-splitting cladogenic events, but not so much that multiple changes at individual sites have erased too much historical information). In flowering plants, chloroplast genes have been very popular, with different sequences used for studies at different taxonomic levels—relatively slowly evolving genes such as *rbcL* (encoding the large subunit of ribulose biphosphate carboxylase/oxygenase) for family relationships and more rapidly evolving spacers of various genes are used for studies of genera and species. Different portions of the nuclear-encoded ribosomal cistron are used for different purposes: The highly conserved small subunit (18S) gene is used at high taxonomic levels, whereas the internal transcribed spacer is much more variable and is used at lower levels.

With the exponential increase in molecular phylogenies has come a parallel efflorescence of systematic and evolutionary theory, addressing questions as fundamental as how to infer relationships of a group of organisms (a "species tree") from a tree that describes the relationships of a set of homologous sequences sampled from those organisms (a "gene tree")—the two are not necessarily the same! There is also intense, often acrimonious debate about the degree to which formal taxonomic groupings should reflect this newly produced knowledge of phylogenetic relationships, beyond correcting egregious mistakes or cleaning up the "missing persons" list of enigmatic, previously ungrouped taxa. The result of all of this has been a revolution in our understanding of flowering plant relationships. At higher taxonomic levels, intuitive classifications, with their interpretive

diagrams of relationships among families, have given way to objectively constructed phylogenetic trees.

Within this overall revolution, the legumes have been the focus of an active and collaborative international group of researchers in the area of systematics and evolution, for years nucleated most actively by Roger Polhill of the Royal Botanic Gardens, Kew in England. Beginning with the legume "bible"—the 1981 *Advances in Legume Systematics, Part 1* (Polhill and Raven, 1981)—and extending through the latest *ALS Part 10* (Klitgaard and Bruneau, 2003), there has been a steady stream of volumes dedicated to legume phylogeny and evolution. This year will see not only the publication of *ALS, Part 10*, but also the comprehensive *Legumes of the World* (Lewis et al., 2003).

THE BIG PICTURE: LEGUMES AMONG THE ROSIDS AND THE MAJOR DIVISIONS IN THE FAMILY

Traditional classifications viewed the legumes as a very distinctive family, largely due to its legume fruit, with affinities somewhere among an extremely large and ill-defined group of dicots called "rosids" (named for the rose family, Rosaceae). Numerous suggestions for closest allies were made by various authors, but molecular phylogenies have turned out not to support any of these, instead grouping legumes with the small tropical family Surianaceae, the genus *Quillaja* (the Chilean soap tree), and the family Polygalaceae, some of whose members have bilaterally symmetrical flowers superficially like those of papilionoid legumes (Fig. 2; Soltis et al., 2000). This group (order Fabales) is in turn part of a large "eurosoid I" group that contains, among other things, all of the families with nodulation symbioses. Right next door is the eurosoid II group, which includes the mustards (Brassicaceae), to which *Arabidopsis* belongs. *Arabidopsis* is thus not nearly as distantly related to legumes as older classifications suggested.

The Leguminosae traditionally has been divided into three major groups of very unequal size on the basis of major morphological characters, particularly floral ones. These are most commonly recognized as subfamilies Caesalpinioideae, Mimosoideae, and Papilionoideae of a single family (Leguminosae or Fabaceae) or, less frequently, as three separate but

Figure 1. (Legend continued from previous page.)

D. regia above, but it has a quite different floral structure and is bird-pollinated; *Calliandra tergemina* (Mimosoideae) is native to Latin America; and *Dichrostachys cinerea* (Mimosoideae) has sterile flowers that turn from rose to white as the inflorescence ages, presumably as a signal to pollinators. Third row, left to right, *Acacia* sp. (Mimosoideae) are often armed with thorns or spines and dominate the dry savannahs and deserts of the world; *Mimosa dysocarpa* (Mimosoideae) is from southern Arizona and is related to the sensitive plant often used to demonstrate thigmotropism; and *Amorpha fruticosa* (Papilionoideae) is native to North America and has small un-papilionoid flowers that vary in color from deep purple to white, even within a single population. Bottom row, left to right, *Clitoria lasciva* (Papilionoideae) is native to Madagascar, where an infusion of the root is used to treat constipation; *Erythrina crista-galli* (Papilionoideae), native to tropical South America but now widely cultivated throughout the tropics, has close relatives with tubular, hummingbird-pollinated flowers; and *Vigna* sp. (Papilionoideae) has a coiled keel that retracts to expose the anthers when a pollinator lands on the flower. Photo of *D. regia* is courtesy of Jacqueline Salazar; photo of *Acacia* sp. is courtesy of Jerrold Davis.

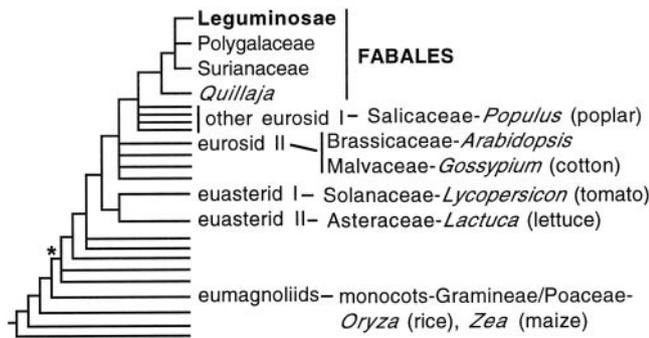


Figure 2. Relationships of legumes and allies (Fabales, euroid I clade) to other major clades of flowering plants. The tree is summarized from (Soltis et al., 2000) and is rooted with gymnosperms. All taxa above the ancestor marked with an asterisk are classified as “eudicots” by (Soltis et al., 2000). A few important model plant groups representing other major clades are indicated. Eumagnoliids include numerous other dicots as well as monocots.

closely related families (Caesalpiniaceae, Mimosaaceae, and Fabaceae). The papilionoids, by far the largest subfamily with 476 genera and about 14,000 species (Lewis et al., 2003), are the familiar “beans,” to which all of the model taxa belong. Mimosoideae, with 77 genera and around 3,000 species, includes such groups as the acacias, and Caesalpinoideae is a diverse, mostly tropical assemblage of 162 genera and approximately 3,000 species. The three subfamilies are broken into groups of genera called tribes.

