# Deep Homology and Evolutionary Tinkering in the Origins of Nodulation

by

Alexander B. Taylor

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in the University of Michigan
2018

Doctoral Committee:

Associate Professor Yin-Long Qiu, Chair
Professor Timothy Y. James
Professor Patricia Wittkopp
Professor Donald R. Zak

Alexander B. Taylor

abtaylor@umich.edu

ORCID iD: 0000-0003-2741-1224

## DEDICATION

For my incredible wife Jen – your constant love and support made this possible

## ACKNOWLEDGEMENTS

I'm deeply grateful for all the help I've gotten in making this dissertation. First and foremost, I thank Dr. Yin-Long Qiu for all his advice and support, and for bringing me to evolutionary biology in the first place. His big-picture thinking help guide me along the way. I'm also grateful to my dissertation committee members: Dr. Tim James, Dr. Trisha Wittkopp and Dr. Donald Zak, whose advice and guidance has been invaluable, and led me to conduct my "wet" chapter on *Elaeagnus umbellata.*

I would like to thank the James lab, and particularly Dr. Tim James (again), Dr. Tommy Jenkinson, Jill Myers, Buck Castillo, Dr. Alisha Quandt and Anat Belasen, for letting me hang around – being in a small lab can be lonely, and you all gave me a second home. Likewise, the Smith Lab folks – Dr. Stephen Smith, Dr. Cody Hinchcliff, Dr. Ya Yang, Dr. Joseph Brown, and Joe Walker – made me feel welcome at lab meetings and helped guide me through some tricky analyses. I'd like to thank the undergraduate researchers in the Qiu lab, and particularly Miranda Niemiec, for fun and insightful conversations.

I'd like to thank the graduate students, faculty and staff in the EEB department, for being great friends and colleagues. I'm thankful for the donors and managers of the funding sources that made this work possible: the Rackham Graduate School, Matthaei Botanical Gardens, EEB Block Grant and Emma J. Cole Fellowship.

## TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

## ABSTRACT

Nodulation is the mutualistic symbiosis in which plants house nitrogen-fixing bacteria in specialized root nodule organs, and exchange plant photosynthate for bacteria fixed nitrogen. This symbiosis occurs only in the Nitrogen-Fixing Clade (NFC) of rosids – the Fagales, Fabales, Cucurbitales and Rosales. Within the NFC, nodulation evolved multiple times independently, and different nodulating lineages show substantial differences in nodule morphology and development, as well as bacterial symbiont and infection mechanism. Despite the apparent nonhomology of nodulation, each examined instance of its evolution involved the recruitment of the same homologous genes, meaning that nodulation is deeply homologous. The nodulation pathway evolved by the recruitment of genes from a variety of pathways, in an example of evolutionary tinkering. This dissertation examines the precise pattern of gene recruitment that created nodules in different lineages, with a focus on the subtilase gene family. A phylogenetic analysis of the subtilase gene family reveals that the subtilase ortholog recruited for nodulation in the legumes is paralogous to the one recruited in the Fagales, and that these two lineages diverged before the origin of the NFC. A synteny analysis of symbiotic subtilases in the legumes shows that a lineage of subtilases mediating arbuscular mycorrhization originated by whole genome duplication, followed by lineage-specific tandem duplication. A transcriptomic study of nodulation in *Elaeagnus umbellata*, a nodulating actinorhizal shrub in the Elaeagnaceae for which the genetic basis of nodulation is poorly

understood, assembled 13 genes homologous with those involved in nodulation in other lineages, showing deep homology. 12 of these 13 genes were orthologous, showing congruence with the species tree in phylogenetic analysis, but our assembled *E. umbellata* subtilase was paralogous with those mediating nodulation in the actinorhizal Fagales and orthologous with those mediating nodulation in rhizobial legumes. In showing differential recruitment of subtilases in the independent origins of nodulation, this dissertation provides further evidence of the nonhomology of nodulation, and presents nodulation as a fertile system to examine the ortholog conjecture, deep homology, and evolutionary tinkering in the origin of a complex and important symbiotic organ.

# CHAPTER 1

# Introduction

## 1.1 Complexities of Homology Inference

Homology is one of the most powerful and animating ideas in describing the diversity of life on earth (Simpson, 1961; Van Valen, 1982; Patterson, 1982; Roth, 1984; Patterson, 1988; Wagner, 1989). The term was first defined by Richard Owen in 1843 as "the same organ in different animals under every variety of form and function" (Owen, 1855). After the introduction of formal evolutionary theory, it came to mean a similarity of corresponding structures in different species, based on continuous inheritance from a common ancestor (Patterson, 1988). A shared developmental basis and anatomical position has long been recognized to be an important part of the similarity criterion of homology (Roth, 1984; Wagner, 1989). Even in the 19th century, Karl Gegenbaur specified that homologous structures should be derived from the same primordium or *anlage* (quoted by Spemann, 1915 translated by Wagner, 1989). If the same developmental program and constraints yield corresponding structures in different organisms, the thinking goes, the similarity in those structures is more likely to reflect a shared inheritance and common descent (Wagner, 1989). However, as the genetic underpinnings of various traits and their development have been revealed in recent

decades, this developmental reasoning has been complicated (Abouheif *et al.,* 1997; Shubin *et al.,* 2009; Scotland, 2010; Mozcek *et al.,* 2015).

Continuously inherited orthologous genes can be independently recruited for convergent functions in different lineages, a phenomenon that has been called "deep homology" (Shubin *et al.,* 1997). Structures traditionally viewed as analogous, such as the eyes of insects and vertebrates, are directed in part by orthologous genes (such as *PAX6*) that were independently recruited in these lineages (Shubin *et al.,* 2009). Furthermore, genes may be recruited from multiple different pathways and neofunctionalized to form a new one, a process called "tinkering" or "bricolage" (Jacob, 1977; Wilkins & DuBoule, 1998; Wilkins, 2007). Therefore, there needn't be a one-to-one correspondence between the evolutionary histories of different elements of a trait or structure, nor between the different levels (genetic, cellular, developmental, and anatomical) at which that structure can be assessed (Roth, 1984; Shubin *et al.,* 1997). Since organs and their development emerge from the interactions of multiple genes, each of which may have a different evolutionary history, complex structures should be broken up into their component parts when assessing homology (Roth, 1988). This view has led to a more nuanced assessment of homology, and greater insight into the evolutionary processes that create complex traits. In plant biology, nodulation affords one of the most fertile case studies in the complexities of inferring homology (Markmann & Parniske, 2009; Doyle, 2011).

## 1.2 Nodules are not Simply Homologous

Nodulation is a mutualistic symbiosis in which nitrogen-fixing bacteria, housed in specialized nodule organs on the plant root, exchange bacterially fixed nitrogen for plant photosynthate (Hellriegel & Wilfarth, 1888; Sprent, 2001). This association has been studied in detail due to its important role in terrestrial nutrient cycling in both wild and agricultural ecosystems, and in providing dietary protein for human nutrition (Galloway *et al.,* 1995; Smil 1999; Sprent, 2001; Oldroyd 2013). In angiosperms, nodulation occurs only in the Nitrogen Fixing Clade (NFC) of rosids (the Fagales, Fabales, Cucurbitales and Rosales) (Soltis *et al.,* 1995). Elements of the nodulation program have a shared genetic basis and developmental pathway across examined nodulating lineages (Swensen, 1996; Gualtieri & Bisseling, 2000; Markmann *et al.,* 2008; Gherbi *et al.,* 2008; Svistoonoff *et al.,* 2014).

