

CURRENT REVIEW

Phylogenetic Perspectives on the Origins of Nodulation

Jeff J. Doyle

Department of Plant Biology, 412 Mann Library Building, Cornell University, Ithaca, NY 14853, U.S.A.

Submitted 9 May 2011. Accepted 21 June 2011.

Recent refinements to the phylogeny of rosid angiosperms support the conclusion that nodulation has evolved several times in the so-called N_2 -fixing clade (NFC), and provide dates for these origins. The hypothesized predisposition that enabled the evolution of nodulation occurred approximately 100 million years ago (MYA), was retained in the various lineages that radiated rapidly shortly thereafter, and was functional in its non-nodulation role for at least an additional 30 million years in each nodulating lineage. Legumes radiated rapidly shortly after their origin approximately 60 MYA, and nodulation most likely evolved several times during this radiation. The major lineages of papilionoid legumes diverged close to the time of origin of nodulation, accounting for the diversity of nodule biology in the group. Nodulation symbioses exemplify the concept of “deep homology,” sharing various homologous components across nonhomologous origins of nodulation, largely due to recruitment from existing functions, notably the older arbuscular mycorrhizal symbiosis. Although polyploidy may have played a role in the origin of papilionoid legume nodules, it did not do so in other legumes, nor did the pre-rosid whole-genome triplication lead directly to the predisposition of nodulation.

How novel organs come into being is one of the central questions of evolutionary developmental biology. Complex structures such as the camera-type eye or the angiosperm flower pose fascinating questions to evolutionary biologists and philosophers alike—how can a complex organ originate, and is its complexity, to use the term coined by the “intelligent design” camp, “irreducible”? When such complex organs evolve multiple times independently in different lineages, additional intriguing questions arise.

Nitrogen-fixing nodules, though less widely known outside the plant sciences, pose the same kinds of questions as do eyes or flowers. What is a nodule, fundamentally, and what is its origin? Hirsch and Larue (1997) asked whether the nodule is a transformed root or shoot, or a novel organ in its own right. More recently, Markmann and Parniske (2009) asked, “How novel are nodules?” Such questions bring into play a widely misunderstood and often misused word: “homology,” which evolutionary biologists define as “similarity due to common descent.” Similarity, whether in function, morphology, anatomy, or gene expression, is required for hypothesizing homology but similarity is not synonymous with homology in this evolutionary sense of the term. Only when similar traits are inherited from a common ancestor is the trait homologous across

taxa. Unlike similarity, homology is an all-or-nothing term—a trait is homologous in two or more taxa if inherited in its original or derived form from their common ancestor, or nonhomologous if it arises independently by parallel or convergent evolution. The classic example of wings in birds and bats illustrates this point: “wings” are a nonhomologous trait, being functionally analogous structures derived independently in two phylogenetically distinct vertebrate lineages whose common reptilian ancestor lacked wings. It also illustrates the point that homology is a hierarchical term and depends on the level at which it is being considered and how the trait is defined. Thus, as wings, these appendages are nonhomologous but, as modifications of the ancestral vertebrate forelimb, they are homologous, along with the arms of humans.

Thus, the assessment of homology requires knowledge of phylogenetic relationships. Advances in systematics and evolutionary biology have continued to refine our understanding of angiosperm phylogeny, with implications for understanding the evolution of nodulation.

Nonhomologous origins of nitrogen-fixing symbioses.

Early angiosperm molecular phylogenies provided insights into the evolution of nodulation by revealing that families of nodulating plants such as Casuarinaceae and Leguminosae were much more closely related than had been hypothesized based on morphology. Root nodule symbioses (RNS) turned out to be confined to a single large clade termed the N_2 -fixing clade (NFC) (Soltis et al. 1995).

Nodulation occurs in all four orders that compose the NFC (Fig. 1A). However, nodulation is the exception in all of these orders except Fabales where, although only one of the four families contains nodulators, that family, the Leguminosae (or Fabaceae), is the third largest family of flowering plants and is dominated by nodulators among its 700+ genera and approximately 20,000 species. Elsewhere in the NFC, families in which nodulation is universal or widespread are mostly small, such as the Elaeagnaceae and the monotypic Coriariaceae (Swensen and Benson 2008). At the other extreme, among the 100+ genera and 3,000+ species of the large family Rosaceae, only nine species in five genera nodulate. Hypothesizing homology of rosid nodulation is unparsimonious, because a single origin in the common ancestor of the NFC with transmission to each of the phylogenetically diverse families in which nodulation occurs would require many independent losses. Moreover, nodules in different lineages are structurally and anatomically quite different from one another (Pawlowski and Sprent 2008) (for example, legume nodules are anatomically more similar to shoots whereas other nodules are more similar to roots); therefore, the similarity criterion is not met on detailed examination.

Nodulation symbioses originated relatively recently in angiosperm history.

Recent studies (Bell et al. 2010; Wang et al. 2009) have resolved previously ambiguous relationships in the rosids, although questions remain (Burleigh et al. 2011); dates from Bell and associates (2010) are used throughout this article. Within the eudicot lineage, the rosids appeared approximately 110 million years ago (MYA), 60 to 90 million years (MY) after the origin of angiosperms, and underwent a rapid diversification, splitting into several major lineages within a few million years. At approximately 105 MYA, a common ancestor gave rise to the large class of malvids (formerly eurosid II, which includes *Arabidopsis*) and fabids (eurosid I). Modern fabids radiated from a common ancestor approximately 100 MYA, splitting soon after into two major clades. One of these comprises three orders, including Malpighiales, of which poplar is a member. The other clade diversified from this common ancestor into the NFC (Fig. 1A); it was this NFC ancestor that is presumed to have evolved the still-unknown predisposition that allowed nodulation to evolve independently in several of its descendant lineages.