On the basis of molecular phylogenetic studies, Papilionoideae and Mimosoideae appear to represent lineages that are fully or nearly monophyletic—natural in the systematic sense, meaning that they include an ancestor plus all of its descendants and no extraneous, unrelated taxa descended from other ancestors (Fig. 3; e.g. Doyle et al., 2000; Kajita et al., 2001; Wojciechowski, 2003). The third subfamily, though, has long been suspected to be unnatural (paraphyletic), comprising a diverse assemblage of unrelated “caesalpinioid” lineages mostly diverging relatively early in the history of the family and lacking the distinctive floral features used to group genera into the other two families. Among these lineages is the one that molecular data suggest was earliest to diverge in the family, the caesalpinioid tribe Cercideae. Readers in temperate climates may be familiar with the genus *Cercis*, whose species include the redbuds of North America and the Judas tree of Mediterranean Europe, small trees whose flowers are superficially similar to those of papilionoid legumes. The largest genus in the Cercideae, however, is the tropical genus *Bauhinia*, comprising trees, shrubs, and lianas with exotic-looking flowers that bear more resemblance to orchids than to peas (Fig. 1).

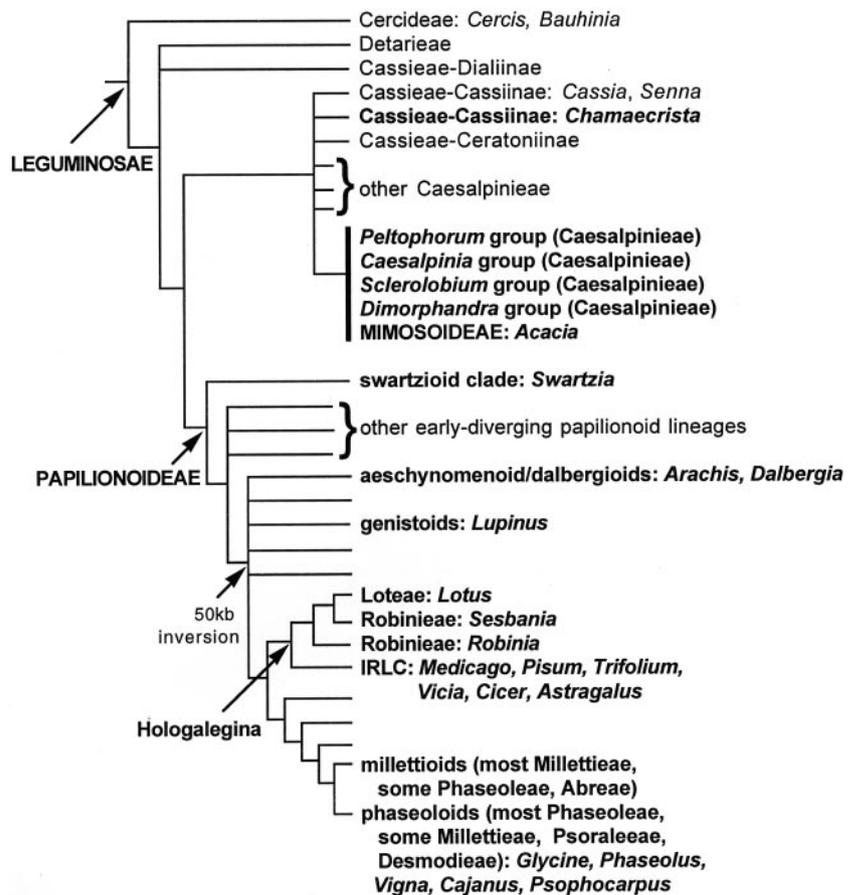
There is some disagreement among molecular analyses about the branching order of the two next-diverging lineages in the family, and so this is shown as an unresolved trichotomy (Fig. 3). Neither will be known to most readers, but the Detarieae/Amher-

stieae clade contains the largest group of genera in the Caesalpinoideae, is an important group in tropical ecosystems of South America and particularly Africa, and is tremendously diverse in chemistry and floral morphology. Considerable progress has been made in understanding relationships among its genera (Bruneau et al., 2001; Herendeen et al., 2003).

The next dichotomy in the tree separates the papilionoid legumes from a clade that includes the remaining Caesalpinoideae and all members of the Mimosoideae. Molecular data have yet to provide much overall resolution in this caesalpinioid/mimosoid clade, and it is to be hoped that this will be remedied as additional genes are added, because the group is very interesting for several reasons. First, there is the evolution of the very distinctive Mimosoideae, major groups of which are characterized by small, clustered flowers with numerous showy stamens and pollen shed in aggregates. Although most “typical” mimosoids form a very strongly supported group, relationships among early-diverging members of the subfamily and members of the *Dimorphandra* group of caesalpinioids are hazy (Luckow et al., 2000; Bruneau et al., 2001). Second, as is discussed below, the caesalpinioid/mimosoid clade contains all of the genera outside of Papilionoideae that are definitively known to nodulate. Finally, even excluding the mimosoids from consideration, this clade is tremendously diverse morphologically. For example, flowers vary from reduced and actually or functionally unisexual in *Ceratonia* (carob), *Gymnocladus* (Kentucky coffee tree), and *Gleditsia* (honeylocust) to large and showy in ornamental trees such as *Delonix* (poinciana, flamboyant).