Despite the unifying definition, phylogenetic clustering, and appreciable genetic and developmental similarities among different instances of nodulation, nodulation fails both the congruency and similarity tests of homology (Swensen *et al.,* 1996; Doyle, 2011). Multiple lines of phylogenetic, genetic, morphological and developmental evidence suggest that nodules are not strictly homologous and instead were derived from multiple independent origins. First, within the NFC, the phylogenetic distribution of nodulation is quite incongruent. A strict parsimony analysis of ancestral state infers five independent origins within the legumes and nine origins in the Rosales, Fagales and Cucurbitales. The predominant view is that nodulation evolved multiple times independently within the NFC with relatively few losses (Soltis *et al.,* 1995; Swensen, 1996; Swensen & Mullin, 1997; Gualtieri & Bisseling, 2000; Markmann & Parniske,

2009; Doyle 2011; Svistoonoff *et al.,* 2014; Werner *et al.,* 2014; Li *et al.,* 2015; Doyle, 2016), though there is some evidence for the alternative view that nodulation evolved once with massive subsequent losses (Soltis *et al.,* 1995; Van Velzen *et al.,* 2018). Second, there are differences in nodule development among nodulating lineages (**Fig. 1.1, Fig. 1.2**). In non-legumes, the nodule is indeterminate (retaining a meristem) and arises from pericycle cells (Racette & Torrey, 1989; Pawlowski & Sprent, 2008), whereas in legumes, the nodule can be determinate or indeterminate in different lineages, and is derived from cortical cells (Sprent, 2001; Sprent & James, 2007). Legumes grow peripheral vasculature around the nodule like a stem, while non-legumes grow a central vasculature like a root (Pawlowski & Sprent, 2008). Third, some nodulating lineages are infected through root hair curling, while others are infected through the middle lamellae between epidermal cells or even wounds or cracks in the epidermis; all three infection mechanisms have a polyphyletic distribution (Sprent, 2001; Pawlowski & Sprent, 2008). Fourth, different paralogs have been recruited in the independent origins of nodulation, which would not be expected in nodules were homologous (Taylor & Qiu, 2017; Sturms *et al.,* 2010).

Additionally, different nitrogen-fixing bacteria infect different nodulating plant species (Hirsch & LaRue*,* 1997; Sprent, 2001; Pawlowski & Sprent, 2008). The actinobacterial genus *Frankia* infects the actinorhizal plants, which constitute about 220 species spanning eight families and 25 genera in the Rosales, Fagales, and Cucurbitales (Fig. 1A, Wall, 2000). Different species of bacteria collectively called "rhizobia," which comprise a polyphyletic group of 15 genera in both the alpha- and beta-proteobacteria, nodulate different legumes, as well as the genus *Parasponia* in the Cannabaceae

(Rosales) (Swensen, 1996; Sprent, 2001). All rhizobia have both nitrogen-fixing genes (such as *nifH*) and *nod* genes that encode enzymes responsible for making lipo-chito-oligosaccharide (LCO) signaling molecules (Velázquez *et al.,* 2010; Gyaneshwar *et al.,* 2011).

### 1.3 The Evolution of Nodulation

While nodules are likely not homologous, the absence of nodulation outside of the NFC suggests that there is some synapomorphic genetic endowment or "predisposition" for nodulation unique to this clade (Soltis *et al.,* 1995; Swensen, 1996; Doyle, 1998; Markmann & Parniske, 2009). The shared genetic basis across independent instances of nodulation strengthens this case; more than 290 orthologous genes have been found to be induced during nodulation in both *Medicago* (Fabales) and *Parasponia* (Rosales) (van Velzen *et al.,* 2018).

Our understanding of the shared genetic basis for nodulation has been growing since the late 1980's, when legume mutants incapable of forming nodules were discovered to be deficient in forming Arbuscular Mycorrhizal (AM) associations with glomeromycete fungi as well (Duc *et al.*, 1989). Nodulation was hypothesized to be derived from AM (LaRue & Weeden, 1994), and subsequent investigation revealed that these dual deficiencies were due to mutations in orthologous genes involved in a signal transduction pathway mediating both AM and nodulation (e.g., van Rhijn et al., 1997; Catoira *et al.,* 2000; Kistner *et al.,* 2002; Kistner *et al.,* 2005). This pathway has been called the Common Symbiotic (*sym*) Pathway or the Common Symbiotic Signaling Pathway (CSSP).

The CSSP genes encode a pathway for the perception of microsymbiont signals and subsequent cellular signal transduction events including nuclear calcium spiking, leading to the reception of the microsymbiont in the plant root (for reviews, see Parniske, 2008; Yokota & Hayashi, 2011; Oldroyd, 2013; Genre & Russo, 2016). Ten CSSP genes have been shown to be required for both nodulation and AM (Table 2.1, Genre & Russo, 2016), with dozens shown to be induced during both symbioses (van Velzen *et al.,* 2018). Many of these genes, as well as the nuclear calcium spiking phenotype, have been shown to be induced during nodulation in both rhizobial and actinorhizal lineages (Gherbi *et al.,* 2008; Hocher *et al.,* 2011; Svistoonoff *et al.,* 2014; Granqvist *et al.,* 2015). Nodules thus exemplify "deeply homologous" structures (Doyle, 2011), in that phylogenetically distinct instances of nodulation originated by the repeated independent recruitment of homologous genes from an AM signaling pathway, which is symplesiomorphic relative to nodulating lineages.

It is not surprising that the repeated independent evolution of a complex organ such as the nodule would involve the recruitment of a "genetic toolkit" comprising many genes for nodulation, as is seen with the repeated independent evolution of C4 photosynthesis in grasses (Christin *et al.,* 2013). However, while the CSSP is the kind of "genetic toolkit" that can help account for the repeated evolution of an organ as complex as the nodule, the CSSP itself cannot be the genetic endowment or predisposition for nodulation. The CSSP is not restricted to the NFC, and its parallel recruitment does not explain why nodulation only occurs in the NFC; CSSP genes mediate the AM association, which is synapomorphic to all land plants (Wang & Qiu, 2006; Wang *et al.,* 2010). If the CSSP is present and conserved across land plants, what is the special

"predisposition" for nodulation in the NFC? One possibility lies in NFC-specific increases in the "bandwidth" of the CSSP, or the ability to discriminate between signals from nodule bacteria and AM fungi. Some CSSP genes show functional equivalence in NFC and non-NFC lineages; for example, the *Oryza sativa* ortholog of the CSSP gene *CYCLOPS* can restore nodulation in *Lotus japonicus* (Yano *et al.,* 2008). However, other CSSP genes like *SYMRK* and *CCAMK* from non-NFC lineages can restore AM but not nodulation phenotypes in legume *symrk* and *ccamk* mutants (Markmann *et al.,* 2008; Wang *et al.,* 2010).

Further complicating the picture, the genetic machinery of nodulation was not only recruited from the AM symbiosis, but also assembled from genes recruited from multiple pathways (Szczyglowski & Amyot, 2003; Yokota & Hayashi, 2011; Soyano & Hayashi, 2014; Doyle 2016). Duplication and neofunctionalization or subfunctionalization in these non-CSSP genes may also play a role in an increased bandwidth, as different paralogs can mediate AM and nodulation (Vanneste *et al.,* 2014; Taylor & Qiu, 2017). Genes from innate immune pathways (Nakagawa *et al.,* 2011), lateral root development (Soyano *et al.,* 2013; Soyano & Hayashi, 2014), and even pollen tube development (Tansengco *et al.,* 2004; Nouri & Reinhardt, 2015) were also recruited for nodulation.