On a strict parsimony basis and with no losses hypothesized, at least nine origins of nodulation would be inferred outside of the legumes (Fig. 1A, boxes). The major groupings listed by Swensen and Benson (2008) (Fig. 1A, groups A to E plus *Parasponia*) could be taken as a more conservative estimate of the number of origins, allowing for the ease with which nodulation could be lost evolutionarily but also accounting for extensive structural differences in nodules among these groups that are consistent with independent derivation from non-nodulating ancestors.

How recently RNS arose in angiosperm history depends on the number of origins assumed but, using even the more conservative estimates for the number of origins and maximum ages for nodulation in each clade, no group is inferred to have evolved its symbiosis earlier than approximately 70 MYA (Fig. 1A). Thus, over 30 MY elapsed between the common ancestor that acquired the nodulation predisposition and the evolution of nodulation in the earliest nodulating lineage. Retention over such a long period of time suggests that the trait was functional in non-nodulators and was maintained by purifying natural selection. Moreover, its retention independently for millions of years in different lineages prior to recruitment for a similar function (nodulation) in RNS lineages suggests that the trait remained similar across lineages. Thus, this non-nodular function should still exist, and should occur throughout the

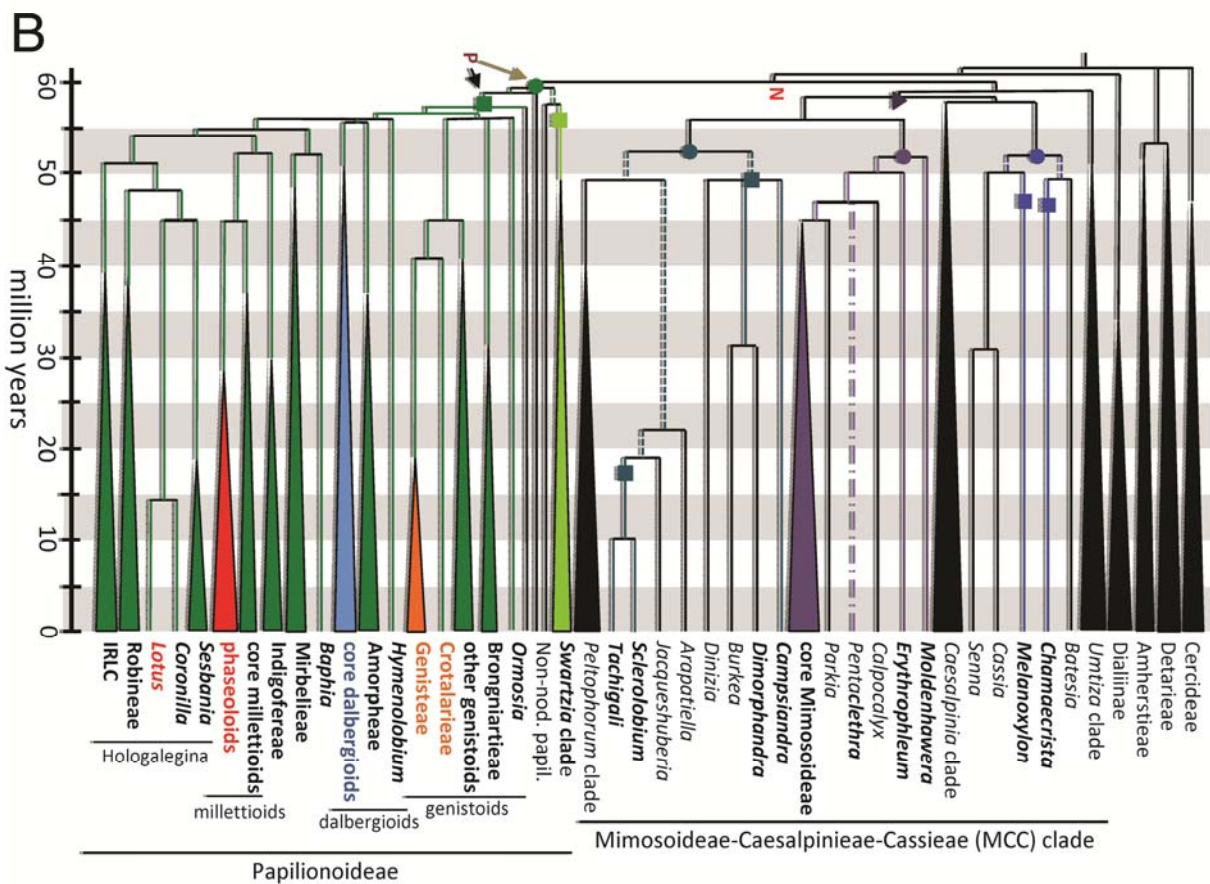
NFC in non-nodulating species; its non-nodulation function should also be maintained in nodulating species, unless its newer function in nodulation made its previous function superfluous.

Origins and evolution of nodulation within legumes.

Molecular phylogenies have also revolutionized our understanding of relationships in legumes, leading to a reorganization of traditional genera and tribes, and confirming the unnaturalness (non-monophyly) of one of the three subfamilies (Caesalpinioideae) which is now best appreciated as a grade comprising the earliest branching lineages of the family (Bruneau et al. 2008; Lewis et al. 2005). The majority of legumes nodulate but nodulation is absent from the earliest diverging lineages of the family. Thus, nodulation evolved in legumes sometime after the origin of the family approximately 60 MYA (Fig. 1B) (Bruneau et al. 2008; Lavin et al. 2005). Nodulation occurs in the two largest clades: i) the monophyletic subfamily Papilionoideae and ii) the Mimosoideae-Caesalpineae-Cassieae clade (MCC) comprising the monophyletic subfamily Mimosoideae, the paraphyletic tribe Caesalpinieae (which includes such genera as *Gleditsia* and *Gymnocladus*), and the core Cassieae (which includes the genus *Chamaecrista*). However, nodulation is not universal in either clade. In the papilionoids, although most species nodulate, some of the earliest-diverging lineages appear to be incapable of nodulation. Lack of resolution in that part of the phylogeny and absence of nodulation data for several genera makes it impossible at present to determine whether there has been more than one origin of nodulation in this largest subfamily of legumes (Fig. 1B).