Progress in understanding phylogenetic relationships in two of the large informal “groups” of tribe Caesalpinieae is reported in *ALS Part 10* (*Peltophorum* group, Haston et al., 2003; *Caesalpinia* group, Simpson et al., 2003); both studies illustrate the complexity of relationships and the inadequacy of current concepts of some large genera. Phylogenetic relationships are better-resolved among the core Mimosoideae (Luckow et al., 2000, 2003), and although not congruent with traditional generic groupings, they correspond rather well with general ideas about mimosoid character evolution (Luckow et al., 2000, 2003). For example, phylogenetic studies support a progression from flowers with 10 free stamens to those with many free stamens, and eventually to fusion of many stamens. Also, rainforest trees of Africa and South America tend to be early offshoots in the mimosoid radiation (“relicts” in prephylogenetic terminology), whereas Asian taxa are confined to more recently diverging lineages. But new conclusions can be drawn from these phylogenies as well. For example, the large genus *Acacia*, whose approximately 1,500 species were grouped on the basis of having many free stamens, does not form a monophyletic group. The Australian acacias are more

Figure 3. Summary of phylogenetic relationships in Leguminosae based on recent molecular analyses (see text for details). Major lineages are indicated along the backbone of the tree by arrows. Leguminosae not indicated as being either Mimosoideae or Papilionoideae are classified as Caesalpinioideae (a classic paraphyletic group). Only groups mentioned in the text are included; some other lineages are indicated but not named. Genera that include model species are indicated. Taxa in bold type are dominated by nodulating species.



closely related to genera with many fused stamens (tribe Ingeae) than they are to the New World and African species of *Acacia* (Luckow et al., 2003). Thus, this ubiquitous and ecologically important genus will undoubtedly be partitioned into a number of smaller genera in the near future.

MAJOR BRANCHES IN THE PAPILIONOID TREE

Comprehensive molecular analyses of the legumes resolve Papilionoideae as a monophyletic group, although with rather weak support in the most comprehensive analyses (Kajita et al., 2001), and suggest that it had diverged from other legumes as early as 45 to 50 million years ago (mya; Wojciechowski, 2003). Papilionoid legumes are distinguished from other legumes by several morphological characters, mainly rather technical ones such as the orientation of the seed hilum and unidirectional initiation of sepals (Doyle et al., 2000). Interestingly, the classic “papilionoid” flower is not found in many of the early-diverging lineages of the Papilionoideae. The ancestral floral condition for the family is unclear, in part because of uncertainties about relationships near the base of the papilionoid tree (Pennington et al., 2001). Resolving relationships among these early-diverging lineages is important for understanding not only flo-

ral evolution, but also the evolution of nodulation (see below).

Four major groups of papilionoids, each containing important domesticated and/or model taxa, are supported by several sources of molecular data, but relationships among some of them are unclear (Fig. 3). The aeschynomenoid/dalbergioid clade is a diverse group that includes the peanut (*Arachis hypogaea*), and other herbaceous genera, but also woody plants such as the Brazilian rosewood (Lavin et al., 2001). Members of the genistoid group all produce characteristic quinolizidine alkaloids. The group’s main lineages probably originated in the Old World tropics and radiated into north temperate regions, or in Africa with secondary radiations from Africa into north temperate regions (Crisp et al., 2000). The most familiar genistoid genus is *Lupinus*, but like the aeschynomenoid/dalbergioid group, the genistoids are quite diverse.

The two remaining large clades, Hologalegina and the phaseoloid/millettoid group, are sister taxa in molecular phylogenies, meaning that they share a most recent common ancestor. Both are large and diverse groups, and between them, they include the most important legume model taxa, *L. japonicus*, *M. truncatula*, and pea in Hologalegina and soybean in the phaseoloid/millettoid group. Hologalegina is it-

self split into two lineages, one of which includes Loteae and Robinieae (*Robinia* spp., e.g. black locust; and *Sesbania* spp., of interest because of stem-nodulation in some species). The most recent phylogeny of Loteae (Allan et al., 2003) does not include *L. japonicus*, but does include another member of its species complex, the Eurasian birdsfoot trefoil (*Lotus corniculatus*), which is commonly found as a roadside weed in much of eastern North America. The second clade of Hologalegina is marked by the loss of one copy of the large (approximately 25 kb) inverted repeat commonly found in chloroplast genomes of angiosperms and is accordingly called the inverted repeat-loss clade (IRLC; Wojciechowski et al., 2000). The IRLC is dominated by temperate, herbaceous genera, including familiar plants such as *Pisum*, *Vicia* (vetch and broadbean), *Cicer arietinum* (chickpea), *Medicago* (alfalfa, medics), and *Trifolium* (clovers) but also less familiar genera such as *Astragalus* (locoweeds), which is the largest legume genus with over 2,000 species. In a recent phylogenetic analysis, *Medicago truncatula* and *M. italica* were strongly supported as sister species, with this group embedded with other *Medicago* spp. in a series of well-supported relationships (Steele and Wojciechowski, 2003).

The largest papilionoid subgroup in number of genera is the phaseoloid/millettioid group, which, like Hologalegina, includes a number of domesticated taxa such as soybean, *Phaseolus* (garden bean and runner bean), *Vigna* (cowpea and mungbean), pigeon pea (*Cajanus cajan*), and *Psophocarpus* (winged bean). Relationships in the group are complex and include elements of several tribes (e.g. Kajita et al., 2001; Hu et al., 2002). Within the phaseoloid group, the closest generic ally of *Glycine* spp. appears to be the pantropical genus *Teramnus* (Lee and Hymowitz, 2001).

WHERE AND WHEN DID LEGUMES ORIGINATE?

The legumes most familiar to the reader are likely to belong to relatively recent northern hemisphere radiations of the family that includes many members of the genistoid and IRLC groups (Fig. 3). However, legumes dominate the lowland tropical rainforests of Africa and South America (but not Asia) and are a major component of dry and seasonally dry tropical forests as well. For example, Gentry (1993) reported that legumes outnumbered all other plant families both in numbers of individuals and of species in his survey plots at various South American sites and in the lowland rainforests of Africa and Madagascar. As diverse as the family is today, all of its members trace their origin to a single species that lived a long time ago in some specific place and adapted to a particular environment, and all of the incredible diversity of the family was built, step by mutational step, from the genome of this species. Where and when was this, and what might the conditions of this common an-

cestor and of subsequent founders of major lineages tell us about the biology of modern legumes? Biogeography is a discipline that attempts to explain the current distribution of organisms in terms of historical and ecological factors, and the legumes, constituting such a large and diverse family of worldwide distribution, clearly invite biogeographic inquiry.