The evolutionary origin of nodulation may thus represent an example of "tinkering" or "bricolage" (Jacob, 1977; Wilkins & Duboule, 1998; Soyano & Hayashi, 2014; Doyle, 2016). Roth (1988) called this kind of genetic rewiring "genetic piracy" while arguing that understanding the homology of the resulting traits requires deconstructing their individual aspects and assessing the evolutionary history of each.

While CSSP genes show an evolutionary history of vertical inheritance of single-copy orthologs, many other nodulation genes have more complex evolutionary histories (Wang *et al.,* 2010; Taylor & Qiu, 2017). Different nodulating lineages have recruited different paralogous genes during the independent origins of nodulation (Op den Camp *et al.,* 2011; Vázquez-Limón *et al.,* 2012; Taylor & Qiu, 2017), which could impact the symbiont specificity, development, or morphology of the resulting nodules

Nodulation represents an excellent system to study questions of deep homology and evolutionary tinkering in the origins of a complex trait. The phylogenetic distribution of nodulation is well understood (Soltis *et al.,* 1995; Li *et al.,* 2015; LPWG, 2017). The morphology and development of nodules has been described extensively in many lineages (see Sprent, 2001; Pawloski & Sprent, 2008; Pawlowski & Demchenko, 2012). Nodulation has been characterized genetically in model papilionoid legumes, and work is underway to extend that understanding to actinorhizal lineages and the rhizobial genus *Parasponia* (e.g., Gherbi *et al.,* 2008; Op den Camp *et al.,* 2011; Hocher *et al.,* 2011; Svistoonoff *et al.,* 2013; Svistoonoff *et al.,* 2014; Van Velzen *et al.,* 2018). This body of work allows for a phylogenetic approach to the question of deep homology in the origins of evolutionarily independent instances of nodulation, by assessing evolutionary history of genes involved in nodulation (Doyle 1994; De Mita *et al.,* 2014; Taylor & Qiu, 2017). This dissertation examines the evolution of nodulation on a genetic level, by investigating the evolutionary history of nodulation genes, and adds the root transcriptome of *Elaeagnus umbellata*, a species in a nodulating lineage that has not been extensively examined at a genetic level.

## 1.4 Summary of Dissertation Chapters

In **Chapter 2**, I review the literature concerning homology between evolutionarily distinct nodulation symbioses at developmental, morphological, and genetic levels, in a phylogenetic framework. This work draws on over a century of characterization of the morphology and development of nodules in different lineages, as well as more recent work on its genetic basis. Nodulation has a complex evolutionary history of deep homology and evolutionary tinkering, yielding different symbiotic organs that remain variations on a single theme. While a common genetic toolkit underlies the repeated evolution of nodules, the precise orthologous gene lineages recruited in some nodulating lineages are not present in others; different plant lineages differentially recruited divergent paralogs of genes such as subtilases and receptor kinases for nodulation. Finally, some genes repeatedly recruited for nodulation, such as the transcription factor *NIN* and nodule hemoglobins, are not involved in the AM, and were recruited from separate pathways, exemplifying evolutionary tinkering. These differences in genetic material available may help explain differences in host specificity and nodule morphology, as well as the phylogenetic distribution of a "predisposition" to nodulate, and which kinds of nodules different lineages are predisposed to evolve. This chapter provides the intellectual foundation for the rest of the dissertation.

In **Chapter 3**, I analyze the evolutionary history of the subtilase gene family, a group of proteases involved in protein turnover at the symbiotic interface in nodulation and AM symbioses. Phylogenetic analysis of the subtilase gene family in plants shows a pattern of repeated duplication that accelerated during the origin of angiosperms. Phylogenetic and syntenic analysis shows that subtilases required for AM in *Lotus*

*japonicus* arose in a legume-specific whole genome duplication. I identify the orthologous gene lineages of several characterized subtilases, with various symbiotic and non-symbiotic functions. I show that subtilases required for nodulation and AM in *Lotus japonicus* are orthologous with those involved in pathogen defense in the asterid *Solanum lycopersicum,* and these genes are paralogous with a lineage of subtilases required for nodulation in *Casuarina glauca* (Fagales). Different orthologous gene lineages that diverged during the origin of angiosperms have been differentially recruited for nodulation in actinorhizal and legume nodulation. In light of the transcriptional conservation of these subtilases in nodulation in these two lineages (Svistoonoff *et al.,* 2003; Svistoonoff *et al.,* 2004), this pattern of differential recruitment of paralogous subtilases for convergent function counters the ortholog conjecture, that orthologous genes are more likely to be recruited for similar functions (Kondrashov *et al.,* 2002; Gabaldon & Koonin, 2013).

**Chapter 4** concerns the root transcriptome of the actinorhizal shrub *Elaeagnus umbellata* following exposure to its nodule symbiont *Frankia*. While much progress has been made in elucidating the genetic basis of nodulation, most of this research has been restricted to a few model organisms, such as the papilionoid legumes *Medicago truncatula* and *Lotus japonicus*, and the actinorhizal species *Casuarina glauca* and *Alnus glutinosa* (Fagales), both characterized by infection through root hair curling and the formation of transcellular infection threads. Several lineages representing independent origins of nodulation, and particularly actinorhizal and intercellularly infected lineages, remain poorly studied at a genetic level. *E. umbellata* is intercellularly infected with no transcellular infection threads, and is in a lineage (Elaeagnaceae; Rosales) for which the

genetic basis of nodulation formation has not been examined. Thus, *E. umbellata* represents an excellent system for discovering the degree of parallelism and convergence in the evolutionary origins of nodulation.

Transcriptome assembly of *E. umbellata* discovered several genes that are homologous to those involved in nodulation in other lineages, providing another example of deep homology in nodulation. Phylogenetic analyses of these genes show their precise orthologous gene lineage; the majority are orthologous to those mediating nodulation in other lineages, as has been previously reported for most CSSP genes. However, we found two examples of nodulation genes in paralogous lineages that were deeply divergent with those mediating nodulation in other plants. A subtilase upregulated in actinorhizal *E. umbellata* during nodulation is in an orthologous lineage to those required for nodulation in rhizobial legumes, and is distantly related to the subtilases required for nodulation in the actinorhizal Fagales. Additionally, while our *E. umbellata NIN* was found to be orthologous to *NIN* in nodulating species in the Fagales, Fabales and Rosales, we found that a *NIN* homolog previously reported to be upregulated during nodulation in *Datisca glomerata* (Cucurbitales; Demina *et al.,* 2013) was orthologous with the *NLP1* lineage, and paralogous to the *NIN* orthogroup.

**References:**


**Abouheif, E., Akam, M., Dickinson, W.J., Holland, P.W., Meyer, A., Patel, N.H., Raff, R.A., Roth, V.L. and Wray, G.A.,** 1997. Homology and developmental genes. *Trends in genetics*, *13*(11), pp.432-433.

**Abouheif, E.**, 2008. Parallelism as the pattern and process of mesoevolution. *Evolution and Development*, *10*(1), pp.3-5.

**Benson, D.R., Brooks, J.M., Huang, Y., Bickhart, D.M. and Mastronunzio, J.E.,** 2011. The biology of Frankia sp. strains in the post-genome era. *Molecular plant-microbe interactions*, *24*(11), pp.1310-1316.