In the MCC clade, the situation is even more complex, and is also complicated by lack of data on nodulation for many caesalpinoid genera (Sprent 2009) and lack of strong resolution in phylogenetic trees (Bruneau et al. 2008). Nodulation is prevalent only in the clade dominated by the subfamily Mimosoideae (acacias, mimosas, and so on), over 90% of whose species nodulate. As in papilionoids, the situation in the earliest diverging lineages is uncertain, with two nodulating caesalpinoid genera (*Moldenhawera* and *Erythrophleum*) diverging early, followed by non-nodulating members of Mimosoideae (Bruneau et al. 2008; Lavin et al. 2005) (Fig. 1B). The Mimosoideae clade is sister to a clade that includes four other nodulating caesalpinoid genera. However, this clade is dominated by non-nodulating taxa and it diverged from the Mimosoideae clade 56 MYA. Therefore, it could represent one or more addi-

Fig. 1. Chronograms showing the distribution of nodulation symbioses in **A**, the N_2 -fixing clade (NFC) and **B**, the legumes. Nodulating taxa are in bold type and their lineages are colored. **A**, NFCs. Nodes marked A through E are Swensen and Benson's (2008) designations for nodulating lineages; circles associated with letters mark maximum ages for origins of nodulation, with an additional origin in *Parasponia*. Possible additional origins of nodulation within these lineages are indicated by colored boxes, which are placed a short distance from the base of each lineage to denote the fact that the origin could be anywhere along that branch; dashed lines denote uncertainty about the date of origin of nodulation between the oldest (circles) and younger (boxes) dates. The origin of the predisposition for nodulation is indicated by a red star. Red lines trace the minimal retention of this predisposition from its origin to the oldest ancestor of each nodulating lineage: the predisposition must have been retained from its origin at least to the common ancestor of each labeled nodes (e.g., node B in Fagales) and *Parasponia*. Topology and dates are from Bell and associates (2010) (minimum ages of their general ranges of 1 to 4 million years are used here), with the exceptions of *Parasponia* and the nodulating species of Rosaceae, which were not included in their analysis. Consequently, dates for these taxa are approximate. **B**, Legumes. The placement of a single origin of nodulation in the family is indicated by the red "N". Potential additional origins are shown using other shapes, as in A. "Non-nod. papil." comprise a diverse, paraphyletic group of papilionoids apparently unable to nodulate. The situation around core species of Mimosoideae is particularly complex, with *Pentaclethra* having both nodulating and non-nodulating species. The great majority of nodulating species have indeterminate nodules, presumably, in most cases, formed by root hair infection (Sprent 2008). Exceptions, all from papilionoids, are colored: determinate "desmodioid" nodules with root hair infection in phaseoloids (e.g., soybean) and *Lotus* spp.; "aeschynomeneoid" nodules with crack infection in core dalbergioids (e.g., peanut); indeterminate nodules of various types (including collar nodules) with epidermal infection in the tribe *Genisteae* (e.g., lupin) and possibly the related genistoid tribe *Crotalariae*. "P" denotes the papilionoid polyploidy event; the black arrow indicates the minimum age, with the gray arrow marking the maximum possible age. Topology and dates are from Lavin and associates (2005) for papilionoids and Bruneau and associates (2008) for caesalpinoids. The base of the tree as shown is a compromise between these two studies, which differ in some details; for example, Lavin and associates (2005) place the tribe Cercideae as sister to remaining legumes whereas Bruneau and associates (2008) place Amherstieae/Detarieae in that position. Representative genera are selected from both analyses. The depth of triangles denoting clades from these studies indicates the deepest branching for each clade.



tional origins of nodulation. Still more distantly related is the *Cassia* clade, a group that diverged approximately 53 MYA and which includes the remaining two nodulating caesalpinioids. These are the large, widespread genus, *Chamaecrista*, and the monotypic Brazilian genus, *Melanoxylon*, which are not sister taxa in Bruneau and associates (2008) and last shared a common ancestor approximately 50 MYA (Bruneau et al. 2008); each could represent an additional origin of nodulation. Thus, parsimony considerations suggest there could have been five to six independent origins of nodulation within the mimosoid superclade, so that there may have been six to seven separate origins of nodulation in the legumes as a whole.

Given the complexity of nodulation, multiple origins may be less likely, albeit more parsimonious, than a scenario involving many independent losses (Doyle 1994). However, it is clear that nodulation can evolve independently—it has done so numerous times in the NFC. Moreover, in the entire papilionoid clade, only a few cases are known of species that cannot nodulate, and none are reported from the Hologalegina or millettoid clades, which together comprise 263 genera and over 8,000 species (Lewis et al. 2005). The pattern near the base of the *Mimosoideae* may involve losses (Sprent 2009) but, at least in some lineages, loss of nodulation may not be particularly likely.

Early rapid radiations of legumes and the evolution of diverse nodule types.

As with both rosids as a whole and with the NFC in particular, the early history of the legume family and its major lineages is one of rapid radiation (Fig. 1B) (Lavin et al. 2005). Thus, the major lineages have been separated for far longer than the period during which they could have shared a common history of nodulation. Therefore, the existence of major differences in nodule development, morphology, biochemistry, and molecular biology among these lineages (den Herder and Parniske 2009) is not surprising.