The time and place of origin of legumes remains something of a mystery. Legumes have not been among the families represented in the rich mid-Cretaceous (approximately 90 mya) floral fossil record that in recent years has come to include ancient relatives of many modern groups such as Arabidopsis. The fossil record of the legumes is not particularly rich until about 35 to 54 mya (Eocene or mid-Tertiary), when papilionoid and mimosoid legumes become abundant and diverse in both North American and European fossil floras (Taylor, 1990; Herendeen et al., 1992). Legumes are not as well represented in the fossil floras of South America and Africa, but this may reflect less intensive sampling and poorer preservation than in the northern hemisphere, rather than a restricted distribution of the family. That the major lineages had all diverged from one another by around 50 mya is in general agreement with dates based on molecular estimates, which further suggest that the major papilionoid lineages may have diverged from one another earlier than previously suggested—for example, the *Lotus* and *Medicago* lineages may have diverged from one another by around 40 mya (Wojciechowski, 2003).

Until recently, legumes were considered to have originated in Africa in the late Cretaceous (Fig. 4), from whence they migrated to South America and subsequently North America, leaving behind “archaic” genera in Africa (Raven and Polhill, 1981). They were called a “western Gondwanan” family, and often were cited as an example of the biotic connection that existed between Africa and South America during the Cretaceous (65–145 mya), when these continents were in close proximity. Advances in our understanding of continental drift and the availability of more precise phylogenies for legumes have not supported the Gondwanan hypothesis, and most recent biogeographic studies have concentrated on Eocene (35–55 mya) or later events to explain legume distributions. During the Eocene, a land bridge in the North Atlantic joined Africa, Europe, and North America in one continuous land mass (Tiffney, 1985) whose climate was much warmer and wetter than it is today. The Pacific Northwest of North America harbored tropical rainforests, whereas Wyoming was covered with warm-temperate forest similar in composition to that seen today in southeast Asia. This mixed assemblage of plants, containing both temperate and tropical elements, was termed the boreotropical flora (Wolfe, 1975). North American legume fossils from around 50 mya include warm-temperate genera found in today’s forests of the southeastern

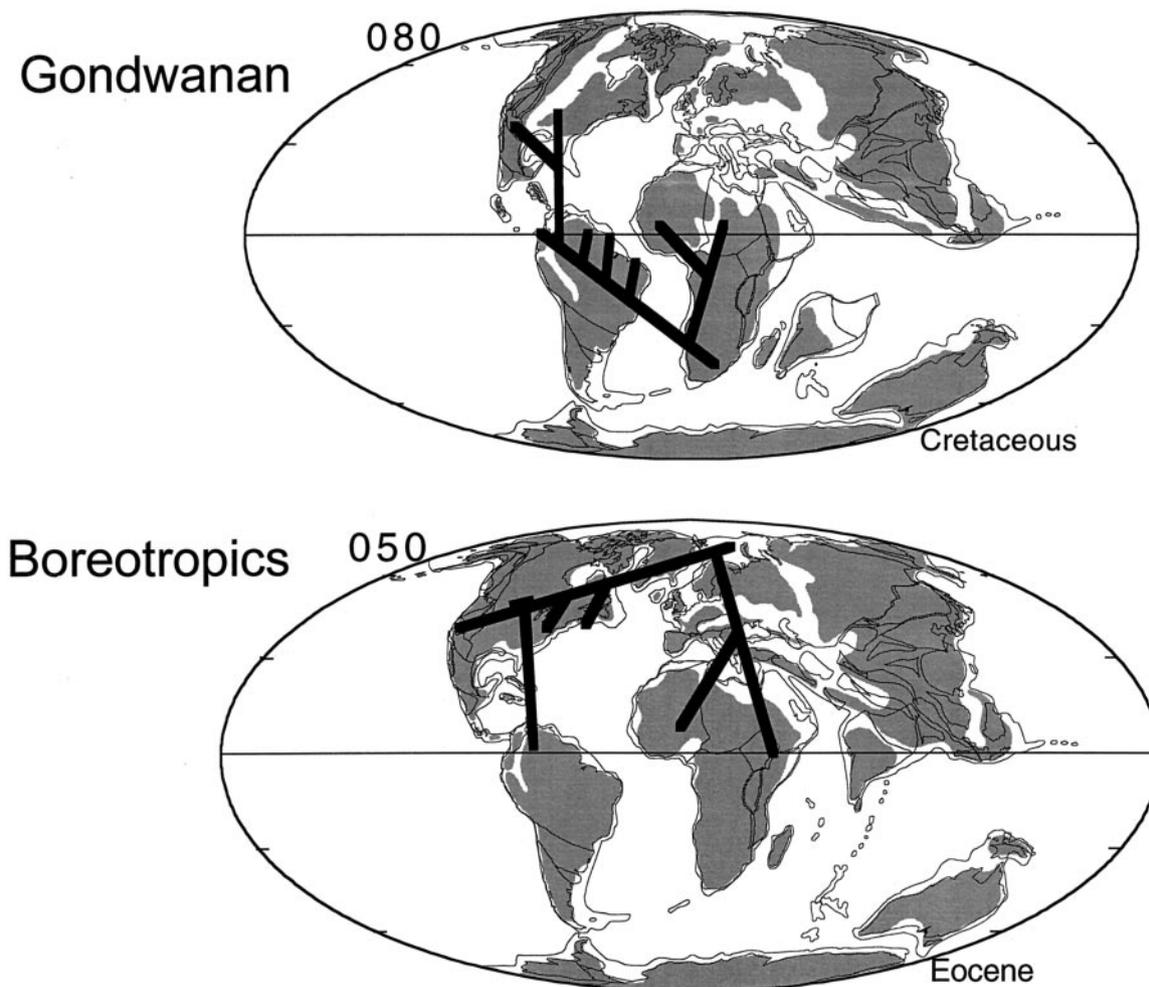


Figure 4. The two most commonly proposed biogeographic scenarios for legumes. The top figure, illustrating the Gondwanan hypothesis, shows a phylogenetic tree superimposed on continental positions from 80 mya. According to this hypothesis, the earliest branching lineages would occur in South America and Africa, and North American taxa would be relatively recent offshoots from South American ancestors. The bottom figure, illustrating the Boreotropics hypothesis, imposes a tree on continental positions from the Eocene (50 mya). Note that the early radiation of the family takes place in the northern hemisphere and that South American taxa are recent offshoots from North American ancestors.