**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*(7), p.e11630.

**Catoira, R., Galera, C., de Billy, F., Penmetsa, R.V., Journet, E.P., Maillet, F., Rosenberg, C., Cook, D., Gough, C. and Dénarié, J.** 2000. Four genes of Medicago truncatula controlling components of a Nod factor transduction pathway. *The Plant Cell*, *12*(9), pp.1647-1665.

**Chabaud, M., Gherbi, H., Pirolles, E., Vaissayre, V., Fournier, J., Moukouanga, D., Franche, C., Bogusz, D., Tisa, L.S., Barker, D.G. and Svistoonoff, S.,** 2016. Chitinase‐resistant hydrophilic symbiotic factors secreted by Frankia activate both Ca2+ spiking and NIN gene expression in the actinorhizal plant Casuarina glauca. *New Phytologist*, *209*(1), pp.86-93.

**Christin, P.A., Boxall, S.F., Gregory, R., Edwards, E.J., Hartwell, J. and Osborne, C.P.,** 2013. Parallel recruitment of multiple genes into C4 photosynthesis. *Genome biology and evolution*, *5*(11), pp.2174-2187.

**De Mita S, Streng A, Bisseling T, Geurts R.** 2014. Evolution of a symbiotic receptor through gene duplications in the legume–rhizobium mutualism. *New Phytologist* **201(3):** 961-972.

**Demina IV, Persson T, Santos P, Plaszczyca M, Pawlowski K. 2013.** Comparison of the nodule vs. root transcriptome of the actinorhizal plant *Datisca glomerata*: actinorhizal nodules contain a specific class of defensins. *PloS one* **8(8):** e72442.

**Doyle JJ.** 1994. Phylogeny of the legume family: an approach to understanding the origins of nodulation. *Annual Review of Ecology and Systematics*, 325-349.

**Doyle, J.J.,** 1998. Phylogenetic perspectives on nodulation: evolving views of plants and symbiotic bacteria. *Trends in Plant Science*, *3*(12), pp.473-478.

**Doyle JJ.** 2011. Phylogenetic perspectives on the origins of nodulation. *Molecular Plant-Microbe Interactions* **24:** 1289–129.

**Doyle, J.J.,** 2016. Chasing unicorns: Nodulation origins and the paradox of novelty. *American journal of botany*, *103*(11), pp.1865-1868.

**Duc G, Trouvelot A, Gianinazzi-Pearson V, Gianinazzi S.** 1989. First report of non-mycorrhizal plant mutants (Myc-) obtained in pea (*Pisum sativum* L.) and fababean (*Vicia faba* L.). *Plant Science* **60:** 215–222.

Gabaldón, T. and Koonin, E.V., 2013. Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics*, *14*(5), p.360.

**Galloway, J.N., Schlesinger, W.H., Levy, H., Michaels, A. and Schnoor, J.L.,** 1995. Nitrogen fixation: Anthropogenic enhancement‐environmental response. *Global biogeochemical cycles*, *9*(2), pp.235-252.

**Genre, A. and Russo, G.**, 2016. Does a common pathway transduce symbiotic signals in plant–microbe interactions? *Frontiers in plant science*, *7*.

**Gherbi H, Markmann K, Svistoonoff S, Estevan J, Autran D, Giczey G, Auguy F, *et al.*** 2008. SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and Frankia bacteria. *Proceedings of the National Academy of Science USA* **105:** 4928–4932.

**Granqvist, E., Sun, J., Op den Camp, R., Pujic, P., Hill, L., Normand, P., Morris, R.J., Downie, J.A., Geurts, R. and Oldroyd, G.E.,** 2015. Bacterial‐induced calcium oscillations are common to nitrogen‐fixing associations of nodulating legumes and non‐legumes. New Phytologist, 207(3), pp.551-558.

**Gualtieri, G. and Bisseling, T.** 2000. The evolution of nodulation. Plant Mol. Biol. 42: 181–194.

**Gyaneshwar, P., Hirsch, A.M., Moulin, L., Chen, W.M., Elliott, G.N., Bontemps, C., Estrada-de los Santos, P., Gross, E., dos Reis Jr, F.B., Sprent, J.I. and Young, J.P.W.,** 2011. Legume-nodulating betaproteobacteria: diversity, host range, and future prospects. *Molecular plant-microbe interactions*, *24*(11), pp.1276-1288.

**Hellriegel H, Wilfarth H,** 1888. Untersuchungen uber die Stickstoffnahrung der Grammeen und Legummosen, Beilageheft zu der Zeitschrift des Vereins Rubenzucker-Jndustrie Deutschen Reiches 1-234.

**Hirsch, A.M., Larue, T.A.,** 1997. Is the legume nodule a modified root or stem or an organ sui generis?. *Critical Reviews in Plant Sciences*, *16*(4), pp.361-392.

**Hocher V, Alloisio N, Auguy F, Founier P, Doumas P, Pujic P, Gherbi H.** 2011. Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant Physiology* Online publication.

**Jacob, F.,** 1977. Evolution and tinkering. *Science (New York, NY)*, *196*(4295), pp.1161-1166.

**Kistner C, Parniske, M. 2002.** Evolution of signal transduction in intracellular symbiosis. *Trends in Plant Science* **7:** 511–518.

**Kistner C, Winzer T, Pitzschke A, Mulder L, Sato S, Kaneko T, Tabata S. 2005.** Seven Lotus japonicus genes required for transcriptional reprogramming of the root during fungal and bacterial symbiosis. *The Plant Cell **17*(8):** 2217-2229.

**Kondrashov, F.A., Rogozin, I.B., Wolf, Y.I. and Koonin, E.V.,** 2002. Selection in the evolution of gene duplications. *Genome biology*, *3*(2), pp.research0008-1.

**Laplaze, L., Duhoux, E., Franche, C., Frutz, T., Svistoonoff, S., Bisseling, T., Bogusz, D. and Pawlowski, K.,** 2000. Casuarina glauca prenodule cells display the same differentiation as the corresponding nodule cells. *Molecular plant-microbe interactions*, *13*(1), pp.107-112.

**LaRue, T.A. and Weeden, N.F.,** 1994, August. The symbiosis genes of the host. In *Proceedings of the 1st European Nitrogen Fixation Conference. Officina Press, Szeged, Hungary* (pp. 147-151).

**Li , H. L. , W. Wang , P. E. Mortimer , R. Q. Li , D. Z. Li , K. D. Hyde , J. C. Xu , et al.** 2015. Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. *Scientific Reports* 5: 14023.

**The Legume Phylogeny Working Group (LPWG),** 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon*, *66*(1), pp.44-77.

**Maillet F, Poinsot V, André O, Puech-Pagès V, Haouy A, Gueunier M, Cromer L *et al*. 2011.** Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature*, **469(7328):** 58-63.

**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 biology*, *6*(3), p.e68.

**Markmann, K. and Parniske, M.,** 2009. Evolution of root endosymbiosis with bacteria: How novel are nodules?. *Trends in plant science*, *14*(2), pp.77-86.

**Moczek, A.P., Sears, K.E., Stollewerk, A., Wittkopp, P.J., Diggle, P., Dworkin, I., Ledon‐Rettig, C., Matus, D.Q., Roth, S., Abouheif, E. and Brown, F.D.,** 2015. The significance and scope of evolutionary developmental biology: a vision for the 21st century. *Evolution & Development*, *17*(3), pp.198-219.