The widespread distribution of indeterminate nodules (Fig. 1B) suggests that, if all legume nodules are homologous, this is the ancestral condition. If multiple origins are inferred, then structurally similar indeterminate types in different lineages are nonhomologous, and their distribution then suggests that this type of nodule may be the easiest type to form or may be more versatile than shorter-lived determinate nodules (Sprent 2008). Determinate nodules are derived from indeterminate types and arose independently and at different times in legume evolution. The determinate “aeschynomenoid” nodule type found in peanut (groundnut, *Arachis hypogea*) and its relatives may be as old as 50 MY. In contrast, the “desmodioid” nodules of *Lotus* spp. and of phaseoloid nodules are nonhomologous, more recent in origin, and appear to be parallel innovations in these two different lineages (Fig. 1B). Their nonhomology is reflected in biochemical differences between them.

Root hair versus intercellular infection: which was the ancestral condition in legumes?

A difference among legume nodules with important evolutionary implications involves the earliest stages of infection and rhizobial growth. The best-characterized infection process involves root hairs, and is a common mechanism among the estimated 75% of all legumes that form infection threads (Sprent 2008). However, among papilionoids, peanut and other legumes with aeschynomenoid nodules are infected through cracks or wounds, and epidermal infection is common in many members of the genistoid clade. Some genera use both root hair and direct infection (e.g., *Sesbania*) (Capoen et al. 2010). It has been hypothesized that direct intercellular infection was a precursor to both crack and root hair infection (Sprent 2007; Sprent and

James 2007), and Madsen and associates (2010) have suggested that this hypothesis is supported by their studies of spontaneous nodulation mutants in which the various components of the infection and nodule organogenesis processes are decoupled.

Unfortunately, testing this hypothesis requires identifying modes of infection in a number of taxa for which little information about nodulation is available, as well as knowing the number of independent origins of nodulation. Information on members of *Ormosia*, Brongniartieae, and other genistoids is needed to determine whether epidermal entry is ancestral in the genistoid group as a whole, and information on members of *Amorpheae* and *Hymenolobium* could indicate whether crack entry is ancestral in the dalbergioid clade (Fig. 1B). If non-root hair infection is shown to be ancestral in these two groups, then it was likely ancestral for the core papilionoids as a whole.

Homoplasy, deep homology, and recruitment in nodulation.

A nonhomologous trait arising independently in two lineages through parallel or convergent evolution is called a homoplasy, and such features are of great interest in evolutionary biology (Sanderson and Hufford 1996; Wake et al. 2011). As more is learned about the details of how complex organs are constructed, it becomes increasingly apparent how thrifty evolution is. This realization underlies the concept of “deep homology” (Shubin et al. 2009), which simply means that, although a complex organ may evolve independently in multiple lineages (and thus be nonhomologous), each origin may involve assembly from homologous components. Camera-type eyes have evolved independently in mollusks and vertebrates and, thus, are nonhomologous. But the ancestors of these taxa independently recruited homologous cell types and signaling networks inherited from their remote common ancestor to construct their complex eyes. RNS is a homoplasy among lineages A to E (Fig. 1A). Yet, as with eyes, there are deep homologies across lineages.

Most “nodulin” proteins—even the archetypical nodulin, leghemoglobin, once thought to be unique to legumes—have been shown to be encoded by genes that are not unique to nodulating plants. The current understanding is summarized by Markmann and Parniske (2009): “The emergence of RNS seems to be based on the evolutionary concept of recruitment, in that many RNS-related processes rely on conserved genetic programs that also support preexisting plant developmental processes and that later acquired additional roles in RNS.” Thus, it is not surprising that, as the molecular underpinnings of nodulation are dissected, it is discovered that the genes underlying every aspect of nodulation, from the earliest signaling stages to senescence in indeterminate nodules, have roles elsewhere in the same plant species (Yokota and Hayashi 2011). Given the thriftiness of evolution, it is not surprising to find that the functional similarities between infection thread growth and tip growth in pollen tubes, root hairs, and other structures are underlain by homologous genes involved in cytoskeleton repatterning (Yokota et al. 2009), or that resistance genes (Yang et al. 2010) or defensin-like proteins (van de Velde et al. 2010) and their secretory apparatus (Wang et al. 2010) are involved in the interaction of the host plant with their microsymbionts.

Molecular evolution of recruited genes.

Recruitment, in the strictest and simplest sense, involves a gene acquiring a new context for its existing function as part of a novel phenomenon such as nodulation. This would result in a single gene being involved in two different processes, such as root and nodule development. Alternatively, gene duplication could be involved: if duplication preceded the origin of nodulation, then potentially only one copy would need to be recruited for nodulation, resulting in “regulatory neofunctionali-

zation” (Tirosh and Barkai 2007) to produce a “nodulin” paralogue. A duplication coinciding with the origin of nodulation would be a candidate for a causative innovation enabling nodulation to evolve; a still-earlier duplication could provide a trait “predisposing” a lineage for nodulation. The novelty of nodulation might rest in the new associations of different components, with the underlying genes retaining their original functions; alternatively, in some cases, neofunctionalization of paralogues could generate new catalytic functions or regulatory associations.

Duplication of a gene having both nodular and non-nodular functions (by definition occurring after the origin of nodulation) could partition these functions between the two paralogues, resulting in a nodular and a non-nodular copy. Such a subfunctionalized gene could underlie refinements of nodulation but the subfunctionalization event could not itself be critical for the origin of nodulation.