United States such as *Cladrastis* (yellowwood), *Robinia* (locust), and *Gymnocladus* (Kentucky coffee tree) but also genera that today are exclusive to the tropics (Taylor, 1990) such as *Dalbergia* (rosewood), *Bauhinia* (orchid tree), and *Swartzia* (an early diverging papilionoid).

Thus, an alternative general explanation for the distribution of legumes hypothesizes that continental breakup and progressively cooler climates since the Eocene have led to the distribution patterns we see today (Fig. 4). Under this model, any archaic African taxa are interpreted as relicts of the once widespread boreotropical forest that have managed to survive the increased aridity and cooler temperatures that led to their extinction in the northern hemisphere (Lavin et al., 2000). This boreotropical hypothesis is also compatible with the finding that the Cercideae is one of the first legume clades to have differentiated: The boreotropical flora contains both *Cercis* and *Bauhinia*,

and *Cercis* (redbud) has persisted in the modern temperate forests of Europe, North America, and east Asia.

Phylogenetic evidence (e.g. Lavin and Luckow, 1993) and a reinterpretation of Eocene fossils (Taylor, 1990) also indicate that many putatively "archaic" South American taxa are in fact recent offshoots from northern hemisphere radiations; in contrast to the Gondwanan hypothesis, the direction of migration of legumes in the Americas is north to south rather than south to north (Fig. 4). Finally, the few phylogenetic studies that have used molecular approaches to date legume radiations have found them to have Tertiary rather than late Cretaceous ages (Wojciechowski, 2003, and refs. therein). There are unquestionably "archaic" lineages of legumes in South America as well as in Africa, but whether they are relictual boreotropical taxa pushed south by an increasingly inhospitable climate, Gondwanan ele-

ments that managed to surmount a large water barrier, or more recent examples of long-distance dispersal will probably never be known. What has become clear is that such distributions are the exception rather than the rule, and that a general historical explanation for legume distributions is best sought in events during and after the Tertiary.

HOW MANY TIMES HAS NODULATION EVOLVED IN LEGUMES?

Although for many people nodulation is likely to be the first thing that comes to mind when legumes are mentioned, nodulation is neither universal among legumes nor confined to the family. Instead, the pattern of distribution of nodulation is complex, and has thus far defied any simple explanation for its origin or origins within the family or among flowering plants (Doyle, 1998; Sprent, 2001). Nodulation symbioses with either rhizobia or actinorhizal bacteria are found in nine other families outside the Leguminosae, all of which, like the legumes, have been shown to belong to the eurosoid I group, suggesting that there arose in this group some factor enabling, but not committing, plants to form nitrogen-fixing symbioses (Soltis et al., 1995). Thus, within this overall clade, the nitrogen-fixing families do not appear to share a common ancestor, and nodulation has certainly arisen quite independently among—and even within—particular families.

Within the legumes themselves, nodulation occurs in more than 90% of papilionoid genera and just under that percentage of mimosoid genera, but it is much more rare (around 5%) among the diverse and unrelated genera referred to as Caesalpinioideae (Sprent, 2001). There is no evidence for nodulation at the base of the legume tree: first, because no Cercideae nodulate, and second, because none of the other families in the Fabales (Polygalaceae, Surianaceae, or *Quillaja*) are nodulators. Thus it can be postulated with some confidence that nodulation arose within the legumes and only after the earliest lineages had already diverged. Nor do the next-diverging lineages—regardless of their order of divergence (Fig. 3)—include the ancestors of the bulk of nodulating legumes. The Detarieae and Daliinae lineages each contain a single genus with an unconfirmed report of nodulation (Sprent, 2001), but in both cases, these genera are nested among other taxa that do not appear to nodulate (Bruneau et al., 2001), suggesting that they could represent additional independent origins of nodulation in the family.

It is at the next dichotomy that the action begins in earnest, with the Papilionoideae branching off from the remaining lineages, including Mimosoideae and many Caesalpinioideae. The mimosoid/caesalpinoid clade may itself contain several origins of nodulation. One origin certainly involves the core Mimosoideae plus at least some caesalpinoids (Sprent,

2001). *Chamaecrista*, a large caesalpinoid genus previously lumped with non-nodulating *Senna* and *Cassia* as *Cassia*, may well represent an independent origin of nodulation, and there are a handful of other nodulators that could represent still additional origins. Phylogenetic relationships of this entire clade, particularly at its base, require resolution by increased sampling of genes and key taxa, and more caesalpinoid genera need to be tested for their ability to nodulate.

The situation in Papilionoideae is also complex, again due to poor resolution in trees of the subfamily (Fig. 3), especially involving early-diverging lineages. All of the major lineages of the subfamily are overwhelmingly dominated by nodulating taxa, but many earlier diverging lineages do not appear capable of nodulating. However, the very first dichotomy in the subfamily appears to split a group of nodulating “swartzoid” genera from the remainder of the subfamily (Pennington et al., 2001). Thus it is likely that the first papilionoids had already evolved the capacity to nodulate, and the absence of nodulation in other early-diverging lineages is presumably due to evolutionary loss of this complex and energetically costly phenomenon, such as is known to have occurred elsewhere in the family (e.g. Luckow et al., 2003).

There is an enormous diversity of nodule types in the family, particularly among Papilionoideae, differing in morphology, anatomy, and chemistry (Sprent, 2001). As relationships among the family’s constituent lineages become resolved more precisely, it will be possible to track the evolution of this diversity in detail. Early advances in this area include the recognition that the very characteristic “aeschynomenoid” type of nodule typifies the newly identified aeschynomenoid/dalbergioid lineage (Lavin et al., 2001) and that determinate “desmodioid” nodules are likely to have evolved independently in the phaseoloid groups and in Loteae (Sprent et al., 2001). Knowing where in the phylogeny such innovations occurred should provide valuable clues concerning their biological significance.