**Nakagawa, T., Kaku, H., Shimoda, Y., Sugiyama, A., Shimamura, M., Takanashi, K., Yazaki, K., Aoki, T., Shibuya, N. and Kouchi, H.,** 2011. From defense to symbiosis: limited alterations in the kinase domain of LysM receptor‐like kinases are crucial for evolution of legume–Rhizobium symbiosis. *The Plant Journal*, *65*(2), pp.169-180.

**Normand, P., Lapierre, P., Tisa, L.S., Gogarten, J.P., Alloisio, N., Bagnarol, E., Bassi, C.A., Berry, A.M., Bickhart, D.M., Choisne, N. and Couloux, A.,** 2007. Genome characteristics of facultatively symbiotic Frankia sp. strains reflect host range and host plant biogeography. *Genome research*, *17*(1), pp.7-15.

**Nouri, E., and D. Reinhardt.** 2015. Flowers and mycorrhizal roots—Closer than we think? *Trends in Plant Science* 20 : 344 – 350.

**Oldroyd GE. 2013.** Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews Microbiology* **11(4):** 252-263.

**Op den Camp R, Streng A, De Mita S, Cao Q, Polone E, Lie W, Ammiraju JSS *et al.* 2011.** LysM-type mycorrhizal receptor recruited for Rhizobium symbiosis in nonlegume Parasponia. *Science* **331:** 909–912.

**Owen, R.,** 1855. *Lectures on the comparative anatomy and physiology of the invertebrate animals*. Longman.

**Parniske M. 2008.** Arbuscular Mycorrhiza: the Mother of Plant Root Endosymbioses. *Nature Reviews of Microbiology* **6:** 763–775.

**Patterson, C.**, 1982. Morphological characters and homology. *Problems of phylogenetic reconstruction*, pp.21-74.

**Patterson, C.,** 1988. Homology in classical and molecular biology. *Molecular biology and evolution*, *5*(6), pp.603-625.

**Pawlowski, K. and Demchenko, K.N.**, 2012. The diversity of actinorhizal symbiosis. *Protoplasma*, *249*(4), pp.967-979.

**Pawlowski K, Sprent JI. 2008.** Comparison between actinorhizal and legume symbiosis. In *Nitrogen-fixing actinorhizal symbioses* (pp. 261-288). Springer Netherlands.

**Racette, S. and Torrey, J.G.,** 1989. Root nodule initiation in Gymnostoma (Casuarinaceae) and Shepherdia (Elaeagnaceae) induced by Frankia strain HFPGpI1. *Canadian Journal of Botany*, *67*(10), pp.2873-2879.

**Roth, V.L.,** 1984. On homology. *Biological Journal of the Linnean Society*, *22*(1), pp.13-29.

**Roth, V.L.,** 1988. The biological basis of homology. *Ontogeny and systematics*, pp.1-26.

**Scotland, R.W.,** 2010. Deep homology: a view from systematics. *Bioessays*, *32*(5), pp.438-449.

**Shubin, N., Tabin, C. and Carroll, S.**, 1997. Fossils, genes and the evolution of animal limbs. *Nature*, *388*(6643), p.639.

**Shubin, N., Tabin, C. and Carroll, S.,** 2009. Deep homology and the origins of evolutionary novelty. *Nature*, *457*(7231), p.818.

**Simpson, G.G.,** 1961. Principles of animal taxonomy.

**Smil, V.,** 1999. Nitrogen in crop production: An account of global flows. *Global biogeochemical cycles*, *13*(2), pp.647-662.

**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. *Proceedings of the National Academy of Sciences*, *92*(7), pp.2647-2651.

**Soyano, T., Kouchi, H., Hirota, A. and Hayashi, M.,** 2013. Nodule inception directly targets NF-Y subunit genes to regulate essential processes of root nodule development in Lotus japonicus. *PLoS Genetics*, *9*(3), p.e1003352.

**Soyano , T. , and M. Hayashi.** 2014 . Transcriptional networks leading to symbiotic nodule organogenesis. Current Opinion in Plant Biology 20 : 146 – 154 .

**Spemann, H.,** 1915. Zur Geschichte und Kritik des Begriffs der Homologie. *Die Kulturen der Gegenwart*, *1*, pp.63-86.

**Sprent JI.** 2001. Nodulation in legumes. London: Royal Botanic Gardens Kew.

**Sprent, J.I. and James, E.K.,** 2007. Legume evolution: where do nodules and mycorrhizas fit in?. *Plant Physiology*, *144*(2), pp.575-581.

**Sturms, R., Kakar, S., Trent III, J. and Hargrove, M.S.,** 2010. *Trema* and *Parasponia* hemoglobins reveal convergent evolution of oxygen transport in plants. *Biochemistry, 49*(19), pp. 4085-4093.

**Svistoonoff, S., Laplaze, L., Auguy, F., Runions, J., Duponnois, R., Haseloff, J., Franche, C. and Bogusz, D.,** 2003. cg12 expression is specifically linked to infection of root hairs and cortical cells during Casuarina glauca and Allocasuarina verticillata actinorhizal nodule development. *Molecular plant-microbe interactions*, *16*(7), pp.600-607.

**Svistoonoff, S., Laplaze, L., Liang, J., Ribeiro, A., Gouveia, M.C., Auguy, F., Fevereiro, P., Franche, C. and Bogusz, D.,** 2004. Infection-related activation of the cg12 promoter is conserved between actinorhizal and legume-rhizobia root nodule symbiosis. *Plant physiology*, *136*(2), pp.3191-3197.

**Svistoonoff, S., Benabdoun, F.M., Nambiar-Veetil, M., Imanishi, L., Vaissayre, V., Cesari, S., Diagne, N., Hocher, V., De Billy, F., Bonneau, J. and Wall, L.,** 2013. The independent acquisition of plant root nitrogen-fixing symbiosis in Fabids recruited the same genetic pathway for nodule organogenesis. *PLoS One*, *8*(5), p.e64515.

**Svistoonoff, S., Hocher, V. and Gherbi, H.,** 2014. Actinorhizal root nodule symbioses: what is signalling telling on the origins of nodulation?. *Current opinion in plant biology*, *20*, pp.11-18.

**Swensen, S.M.,** 1996. The evolution of actinorhizal symbioses: evidence for multiple origins of the symbiotic association. *American Journal of Botany*, pp.1503-1512.

**Swensen, S.M. and Mullin, B.C.,** 1997. Phylogenetic relationships among actinorhizal plants. The impact of molecular systematics and implications for the evolution of actinorhizal symbioses. *Physiologia plantarum*, *99*(4), pp.565-573.

**Szczyglowski, K. and Amyot, L.** 2003. Symbiosis, inventiveness by recruitment?. *Plant Physiology*, *131*(3), pp.935-940.

**Tansengco, M.L., Imaizumi-Anraku, H., Yoshikawa, M., Takagi, S., Kawaguchi, M., Hayashi, M. and Murooka, Y.,** 2004. Pollen development and tube growth are affected in the symbiotic mutant of Lotus japonicus, crinkle. *Plant and cell physiology*, *45*(5), pp.511-520.

**Taylor, A. and Qiu, Y.L.,** 2017. Evolutionary history of subtilases in land plants and their involvement in symbiotic interactions. *Molecular Plant-Microbe Interactions*, *30*(6), pp.489-501.

**Van Rhijn, P., Fang, Y., Galili, S., Shaul, O., Atzmon, N., Wininger, S., Eshed, Y., Lum, M., Li, Y., To, V. and Fujishige, N.,** 1997. Expression of early nodulin genes in alfalfa mycorrhizae indicates that signal transduction pathways used in forming arbuscular mycorrhizae and Rhizobium-induced nodules may be conserved. *Proceedings of the National Academy of Sciences*, *94*(10), pp.5467-5472.