In theory, the elucidation of the “original” functions of genes expressed in nodules could identify the sources from which recruitment occurred. Buried in the phylogenies of complex gene families such as those of Nod26-like intrinsic proteins (Liu et al. 2009), *LysM* (Zhang et al. 2009), or plant globins (Gopalasubramanian et al. 2008) is information about their recruitment; however, extracting that information is not trivial for a variety of technical and biological reasons (Philippe et al. 2011; Studer and Robinson-Rechavi 2009). Also, gene expression in multiple tissues represents a continuum, and it is seldom easy to score genes as “on” or “off” in a particular tissue; up- or downregulation is a commonly used criterion but there is no reason why a gene recruited for nodulation would need to be upregulated relative to its expression level in the tissue from which the gene was recruited.

In theory, it should be possible to test whether nodulation is homologous in two lineages by determining whether genes with common roles across lineages are homologous and, if so, orthologous (Doyle 1994). Homologous nodulation symbioses originated, by definition, in a common ancestor and, therefore, involved only a single recruitment event for each gene, so that all genes in homologous nodulation symbioses should not only be homologous but also orthologous (derived by speciation, not duplication). Nonhomologous nodulation symbioses could involve convergent recruitment of nonhomologous genes to perform similar functions, or parallel recruitment of either orthologues or paralogues. The finding that nonhomologous or paralogous genes perform common roles in nodules of two lineages provides evidence of nonhomology (Doyle 1994).

The above approach is most likely to produce definitive evidence concerning nonhomology if genes are recruited individually, in which case each gene family provides one or more independent tests of nodule homology. Nonindependence becomes a problem when the unit of recruitment is an entire biochemical or developmental network as opposed to an individual gene. Nodulation involves the integration of various developmental processes, with separable stages controlled by different sets of genes (Groth et al. 2010; Moreau et al. 2011), but how modular the evolution of nodulation was remains an open question.

The classic case of recruitment of an entire group of functionally related genes is the common symbiosis (Sym) pathway, controlling the early signaling stages of nodulation (Markmann and Parniske 2009; Oldroyd et al. 2009). It has been known for some time now that, in legumes, the genes of this pathway were recruited from the much older arbuscular mycorrhizal symbiosis, and new details continue to be reported (Groth et al. 2010). Just how interrelated these symbioses are has been shown recently by the characterization of previously elusive mycorrhizal (Myc) factors which, like Nod factors, are lipochitooligosaccharide (LCO) compounds but are structur-

ally simpler (Maillet et al. 2011). Maillet and associates (2011) also show that the interconnection between the Nod and Myc pathways subsequent to elicitation by these LCO is stronger than previously recognized.

A key component of the Sym pathway, *SymRK*, has been recruited for nodulation independently by species of *Casuarina* (Gherbi et al. 2008), *Datisca* (Markmann et al. 2008), and *Parasponia* (Op den Camp et al. 2011). Hocher and associates (2011) have shown that many other genes from the common symbiosis pathway are expressed in the genera *Alnus* and *Casuarina*. Even more interestingly, they show that genes from the downstream nodulation-specific pathway known previously only from legumes are also expressed during nodulation in these actinorhizal plants. Given the overwhelming evidence that actinorhizal and legume–rhizobial symbioses are nonhomologous, it seems likely that these entire pathways have been recruited independently multiple times. This clearly poses problems for the use of individual gene orthology as a criterion for nodule homology. Regardless, the presence of the nodulation signaling pathway in species that diverged 100 MYA (Fig. 1B) suggests functionality in non-nodulating members of the NFC. Perhaps, as Hocher and associates (2011) suggest, this is the nodulation predisposition.

Polyploidy as a source of nodulation genes.

An obvious source of duplicate genes is polyploidy, in which the entire genome is duplicated. There is increasing evidence that polyploidy has occurred throughout the history of seed plants (Jiao et al. 2011; Soltis et al. 2009), and polyploidy is thought to play a major role in the origin of evolutionary novelty (Freeling and Thomas 2006). What role might polyploidy have played in the evolution of nodulation?

There is a hypothesized whole-genome triplication (WGT) that is shared by all rosids, including *Arabidopsis* and *Populus* spp.; by asterids (e.g., tomato and cacao) (Jaillon et al. 2007; Tang et al. 2008); and perhaps by even more angiosperms (Soltis et al. 2009). Therefore, the WGT was not unique to the ancestor of the NFC and, indeed, preceded the emergence of that ancestor by tens of millions of years, and the ancestors of each of the nodulating lineages by at least another 30 MY, as noted above. Certainly, any genes preserved in duplicate or triplicate from that event were available to be recruited for nodulation (or the predisposition) but, given the long periods of time, it seems unlikely that any of the many models proposed to preserve paralogues (Innan and Kondrashov 2010) could have involved selection pressure involving nodulation or its predisposition. Gene phylogenies do provide a possible criterion for identifying candidate genes for the predisposition: a gene whose function has been retained by the diverse nodulating lineages before being recruited by each should be orthologous across the diverse groups.

Cryptic polyploidy (polyploidy not currently accompanied by higher chromosome numbers) can be detected from genomic sequencing or from consideration of peaks in the synonymous distance (K_s) distributions of duplicated gene pairs (Blanc and Wolfe 2004; Schlueter et al. 2004), but these approaches have not been applied as yet to nodulating species outside the legumes. Consequently, it is not known whether polyploidy could have facilitated the evolution of nodulation in non-legume nodulators, though this is being remedied by the 1KP project, which includes several key nodulating taxa.

A series of studies have revealed that most papilionoid legumes are derived from a polyploid ancestor (Bertioli et al. 2009; Cannon et al. 2006; Pfeil et al. 2005) (Fig. 1B). Papilionoids are cryptic polyploids, typically with chromosome numbers of $n = 7$ to 11. Critical species from the early diverging lineages of the papilionoids have yet to be surveyed for

cryptic polyploidy; therefore, it is not known whether all papilionoids are fundamentally polyploid. The ancient (approximately 58 MYA) papilionoid WGD could coincide with the origin of nodulation in the main radiation of papilionoids, and could have provided the raw material for this evolutionary innovation. It is now clear, however, that polyploidy is not a prerequisite for the evolution of nodulation in all legumes. A phylogenomic study did not detect any evidence that the genus *Chamaecrista* shares the papilionoid WGD, suggesting that the ancestor of the entire MCC clade was not polyploid (Cannon et al. 2010). Nor did species of *Chamaecrista* show evidence of more recent polyploidy. Genomic evidence is as yet lacking for other nodulating caesalpinoid or mimosoid legumes but several key nonpapilionoid legumes as well as early-diverging papilionoids are also included in the 1KP project.