Phylogenies of legume genera and species, for all of their value, cannot alone resolve how many times nodules evolved, and they cannot tell us what a nodule fundamentally is—a question that has yet to be answered (Hirsch and LaRue, 1998). An answer to that question at the molecular level should tell us how genes have been recruited to build nodules in different species, which may in turn provide us with new criteria for hypothesizing the number of origins of nodulation (Doyle, 1994, 1998). But even knowing what a nodule is will not tell us why nodulation was invented by legumes. Is the evolutionary success of legumes due to nodulation? It has been pointed out that if all nodulating taxa were removed from the family, the remainder would be a moderately sized, rather unremarkable tropical family. But nodulation

alone is not a guarantor of evolutionary success—none of the nonlegume families that nodulate begin to approach the legumes in size or diversity. Legumes, whether or not they nodulate, have a high-nitrogen “lifestyle,” involving high photosynthetic rates in short-lived leaves filled with nitrogen-rich defensive compounds; maybe this lifestyle is what makes the whole family so successful, and perhaps legumes evolved nodulation to feed this nitrogen habit more effectively (McKey, 1994).

GENOME SIZE EVOLUTION: SIZE MATTERS

Mendel may have been lucky—or crafty—in his choice of pea for his genetic experiments, but regrettably, he did not have access to a flow cytometer, and the large genome of his favorite plant has led to its being eclipsed as a legume model for genomic studies by relatives with genomes one-tenth its size, *M. truncatula* and *L. japonicus*. The other model legume, soybean, has a genome about twice as large as *M. truncatula* and *L. japonicus*, but is known to be an ancient tetraploid, so its fundamental genome might not be much larger than theirs. What of the rest of the family? Is the typical condition to have small genomes or large ones? Do legumes have “a one-way ticket to genomic obesity,” as has been hypothesized in grasses (Bennetzen and Kellogg, 1997), or do legume genomes expand and contract? Is polyploidy common or uncommon? Our growing knowledge of legume phylogeny will allow us to address these and other questions concerning genome evolution.

A search of the Royal Botanic Gardens, Kew, Plant DNA C-value database (<http://www.rbghkew.org.uk/cval/homepage.html>) for Leguminosae turns up around 500 records, ranging from under 400 to over 26,000 Mbp/1C. There are 23 values at or below the sizes given for *M. truncatula* and *L. japonicus* (466 Mbp). The smallest legume genome, at 368 Mbp, belongs to *Lablab niger* (a relative of the hyacinth bean), a member of the phaseoloid/millettioid clade. That group is also represented in the small genome list by three species of *Vigna* (cowpea and allies; 417–466 Mbp) and *Phaseolus macvaughii* (441 Mbp; *Phaseolus vulgaris* is somewhat larger at 588 Mbp). Other papilionoids with small genomes include five other *Lotus* spp. (441–466 Mbp) and two other members of Loteae, *Scorpiurus vermiculatus* (392 Mbp) and *Anthyllis vulneraria* (466 Mbp). Relatives of *Medicago* are also represented: two species of *Trifolium* (417–466 Mbp) and *Biserrula* (417 Mbp). The other two major papilionoid lineages have no members on this list, but the smallest value for the genistoid group is only 539 Mbp (*Cytisus nigricans*). The aeschynom-enoid/dalbergioid group is represented in the entire database only by two *Dalbergia* spp., which have considerably larger genomes (1,078–1,250 Mbp), and by *Arachis* spp., whose diploid ($2n = 20$) species have genomes that range from 1,480 to 3,381 Mbp, so it is

possible that this entire group has large genomes. No data are available from early diverging lineages of Papilionoideae, not even from commonly planted trees such as *Styphnolobium japonicum* (= *Sophora japonica*, pagoda tree); clearly this is an area where more data would be useful.

Relatively few genome size estimates are available for species outside of the Papilionoideae, but among the list of small genomes are several members of the mimosoid genus *Prosopis* (mesquite), whose diploid ($2n = 28$) species range from 392 to 490 Mbp. Other mimosoids also have relatively small genomes, such as *Acacia tortilis*, at 515 Mbp; interestingly, this $2n = 52$ species is listed as a tetraploid, making this effectively a 2C value and suggesting that the 1C value could be just over 200 Mbp, not a great deal larger than that of Arabidopsis. The few other members of the mimosoid/caesalpinoid clade that have C-value estimates include several in the 600 to 900 Mbp range, smaller than soybean. *Bauhinia purpurea*, a member of the earliest-diverging lineage in the family has a genome of only 588 Mbp. Moreover, at $2n = 28$, the genus *Bauhinia* is thought to be tetraploid relative to the genus *Cercis*, all of whose species are $2n = 14$ (Goldblatt, 1981), so it is possible that the original 1C value was much smaller. Obtaining genome size estimates for *Cercis*—a common temperate zone tree—would obviously be desirable. Thus it is at least conceivable that the earliest legumes had small genome sizes. The Kew C-value index has no records for Polygalaceae or other Fabales that could be used to optimize genome size on the legume tree; values would be welcome and not difficult to obtain for a widespread family such as Polygalaceae.

The above emphasis on small genome sizes may give a false impression of genome evolution in the family, which in fact is quite dynamic. Classic cases of rampant genome size evolution occur in the IRLC group of papilionoids: For example, *Lathyrus* has diploid ($2n = 14$) species with genomes listed from 3,357 to 14,308 Mbp. The new phylogeny of the Viciae and Trifolieae (Steele and Wojciechowski, 2003; Fig. 5) allows the low values of *Medicago* spp. and the much larger genomes of *Lathyrus* to be put into a phylogenetic context (Fig. 5). Optimization of “genome size” is difficult because it is not a discrete character (Bennetzen and Kellogg, 1997), but some tentative conclusions can be reached. The ancestor of *Lathyrus*, *Vicia*, *Lens*, and *Pisum* (Viciae) had a large genome ($>>1,000$ Mbp), whereas the ancestor of *Trifolium* appears to have had a considerably smaller genome. None of the species outside of these clades have genomes much in excess of 1,000 Mbp, including the two species from outside the vicioid clade that were used here as outgroups (*Biserrula* and *Colutea*), and this suggests that the ancestor of the entire vicioid group had a small genome, and that genomes of Viciae have increased greatly since diverging from their common ancestor with *Trifolium*.