**Van Valen, L.M.,** 1982. Homology and causes. *Journal of Morphology*, *173*(3), pp.305-312.

**Van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., Santuari, L., Cao, Q., Sharma, T., Shen, D. and Roswanjaya, Y.,** 2018. Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proceedings of the National Academy of Sciences*, p.201721395.

**Vanneste, K., Maere, S. and Van de Peer, Y.,** 2014. Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Phil. Trans. R. Soc. B*, *369*(1648), p.20130353.

**Vázquez-Limón, C., Hoogewijs, D., Vinogradov, S.N. and Arredondo-Peter, R.,** 2012. The evolution of land plant hemoglobins. *Plant science*, *191*, pp.71-81.

**Velázquez, E., García-Fraile, P., Ramírez-Bahena, M.H., Peix, A. and Rivas, R.,** 2010. Proteobacteria forming nitrogen fixing symbiosis with higher plants. *Nova Science Publishers, Inc., NY, USA*, pp.37-56.

**Wagner, G.P.,** 1989. The biological homology concept. *Annual Review of Ecology and Systematics*, *20*(1), pp.51-69.

**Wall, L.G.,** 2000. The actinorhizal symbiosis. *Journal of plant growth regulation*, *19*(2), pp.167-182.

**Wang, B. and Qiu, Y.L.,** 2006. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza*, *16*(5), pp.299-363.

**Wang B, Yeun LH, Xue JY, Yang L, Ane JM, Qiu YL. 2010.** Presence of three mycorrhizal genes in the common ancestor of land plants suggests a key role of mycorrhizas in the colonization of land by plants. *New Phytologist* **186:** 514–525.

**Werner GD, Cornwell WK, Sprent JI, Kattge J, Kiers ET. 2014.** A single evolutionary innovation drives the deep evolution of symbiotic N2-fixation in angiosperms. *Nature Communications* *5*:4087

**Wilkins, A.S. and DuBoule, D.,** 1998. The evolution of 'bricolage'. *Trends Genet*, *14*, pp.54-59.

**Wilkins, A.S.,** 2007. Between "design" and "bricolage": genetic networks, levels of selection, and adaptive evolution. *PNAS*, *104*(suppl 1), pp.8590-8596.

**Yano, K., Yoshida, S., Müller, J., Singh, S., Banba, M., Vickers, K., Markmann, K., White, C., Schuller, B., Sato, S. and Asamizu, E.,** 2008. CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proceedings of the National Academy of Sciences*, *105*(51), pp.20540-20545.

**Yokota, K. and Hayashi, M.,** 2011. Function and evolution of nodulation genes in legumes. *Cellular and molecular life sciences*, *68*(8), pp.1341-1351

# CHAPTER 2

## The Evolution of Nodulation

### 2.1 Nodulation in Legumes

*Infection Mechanism*

Our understanding of the infection and organogenesis of nodules is derived primarily from a few model lineages, foremost among them the model papilionoid legumes *Lotus japonicus* (tribe Loteae) and *Medicago truncatula* (tribe Trifoleae). In *M. truncatula*, perception of compatible Lipo-Chito-Oligosaccharide (LCO) signals by receptor kinases induces a set of coordinated developmental responses (Timmers *et al.,* 1999; Guinel & Gell, 2002; Oldroyd & Downie, 2008). The root hair curls, forming the characteristic "Shepherd's crook" which traps rhizobia in a pocket on the cell surface (Truchet *et al.,* 1991). Calcium influx induces the movement of the nucleus to the tip of the root hair, at which point it is surrounded by a dense microtubule array (Felle *et al.,* 1999; Timmers *et al.,* 1999). Nuclear calcium spiking in the root hair and surrounding epidermal tissue, mediated by the Common Symbiotic Signaling Pathway (CSSP), induces the expression of genes associated with nodule organogenesis and infection, which may be separate but interconnected pathways (Ehrhardt *et al.,* 1996; Madsen *et al.,* 2010; Oldroyd 2013; Soyano & Hayashi, 2014). A "pre-infection thread" (PIT) of cytoplasmic bridges across root cells and an "infection thread" (IT) sheathed in cell wall material begins to form from the center of the curled root hair by invagination of the

plasma membrane (Dart, 1977; Timmers *et al.,* 1999). A dense array of microtubular cytoskeleton surrounds the IT, which connects to the root hair nucleus and then grows in a polar fashion through the PIT (Timmers *et al.,* 1999). Rhizobia grow through the IT, which carries them to the nodule primordium, formed from cortical cell divisions (Timmers *et al.,* 1999).

This canonical root hair infection mechanism appears to be ancestral and widespread across legumes, being found in the basal genus *Chamaecrista* (Cassieae), and the majority of mimosoid and papilionoid legumes (Fig. 1B; Sprent, 2001). Nuclear calcium spiking in response to rhizobial LCOs has also been shown in a variety of legumes (Granqvist *et al.,* 2015). However, root hair curling is not universal: in some legume lineages, such as the genistoid and dalbergioid clades, nodule bacteria enter through the middle lamellae between intact root epidermal cells (Fig. 1B; Sprent, 2001). Other lineages, such as the aeschynomenoid genera *Stylosanthes* and *Arachis,* are infected through cracks in the epidermis, such as where lateral roots emerge (Fig. 1B; Sprent, 2001). The diversity of infection mechanisms, with multiple convergent derivations and reversions (Fig. 1B), suggest that these traits may be fairly labile, and morphology and development may relate more to environmental and ecological pressures than phylogenetic origin (Sprent & James, 2007).

*The Diversity of IT Formation*

In *M. truncatula*, microtubule rearrangements form the PIT, with a set of cytoplasmic bridges across the cortex, leading to the dividing inner cortical cells of the nodule primordium (van Brussel *et al.,* 1992; Yang *et al.,* 1994; Timmers *et al.,* 1999).

As the IT proceeds to grow through the cortical PIT, microtubules accumulate around it and particularly around the cytoplasmic bridges between PIT cells (Timmers *et al.,* 1999), with synthesis of cellulose microfibrils and targeting of pectin-containing vacuoles concentrating around the IT apex (Brewin, 2004). Rhizobia growing in the lumen of the IT are enmeshed in a matrix of plant-derived glycoproteins including Root Nodule Extensin; as the IT continues to grow, older parts of the IT matrix transition from fluid to solid state through hydrogen peroxide-mediated protein cross-linking (Rathbun *et al.,* 2002; Brewin 2004).

Transcellular IT formation likely represents the ancestral condition in legumes and is widely distributed, including in mimosoids and basal lineages such as *Chamaecrista* that may represent independent origins of nodulation (Fig. 1B; Sprent, 2001; Doyle, 2011). IT formation is a trait independent of root hair curling; in species in the Milletteae, a transcellular IT forms in cortical cells after non-root-hair (atrichoblastic) epidermal infection (Cordeiro *et al.,* 1996; Sprent, 2001). Epidermal bacterial entry with no transcellular IT is common in the dalbergioid and genistoid lineages, though extracellular components similar to ITs surround rhizobia as they pass between cells (Fig. 1B; Sprent, 2001; Brewin, 2004; Sprent & James, 2007). In legumes with transcellular ITs, uninfected cells are found in the nodule along with infected cells (Sprent, 2001). In the aeschynomenoid genera *Stylosanthes* and *Arachis* (which are infected through crack entry), no transcellular IT forms, and determinate nodules with uniformly infected tissues arise from infected cortical cells (Lavin *et al.,* 2001).