If papilionoid nodulation is homologous with nodulation in other legumes, the papilionoid WGD could have provided genes for modifying and refining the symbiosis. The recent publication of the soybean genome (Schmutz et al. 2010) provides an opportunity to study the relatively early (<15 MY) stages of genome fractionation following polyploidy. Fates of individual genes, gene families, and functional groups vary considerably for genes involved in photosynthesis (Coate et al. 2011), and the same may also be true for nodulation. Initial reports of functional divergence include the finding of Indrasumunar and associates (2010) that homologous copies of the putative Nod-factor receptor kinase, GmNFR1, have diverged in their expression. Libault and associates (2010), in their transcriptomic study of inoculated and mock-inoculated root hairs and stripped roots, found overall conservation of expression of homologous copies but numerous instances of differential expression between homologues. The *Medicago truncatula* genome project will provide further insights into the involvement of genes duplicated by the papilionoid WGD in the evolution of nodulation in the papilionoids.

So, what are nodules and are they novel?

The phylogenetic distribution of nodulation (Fig. 1A and B) provides two seemingly contradictory answers to one facet of Markmann and Parniske's (2009) question about the novelty of nodules. In one sense, not only are nodules novel but the nodules of legumes and those of species of *Parasponia*, *Casuarina*, and members of several other NFC clades are also each individually novel, originating independently of one another (Swensen and Benson 2008), quite possibly with several origins of this novel organ just within legumes. On the other hand, the more times nodules have originated independently, the less novel each individual nodulation symbiosis becomes.

Their question, however, has more to do with features shared by nodules and mycorrhizal symbioses or pathogen responses. And this connects it to the question asked by Hirsch and Larue (1997): "Is the nodule a transformed root or shoot, or an organ sui generis?", which is a question about serial homology—homology that involves structures on the same organism. In any case, these alternatives are not necessarily mutually exclusive. If the developmental program of the root were modified so as to produce the nodule, the nodule would still be a novel organ; a sepal may be a modified leaf but it is still a novel organ in its own right, and not merely a modified leaf. The concept of deep homology (Shubin et al. 2009) connects these various threads by providing the context of evolutionary processes that are thrifty, recruiting existing gene networks, cell types, and processes to build novelty. Thus, not only do nonhomologous, independently originating nodulation symbioses in diverse species share deep homologies but they also all share still deeper homologies with other plant processes, notably with the ancient arbuscular mycorrhizal symbiosis.

ACKNOWLEDGMENTS

I thank the many people with whom I have discussed nodule evolution over the years, particularly J. Sprent and S. Cannon and, lately, the nodulation authors of the *Medicago truncatula* genome project; two anonymous reviewers for their helpful comments; and the United States National Science Foundation for long-term support of my research on legumes, including current awards from IOS (0744306, 0822258, and 0939423) and DEB (0948800).