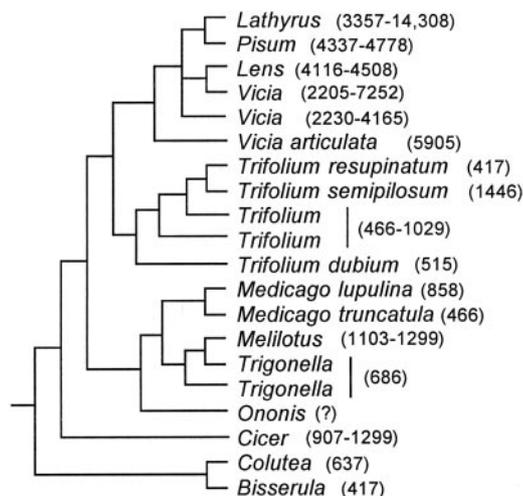


Figure 5. Genome size variation in *Medicago* spp. and allies. The tree topology is summarized from Steele and Wojciechowski (2003). Genome sizes from the Royal Botanic Gardens, Kew, C-value database are shown in parentheses next to generic or species names.

Genomes can expand not only incrementally, but also saltationally, by polyploidy, which is one of the most important forces in genome evolution of plants. From consideration of chromosome numbers, Goldblatt (1981) concluded that “all Mimosoideae, Papilionoideae, and Caesalpinioideae (except *Cercis*)... have most likely had a polyploid ancestry.” This hypothesis, which might once have been met with skepticism, no longer seems contentious, given the polyploid skeletons in the closets of *Arabidopsis* (Vision et al., 2000) and other “diploids.” Nevertheless, a careful consideration of chromosome number in a phylogenetic context could be very productive both for the family as a whole and in some of its groups, notably the cytologically complex genistoid group of Papilionoideae, where high basic numbers are common and are coupled with additional cycles of polyploidy and aneuploidy (Goldblatt, 1981).

CONCLUSIONS

The Leguminosae is an enormous family, incredibly diverse biologically, providing a rich source of potential experimental questions linking form with molecular function for virtually any topic in plant biology. *Medicago*, *Lotus*, and *Glycine* are related relatively closely to most of the major economically important legumes, but encompass only a small fraction of the overall diversity of the family. From giant caesalpinoid rainforest trees to tiny papilionoid annual herbs, legumes are united by descent from a single common ancestor. What makes a legume a legume and what makes the family so diverse are questions that are well worth asking and that can be answered only by working outward from the bridgehead provided by the current model legumes to explore the rest of the iceberg.

ACKNOWLEDGMENTS

We thank several of the authors of *Advances in Legume Systematics, Part 10* for making those and other unpublished papers available to us: Gwil Lewis, Beryl Simpson, Marty Wojciechowski, Kelly Steele, Matt Lavin, Gery Allan, Anne Bruneau, and Patrick Herendeen. Not all of these could be cited due to the limited number of references permitted. Kate VandenBosch, George Weiblen, Marty Wojciechowski, and an anonymous reviewer provided helpful comments on the manuscript.

Received November 22, 2002; returned for revision December 9, 2002; accepted December 29, 2002.

LITERATURE CITED

- Allan GJ, Zimmer EA, Wagner WL, Sokoloff DD (2003) Origin and classification of tribe Loteae: insight based on phylogenetic analysis of internal transcribed spacer regions of nrDNA. *In* BB Klitgaard, A Bruneau, eds, *Advances in Legume Systematics, Part 10*. Royal Botanic Gardens, Kew, UK (in press)
- Bennetzen JL, Kellogg EA (1997) Do plants have a one-way ticket to genomic obesity? *Plant Cell* **9**: 1509–1514
- Bruneau A, Forest F, Herendeen PS, Klitgaard BB, Lewis GP (2001) Phylogenetic relationships in the Caesalpinioideae (Leguminosae) as inferred from chloroplast *trnL* intron sequences. *Syst Bot* **26**: 487–514
- Crisp MD, Gilmore S, van Wyk B-E (2000) Molecular phylogeny of the genistoid tribes of papilionoid legumes. *In* P Herendeen, A Bruneau, eds, *Advances in Legume Systematics, Part 9*. Royal Botanic Gardens, Kew, UK, pp 249–276
- Doyle JJ (1994) Phylogeny of the legume family: an approach to understanding the origins of nodulation. *Annu Rev Ecol Syst* **25**: 325–349
- Doyle JJ (1998) Phylogenetic perspectives on nodulation: an evolving view of plants and symbiotic bacteria. *Trends Plant Sci* **3**: 473–478
- Doyle JJ, Chappill JA, Bailey CD, Kajita T (2000) Towards a comprehensive phylogeny of legumes: evidence from *rbcl* sequences and non-molecular data. *In* P Herendeen, A Bruneau, eds, *Advances in Legume Systematics, Part 9*. Royal Botanic Gardens, Kew, UK, pp 1–20
- Gentry AH (1993) Diversity and floristic composition of lowland tropical forest in Africa and South America. *In* P Goldblatt, ed, *Biological Relationships between Africa and South America*. Yale University Press, New Haven, CT, pp 500–547
- Goldblatt P (1981) Cytology and phylogeny of Leguminosae. *In* RM Polhill, PH Raven, eds, *Advances in Legume Systematics, Part 2*. Royal Botanic Gardens, Kew, UK, pp 427–463
- Haston EM, Lewis GP, Hawkins JA (2003) A phylogenetic investigation of the Peltophorum group (Caesalpinioideae: Leguminosae). *In* BB Klitgaard, A Bruneau, eds, *Advances in Legume Systematics, Part 10*. Royal Botanic Gardens, Kew, UK (in press)
- Herendeen PS, Bruneau A, Lewis GP (2003) Phylogenetic relationships in caesalpinoid legumes: a preliminary analysis based on morphological and molecular data. *In* BB Klitgaard, A Bruneau, eds, *Advances in Legume Systematics, Part 10*. Royal Botanic Gardens, Kew, UK
- Herendeen PS, Crepet WL, Dilcher DL (1992) The fossil history of the Leguminosae: phylogenetic and biogeographic implications. *In* PS Herendeen, DL Dilcher, eds, *Advances in Legume Systematics, Part 4: The Fossil Record*. Royal Botanic Gardens, Kew, UK, pp 303–316
- Hirsch AM, LaRue TA (1998) Is the legume nodule a modified root or shoot or an organ *sui generis*? *Crit Rev Plant Sci* **16**: 361–392
- Hu J-M, Lavin M, Wojciechowski MF, Sanderson MJ (2002) Phylogenetic analysis of nuclear ribosomal ITS/5.8 S sequences in the tribe Millettieae (Fabaceae): *Poecilanthus-Cyclotobium*, the core Millettieae, and the *Callerya* group. *Syst Bot* **27**: 722–733
- Kajita T, Ohashi H, Tateishi Y, Bailey CD, Doyle JJ (2001) *rbcl* and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and allies. *Syst Bot* **26**: 515–536
- Klitgaard BB, Bruneau A, editors (2003) *Advances in Legume Systematics, Part 10*. Royal Botanic Gardens, Kew, UK
- Lavin M, Luckow M (1993) Origins and relationships of tropical North America in the context of the boreotropics hypothesis. *Am J Bot* **80**: 1–14
- Lavin M, Pennington RT, Klitgaard BB, Sprent JL, de Lima HC, Gasson PE (2001) The dalbergioid legumes (Fabaceae): delimitation of a pantropical monophyletic clade. *Am J Bot* **88**: 503–533