In most legumes, rhizobia are endocytosed into legume nodule cells to form bacterioids surrounded by a peribacterioid membrane (symbiosomes) in enlarged,

polyploid nodule cells (Fig. 1B; Sprent, 2001). However, the basal genus *Chamaecrista* (Cassieae) and the papilionoids *Hymenolobium* and *Andira* (Dalbergieae), as well as *Cyclolobium*, *Poecilanthe* and *Dahlstedia* (Milletteae) retain rhizobia in modified ITs called "fixation threads" (Fig. 1B; Naisbitt *et al.,* 1992; Sprent, 2001; Capoen *et al.,* 2005; Limpens *et al.,* 2005; Sprent, 2007).

*Nodule Morphology and Development*

Indeterminate nodules with an apical meristem are the ancestral condition in legumes, and all mimosoid and the former "caesalpinoid" legumes have these types of nodules (Fig. 2.2; Doyle, 1998; Sprent, 2001; Sprent, 2007). As an indeterminate nodule grows (in both legumes and non-legumes), there is a gradation in bacterial activity, with cells being infected nearest the meristem (infection zone), followed by an area where nitrogen-fixation occurs (fixation zone), and finally a senescent zone where cells die (Fig. 2.1, 2.2; Sprent, 2001; Pawlowski & Sprent, 2008; Pawlowski & Demchenko, 2012).

The determinate nodule is a derived character with multiple independent origins in papilionoid legumes, having arisen in the dalbergioid lineage, the desmodioid lineage, and the Loteae (Fig. 1B; Doyle, 1998; Sprent, 2007). Amongst determinate nodules, the major distinction is between the "desmodioid" nodules (with lenticels) found in the Phaseoleae, Desmodieae, and Psoraleae and the "aeschynomenoid" nodules (without lenticels) found in dalbergioid and genistoid legumes. Desmodioid nodules export nitrogen as ureides in the phaseoloids and as amides in the Loteae (Sprent & James, 2007).

There is also variation in the organogenesis of nodules in different legumes; in most lineages that form determinate nodules, the nodule primordium is derived from outer cortical cells, whereas indeterminate nodules are generally formed from inner cortical cells (Gualtieri & Bisseling, 2000). In *M. truncatula,* inner cortical cell divisions occur as the PIT forms; however, in *Phaseolus vulgaris* and *Glycine max* (which bear determinate, desmodioid nodules), the first cells to divide are in the hypodermis, directly beneath the root hair (Taté *et al.,* 1994; Calvert *et al.,* 1984; Brewin 2004). In *L. japonicus* (also bearing determinate, desmodioid nodules), the first divisions occur in the middle cortex (Szczyglowski *et al.,* 1998).

## 2.2 Nodulation in non-Legumes

Non-legume nodules show several consistent differences from legume nodules (Fig. 2.1, 2.2) (for reviews, see Hirsch & LaRue, 1997; Pawlowski & Sprent, 2008; Pawlowski & Demchenko, 2012). The IT in actinorhizal lineages differs from that in legumes in that there is no Infection Thread Matrix, and the *Frankia* hyphae are in direct contact with the IT wall (Pawlowski & Demchenko, 2012). In infected cells, *Frankia* are retained in primary cell wall material similar to the "fixation thread;" no actinorhizal species endocytose bacteria into symbiosomes, as most legumes do (Berg, 1999a,b; Pawlowski & Demchenko, 2012). It has been suggested that legume nodules developmentally resemble a shoot, while actinorhizal nodules resemble a root (eg, Franche *et al.,* 1998a; Doyle, 1998; Laplaze *et al.,* 2000; Pawlowski & Sprent, 2008). While the nodule primordium in legumes is formed from cell divisions in the root cortex and has peripheral vasculature like a stem, in actinorhizal species the nodule is formed

from the pericycle and has a central vasculature, like a lateral root (Laplaze *et al.,* 2000; Pawlowski & Sprent, 2008). All non-legume nodules are indeterminate, while many legume clades have determinate nodules (Pawlowski & Sprent, 2008).

However, non-legume nodules also show substantial variation in morphology (Fig 2.2). Infection mechanism is determined by the host species, and varies across lineages. The same *Frankia* strain (HFPGpI1) infects *Gymnostoma* (Casuarinaceae) intracellularly through deformed root hairs and *Shepherdia* (Elaeagnaceae) intercellularly through the middle lamella of epidermal cells (Racette & Torrey, 1989).

The only non-legume nodulator that associates with rhizobia is the genus *Parasponia,* in the Cannabaceae family of the Rosales (Lancelle & Torrey, 1984). In *Parasponia*, rhizobia enter through cracks in the epidermis subjacent to the formation of multicellular root hairs, and are not endocytosed into symbiosomes but rather form fixation threads (Fig. 2.1; Lancelle & Torrey, 1984). Despite being nodulated by rhizobial bacteria, species in *Parasponia* show a similar organogenetic program to actinorhizal nodulators, with the nodule arising from pericycle cells and a prenodule arising from cortical cell divisions (Fig. 2.1; Lancelle & Torrey, 1985; Laplaze *et al.,* 2000).

In actinorhizal species in the Fagales (Betulaceae, Casuarinaceae and Myricaceae), *Frankia* infect through root hair deformation (curling or branching) and an IT through the cytoplasmic bridges of a PIT, similarly to intracellularly infected legumes (Fig. 2.2, Berg, 1999; Berg, 1999b; Laplaze *et al.,* 2000). *Frankia* hyphae enveloped in IT can cross from one infected cell to another (Berry & Sunell, 1990; Berg, 1999a,b). Outer cortical cell divisions lead to protuberances called "prenodules," which do not form nodules and are distinct from the nodule primordia formed from pericycle cells (Fig 1A;

Laplaze *et al.,* 2000; Pawlowski & Sprent, 2008). *Frankia* infect the prenodule first, and then the nodule primordium; both plant nodulation genes such as the subtilase *CG12* and *Frankia* genes *nifH* are expressed in prenodule tissues, indicating that nitrogen fixation takes place in the prenodule (Franche *et al.,* 1998a; Laplaze *et al.,* 2000). Gualtieri & Bisseling (2000) pointed to the cortical cell divisions of the prenodule as a possible homology between actinorhizal nodulation Fagales and rhizobial nodulation in legumes, though the prenodule does not ultimately become a nodule.

In actinorhizal species in the Rosales, *Frankia* infect intercellularly through the middle lamella of epidermal cells (Racette & Torrey, 1989; Berry & Sunell, 1990; Liu & Berry, 1991; Hirsch & LaRue, 1997; Berg, 1999; Laplaze *et al.,* 2000). IT-like material is deposited in the apoplastic space only after vascular differentiation has occurred in the nodule primordium, and *Frankia* infect the nodule primordium from this space instead of through "invasive hyphae" enveloped in a transcellular IT (Berry & Sunnell, 1990). The plasma membrane of each infected cell invaginates to form new fixation threads in each infected cell, though invasive hyphae have been occasionally observed to cross from one infected cell to another, as in *Ceanothus* (Rhamnaceae) (Berry & Sunnell, 1990; Liu & Berry, 1991; Berg, 1999a,b). In the majority of actinorhizal Rosales, no prenodule is formed; in lineages where a prenodule does arise from cortical cells, such as *Ceanothus* (Rhamnaceae), it is not infected by *Frankia* (Liu & Berry, 1991; Hirsch & LaRue, 1997).