LITERATURE CITED

- Bell, C. D., Soltis, D. E., and Soltis, P. S. 2010. The age and diversification of the angiosperms re-visited. *Am. J. Bot.* 97:1296-1303.
- Bertioli, D., Moretzsohn, M., Madsen, L., Sandal, N., Leal-Bertioli, S., Guimaraes, P., Hougaard, B., Fredslund, J., Schauser, L., Nielsen, A., Sato, S., Tabata, S., Cannon, S., and Stougaard, J. 2009. An analysis of synteny of *Arachis* with *Lotus* and *Medicago* sheds new light on the structure, stability and evolution of legume genomes. *BMC Genomics* 10:45.
- Blanc, G., and Wolfe, K. H. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16:1667-1678.
- Bruneau, A., Mercure, M., Lewis, G. P., and Herendeen, P. S. 2008. Phylogenetic patterns and diversification in the caesalpinoid legumes. *Botany* 86:697-718.
- Burleigh, J. G., Bansal, M. S., Eulenstein, O., Hartmann, S., Wehe, A., and Vision, T. J. 2011. Genome-scale phylogenetics: Inferring the plant tree of life from 18,896 gene trees. *Syst. Biol.* 60:117-125.
- Cannon, S. B., Sterck, L., Rombauts, S., Sato, S., Cheung, F., Gouzy, J., Wang, X., Mudge, J., Vasdewani, J., Scheix, T., Spannagl, M., Monaghan, E., Nicholson, C., Humphray, S. J., Schoof, H., Mayer, K. F. X., Rogers, J., Quetier, F., Oldroyd, G. E., Debelle, F., Cook, D. R., Retzel, E. F., Roe, B. A., Town, C. D., Tabata, S., Van de Peer, Y., and Young, N. D. 2006. Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes (vol 103, pg 14959, 2006). *Proc. Natl. Acad. Sci. U.S.A.* 103:14959-14964.
- Cannon, S. B., Ilut, D., Farmer, A. D., Maki, S. L., May, G. D., Singer, S. R., and Doyle, J. J. 2010. Polyploidy did not predate the evolution of nodulation in all legumes. *PLoS ONE* 5:e11630. Published online.
- Capoen, W., Oldroyd, G., Goormachtig, S., and Holsters, M. 2010. *Sesbania rostrata*: A case study of natural variation in legume nodulation. *New Phytol.* 186:340-345.
- Coate, J. E., Schlueter, J. A., Whaley, A. M., and Doyle, J. J. 2011. Comparative evolution of photosynthetic genes in response to polyploid and nonpolyploid duplication. *Plant Physiol.* 155:2081-2095.
- Den Herder, G., and Parniske, M. 2009. The unbearable naivety of legumes in symbiosis. *Curr. Opin. Plant Biol.* 12:491-499.
- Doyle, J. J. 1994. Phylogeny of the legume family: An approach to understanding the origins of nodulation. *Annu. Rev. Ecol. Syst.* 25:325-349.
- Freeling, M., and Thomas, B. C. 2006. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res.* 16:805-814.
- Gherbi, H., Markmann, K., Svistoonoff, S., Estevan, J., Autran, D., Giczey, G., Auguy, F., Péret, B., Laplaze, L., Franche, C., Parniske, M., and Bogusz, D. 2008. SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and Franki bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 105:4928-4932.
- Gopalasubramanian, S. K., Kovacs, F., Violante-Mota, F., Twigg, P., Arredondo-Peter, R., and Sarath, G. 2008. Cloning and characterization of a caesalpinoid (*Chamaecrista fasciculata*) hemoglobin: The structural transition from a nonsymbiotic hemoglobin to a leghemoglobin. *Proteins Struct. Funct. Bioinf.* 72:252-260.
- Groth, M., Takeda, N., Perry, J., Uchida, H., Dräxl, S., Brachmann, A., Sato, S., Tabata, S., Kawaguchi, M., Wang, T. L., and Parniske, M. 2010. NENA, a *Lotus japonicus* homolog of Sec13, is required for rhizodermal infection by arbuscular mycorrhiza fungi and rhizobia but dispensable for cortical endosymbiotic development. *Plant Cell* 22:2509-2526.
- Hirsch, A. M., and Larue, T. A. 1997. Is the legume nodule a modified root or stem or an organ sui generis? *Crit. Rev. Plant Sci.* 16:361-392.
- Hocher, V., Alloisio, N., Auguy, F., Fournier, P., Doumas, P., Pujic, P., Gherbi, H., Queiroux, C., Da Silva, C., Wincker, P., Normand, P., and Bogusz, D. 2011. Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant Physiol.* Online publication. doi:10.1104/pp.111.174151.
- Indrasumunar, A., Kereszt, A., Searle, I., Miyagi, M., Li, D., Nguyen, C. D. T., Men, A., Carroll, B. J., and Gresshoff, P. M. 2010. Inactivation of duplicated nod factor receptor 5 (NFR5) genes in recessive loss-of-