- Lavin M, Thulin M, Labat J-N, Pennington RT** (2000) Africa, the odd man out: molecular biogeography of dalbergioid legumes (Fabaceae) suggests otherwise. *Syst Bot* **25**: 449–467
- Lee J, Hymowitz T** (2001) A molecular phylogenetic study of the subtribe Glycininae (Leguminosae) derived from the chloroplast DNA *rps16* intron sequences. *Am J Bot* **88**: 2064–2073
- Lewis GP, Schrire BD, Mackinder BA, Lock JM, editors** (2003) Legumes of the World. Royal Botanic Gardens, Kew, UK
- Luckow M, Miller JT, Murphy DJ, Livshultz T** (2003) A phylogenetic analysis of the Mimosoideae (Leguminosae) based on chloroplast DNA sequence data. *In* BB Klitgaard, A Bruneau, eds, *Advances in Legume Systematics, Part 10*. Royal Botanic Gardens, Kew, UK (in press)
- Luckow M, White PW, Bruneau A** (2000) Relationships among the basal genera of mimosoid legumes. *In* P Herendeen, A Bruneau, eds, *Advances in Legume Systematics, Part 9*. Royal Botanic Gardens, Kew, UK, pp 165–180
- McKey D** (1994) Legumes and nitrogen: the evolutionary ecology of a nitrogen-demanding lifestyle. *In* JI Sprent, D McKey, eds, *Advances in Legume Systematics, Part 5: The Nitrogen Factor*. Royal Botanic Gardens, Kew, UK, pp 211–228
- Pennington RT, Lavin M, Ireland H, Klitgaard B, Preston J, Hu J-M** (2001) Phylogenetic relationships of basal papilionoid legumes based upon sequences of the chloroplast intron *trnL*. *Syst Bot* **26**: 537–556
- Polhill RM, Raven PH, editors** (1981) *Advances in Legume Systematics, Part 1*. Royal Botanic Gardens, Kew, UK
- Raven PH, Polhill RM** (1981) Biogeography of the Leguminosae. *In* RM Polhill, PH Raven, eds, *Advances in Legume Systematics, Part 1*. Royal Botanic Gardens, Kew, UK, pp 27–34
- Simpson BB, Larkin LL, Weeks A** (2003) Progress towards resolving relationships of the Caesalpinia group (Caesalpinieae: Caesalpinioideae: Leguminosae). *In* BB Klitgaard, A Bruneau, eds, *Advances in Legume Systematics, Part 10*. Royal Botanic Gardens, Kew, UK (in press)
- Simpson GG** (1961) *Principles of Animal Taxonomy*. Columbia University Press, New York
- Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF et al.** (2000) Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Bot J Linn Soc* **133**: 381–461
- Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG** (1995) Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc Natl Acad Sci USA* **92**: 2647–2651
- Sprent JI** (2001) *Nodulation in Legumes*. Royal Botanic Gardens, Kew, UK
- Steele KP, Wojciechowski MF** (2003) Phylogenetic analyses of tribes Trifolieae and Viciae based on sequences of the plastid gene *matK* (Papilionoideae: Leguminosae). *In* BB Klitgaard, A Bruneau, eds, *Advances in Legume Systematics, Part 10*. Royal Botanic Gardens, Kew, UK (in press)
- Taylor DW** (1990) Paleobiogeographic relationships of angiosperms from the Cretaceous and early Tertiary of the North American area. *Bot Rev* **56**: 279–415
- Tiffney BH** (1985) The Eocene North American land bridge: its importance in Tertiary and modern phytogeography of the northern hemisphere. *J Arnold Arb Harv Univ* **66**: 243–273
- Vision TJ, Brown DG, Tanksley SD** (2000) The origins of genomic duplications in *Arabidopsis*. *Science* **290**: 2114–2117
- Wojciechowski MF** (2003) Reconstructing the phylogeny of legumes (Leguminosae): an early 21st century perspective. *In* BB Klitgaard, A Bruneau, eds, *Advances in Legume Systematics, Part 10*. Royal Botanic Gardens, Kew, UK (in press)
- Wojciechowski MF, Sanderson MJ, Steele KP, Liston A** (2000) Molecular phylogeny of the “temperate herbaceous tribes” of papilionoid legumes: a supertree approach. *In* P Herendeen, A Bruneau, eds, *Advances in Legume Systematics, Part 9*. Royal Botanic Gardens, Kew, UK, pp 277–298
- Wolfe JA** (1975) Some aspects of plant geography of the northern hemisphere during the late Cretaceous and Tertiary. *Ann MO Bot Gard* **74**: 785–803