- function non-nodulation mutants of allotetraploid soybean (*Glycine max* L. Merr.). *Plant Cell Physiol.* 51:201-214.
- Innan, H., and Kondrashov, F. 2010. The evolution of gene duplications: Classifying and distinguishing between models. *Nat. Rev. Genet.* 11:97-108.
- Jaillon, O., Aury, J., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyere, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pe, M. E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A., Weissenbach, J., Quetier, F., Wincker, P., and French-Italian Public. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature (London)* 449:463-467.
- Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., Tomsho, L. P., Hu, Y., Liang, H., Soltis, P. S., Soltis, D. E., Clifton, S. W., Schlarbaum, S. E., Schuster, S. C., Ma, H., Leebens-Mack, J., and dePamphilis, C. W. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473:97-100.
- Lavin, M., Herendeen, P. S., and Wojciechowski, M. F. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Syst. Biol.* 54:575-594.
- Lewis, G., Schrire, B., Mackinder, B., and Lock, M. 2005. Legumes of the World. Kew, Richmond, U.K.
- Libault, M., Farmer, A., Brechenmacher, L., Drnevich, J., Langley, R. J., Bilgin, D. D., Radwan, O., Neece, D. J., Clough, S. J., May, G. D., and Stacey, G. 2010. Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. *Plant Physiol.* 152:541-552.
- Liu Q, Wang H, Zhang Z, Wu J, Feng Y, and Zhu Z. 2009. Divergence in function and expression of the NOD26-like intrinsic proteins in plants. *BMC Genomics* 10:313.
- Madsen, L. H., Tirichine, L., Jurkiewicz, A., Sullivan, J. T., Heckmann, A. B., Bek, A. S., Ronson, C.W., James, E. K., and Stougaard, J. 2010. The molecular network governing nodule organogenesis and infection in the model legume *Lotus japonicus*. *Nat. Commun.* 1:1-12.
- Maillet, F., Poinot, V., Andre, O., Puech-Pages, V., Haouy, A., Gueunier, M., Cromer, L., Giraudet, D., Formey, D., Niebel, A., Martinez, E. A., Driguez, H., Becard, G., and Denarie, J. 2011. Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature* 469:58-63.
- Markmann, K., and Parniske, M. 2009. Evolution of root endosymbiosis with bacteria: How novel are nodules? *Trends Plant Sci.* 14:77-86.
- Markmann, K., Giczey, G., and Parniske, M. 2008. Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. *PLoS Biol.* 6:e68. Published online.
- Moreau, S., Verdenaud, M., Ott, T., Letort, S., de Billy, F., Niebel, A., Gouzy, J., de Carvalho-Niebel, F., and Gamas, P. 2011. Transcription reprogramming during root nodule development in *Medicago truncatula*. *PLoS ONE* 6:e16463. Published online.
- Oldroyd, G. E. D., Harrison, M. J., and Paszkowski, U. 2009. Reprogramming plant cells for endosymbiosis. *Science* 324:753-754.
- Op den Camp, R., Streng, A., De Mita, S., Cao, Q., Polone, E., , W., Ammiraju, J. S. S., Kudrna, D., Wing, R., Untergasser, A., Bisseling, T., and Geurts, R. 2011. LysM-type mycorrhizal receptor recruited for *Rhizobium* symbiosis in nonlegume *Parasponia*. *Science* 331:909-912.
- Pawlowski, K., and Sprent, J. I. 2008. Comparison between actinorhizal and legume symbiosis. Pages 261-288 in: *Nitrogen-Fixing Actinorhizal Symbioses*. K. Pawlowski and W. E. Newton, eds. Springer, Dordrecht, The Netherlands.
- Pfeil, B. E., Schlueter, J. A., Shoemaker, R. C., and Doyle, J. J. 2005. Placing paleopolyploidy in relation to taxon divergence: A phylogenetic analysis in legumes using 39 gene families. *Syst. Biol.* 54:441-454.
- Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Woerheide, G., and Baurain, D. 2011. Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biol.* 9:e1000602. Published online.
- Sanderson, M. J., and Hufford, L., eds. 1996. *Homoplasy: The Recurrence of Similarity in Evolution*. Academic Press, London.
- Schlueter, J. A., Dixon, P., Granger, C., Grant, D., Clark, L., Doyle, J. J., and Shoemaker, R. C. 2004. Mining EST databases to resolve evolutionary events in major crop species. *Genome* 47:868-876.
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., Cheng, J., Xu, D., Hellsten, U., May, G. D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M. K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X. C., Shinozaki, K., Nguyen, H. T., Wing, R. A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R. C., and Jackson, S. A. 2010. Genome sequence of the paleopolyploid soybean. *Nature* 463:178-183.
- Shubin, N., Tabin, C., and Carroll, S. 2009. Deep homology and the origins of evolutionary novelty. *Nature* 457:818-823.
- Soltis, D. E., Soltis, P. S., Morgan, D. R., Swensen, S. M., Mullin, B. C., Dowd, J. M., and Martin, P. G. 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 92:2647-2651.
- Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., Sankoff, D., dePamphilis, C.W., Wall, P. K., and Soltis, P. S. 2009. Polyploidy and angiosperm diversification. *Am. J. Bot.* 96:336-348.
- Sprent, J. I. 2007. Evolving ideas of legume evolution and diversity: A taxonomic perspective on the occurrence of nodulation. *New Phytol.* 174:11-25.
- Sprent, J. I. 2008. 60Ma of legume nodulation. What's new? What's changing? *J. Exp. Bot.* 59:1081-1084.
- Sprent, J. I. 2009. *Legume Nodulation: A Global Perspective*. Wiley-Blackwell, Ames, IA, U.S.A..
- Sprent, J. I., and James, E. K. 2007. Legume evolution: Where do nodules and mycorrhizas fit in? *Plant Physiol.* 144:575-581.
- Studer, R. A., and Robinson-Rechavi, M. 2009. How confident can we be that orthologs are similar, but paralogs differ? *Trends Genet.* 25:210-216.
- Swensen, S. M., and Benson, D. R. 2008. Evolution of actinorhizal host plants and Frankia endosymbionts. Pages 73-104 in: *Nitrogen-Fixing Actinorhizal Symbioses*. K. Pawlowski and W. E. Newton, eds. Springer, Dordrecht, The Netherlands.
- Tang, H. B., Wang, X. Y., Bowers, J. E., Ming, R., Alam, M., and Paterson, A. H. 2008. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* 18:1944-1954.
- Tirosh, I., and Barkai, N. 2007. Comparative analysis indicates regulatory neofunctionalization of yeast duplicates. *Genome Biol.* 8:R50.
- Van de Velde, W., Zehirov, G., Szatmari, A., Debrezeny, M., Ishihara, H., Kevei, Z., Farkas, A., Mikulass, K., Nagy, A., Tiricz, H., Satiat-Jeunemaitre, B., Alunni, B., Bourge, M., Kucho, K., Abe, M., Kereszt, A., Maroti, G., Uchiumi, T., Kondorosi, E., and Mergaert, P. 2010. Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* 327:1122-1126.
- Wake, D. B., Wake, M. H., and Specht, C. D. 2011. Homoplasy: From detecting pattern to determining process and mechanism of evolution. *Science* 331:1032-1035.
- Wang, D., Griffiths, J., Starker, C., Fedorova, E., Limpens, E., Ivanov, S., Bisseling, T., and Long, S. 2010. A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. *Science* 327:1126-1129.
- Wang, H., Moore, M. J., Soltis, P. S., Bell, C. D., Brockington, S. F., Alexandre, R., Davis, C. C., Latvis, M., Manchester, S. R., and Soltis, D. E. 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc. Natl. Acad. Sci. U.S.A.* 106:3853-3858.
- Yang, S., Tang, F., Gao, M., Krishnan, H. B., and Zhu, H. 2010. R gene-controlled host specificity in the legume-rhizobia symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 107:18735-18740.
- Yokota, K., Fukai, E., Madsen, L. H., Jurkiewicz, A., Rueda, P., Radutoiu, S., Held, M., Hossain, M. S., Szczygłowski, K., Morieri, G., Oldroyd, G. E. D., Downie, J. A., Nielsen, M. W., Rusek, A. M., Sato, S., Tabata, S., James, E. K., Oyaizu, H., Sandal, N., and Stougaard, J. 2009. Rearrangement of actin cytoskeleton mediates invasion of *Lotus japonicus* roots by *Mesorhizobium loti*. *Plant Cell* 21:267-284.
- Yokota, K., and Hayashi, M. 2011. Function and evolution of nodulation genes in legumes. *Cell Mol. Life Sci.* 68:1341-1351.
- Zhang, X., Cannon, S., and Stacey, G. 2009. Evolutionary genomics of LysM genes in land plants. *BMC Evol. Biol.* 9:183.

AUTHOR-RECOMMENDED INTERNET RESOURCE

The 1KP project website: www.onekp.com/index.html