Infection mechanism in the Cucurbitales has not been examined in detail due to their unculturable strain of *Frankia* symbiont, but infection has been observed to proceed through epidermal cells and intercellularly (Berg, 1999; Pawlowski & Demchenko, 2012). In actinorhizal Cucurbitales, invasive hyphae enveloped in IT do infect

transcellularly, but are not preceded by a PIT as in the Fagales (Berg, 1999a; Pawlowski & Demchenko, 2012). Infected cells are multinucleate (Calvert *et al.,* 1979), and are separated from uninfected cells by the stele (Akkermans & Van Dijk, 1981; Swensen, 1996).

Nitrogen fixation by nitrogenase requires significant ATP, best provided by aerobic respiration, but nitrogenase is denatured by oxygen, posing the so-called "oxygen dilemma" (Shah & Brill 1977; Pawlowski & Demchenko, 2012). This issue is somewhat less critical in actinorhizal nodules, due to the oxygen-protective vesicles formed by *Frankia* (Pawlowski & Demchenko, 2012). In legumes, the vascular system of the nodule is embedded in nodule parenchyma tissue surrounding the nodule and forms a turgor-controlled O2 barrier, while actinorhizal nodules are surrounded by periderm, and the vascular system runs through the center of the nodule (Fig 2.1, 2.2; Sprent, 2001; Pawlowski and Sprent, 2008; Pawlowski & Demchenko, 2012). Oxygen levels are regulated through lenticels in some actinorhizal lineages such as *Alnus, Coriaria* and *Datisca* (Torrey, 1976; Tjepkema, 1978). Other lineages, such as *Casuarina,* form lobes (called "nodule roots") that grow agravitropically and develop air chambers to facilitate oxygen diffusion (Silvester *et al.*, 1990; Franche *et al.,* 1998a). Hemoglobins regulate oxygen levels in the nodules of legumes as well as actinorhizal nodulators, though the concentrations vary and these two groups use hemoglobins from different gene lineages (Tjepkema 1983; Roberts *et al.,* 1985; Pawlowski & Bisseling, 1996; Gopalasubramaniam *et al.,* 2008; Sturms *et al.,* 2010; Vázquez-Limón *et al.,* 2012).

## 2.3 Genetic Basis of Nodulation

*LCO reception*

In model papilionoid legumes, the first step of infection is the plant perception of LCO signals by Lysin-motif receptor kinases (LysM-RKs) such as LjNFR1/LjNFR5 in *L. japonicus* and their orthologs MtLYK3/MtNFP in *M. truncatula* (Table 2.1, Fig. 2.3; Limpens *et al.*, 2003; Madsen *et al.,* 2003; Zhang *et al.,* 2009; De Mita *et al.,* 2014). These LysM-RKs are not required for AM and are upstream of the CSSP, though they are likely paralogous to LysM-RKs involved in AM symbiosis. LysM receptor structure determines nod factor specificity and thus symbiont recognition: *M. truncatula* transformed to express *L. japonicus LjNFR1* and *LjNFR5* will form nodules with *Mesorhizobium loti*, the symbiont of *L. japonicus,* with which wildtype *M. truncatula* will not nodulate (Radutiou *et al.,* 2007). Nod factors directly bind to LysM-RKs, and specificity is determined by the LysM2 domain of LjNFR5 (Bensmihen *et al.,* 2011; Broghammer *et al.,* 2012).

LysM receptor kinases in the *LYK* family have expanded in number due to retention after successive rounds of whole-genome, segmental and tandem duplication in flowering plants (Arrighi *et al.,* 2006; Zhang *et al.,* 2009; De Mita *et al.,* 2014), yielding three major LysM-RK clades supported by exon structure, kinase domain conservation and sequence phylogeny (Lohmann *et al.,* 2010). The clades containing *LjNFR5* and *LjNFR1* diverged from one another in a whole genome duplication event before the divergence of monocots and eudicots (Zhang *et al.,* 2009; De Mita *et al.,* 2014), with multiple subsequent tandem duplications in each clade yielding multiple paralogs (Zhang *et al.,* 2009; De Mita *et al.,* 2014). LysM-RKs determine symbiont specificity in legumes

(Radutoiu *et al.,* 2007), and their proliferation and divergence may have allowed these paralogous receptors to expand host range and coevolve with different symbionts (Michelmore & Meyers, 1998; Op den Camp *et al.,* 2011). While the *LjNFR1* (*MtLYK3*) and *LjNFR5* (*MtNFP*) clades are under purifying selection, two clades paralogous to *LjNFR5* are undergoing diversifying selection (Zhang *et al.,* 2007).

*NFR5* was likely recruited for nodulation from a LysM-RK mediating mycorrhization; orthologs are present in many non-nodulating lineages (Zhu *et al.,* 2006) and induced during mycorrhization (Gomez *et al.,* 2009). However, recently it has been shown that *NFR5/MtNFP* is pseudogenized in multiple lineages in the Rosales, including *Malus domestica, Morus notabilis, Trema levigata,* and *Prunus persica* (van Velzen *et al.,* 2018). The *NFR5* clade underwent one legume-specific duplication, with several subsequent tandem duplications in different legume lineages (Zhang *et al.,* 2007; Streng *et al.,* 2011; De Mita *et al.,* 2014). In *Parasponia andersonii,* a LysM-RK orthologous with *NFR5* (*PaNFP2*) was found by RNAi to mediate both nodulation and AM (Op den Camp *et al.,* 2011; Streng *et al.,* 2011), though this finding may have been the result of RNAi inhibition of the paralogous gene *PaNFP1* (van Velzen *et al.,* 2018).

In contrast, *NFR1* was likely recruited from innate immune pathway - the *NFR1* clade is sister to *CERK1*, a LysM-RK gene clade involved in pathogen response (Kaku *et al.,* 2006; De Mita *et al.,* 2014). In *Arabidopsis thaliana,* AtCERK1 is a LysM-RK involved in chitin perception for innate immunity against fungal pathogens (Miya *et al.,* 2007; Zhang *et al.,* 2009; Nakagawa *et al.,* 2011). In *Oryza sativa* OsCERK1 mediates chitin-triggered immune responses in a heterocomplex with the LysM-containing protein CEBiP (Kaku *et al.,* 2006; Shimizu *et al.,* 2010). Fascinatingly, OsCERK1 also mediates

AM symbiosis, suggesting several functions for this chitin receptor in hetercomplexes with other receptor kinases, or several neofunctionalization and subfunctionalization events for different symbiotic functions across the plant phylogeny (Miyata *et al.,* 2014). A chimeric receptor with the extracellular region of NFR1 and the intracellular region of CERK1 does not rescue nodulation in deficient *nfr1-4 L. japonicus* mutants, showing that the CERK1 intracellular signaling differs from NFR1 despite high levels of homology in the proteins (Nakagawa *et al.,* 2011). Introducing three amino acids from (YAQ) from the NFR1 EF/F loop to this chimeric protein is sufficient to restore the nodulation phenotype (Nakagawa *et al.,* 2011).

*CSSP: Nuclear Calcium Spiking*

Nuclear calcium spiking is a critical step in the CSSP, leading to the expression of several downstream transcription factor genes involved in nodule and arbuscule induction. The SYMRK kinase domain interacts with the mevalonate synthase HMGR1 to initiate nuclear calcium spiking in response to rhizobial and AM fungal LCOs (Venkateshwaran *et al.,* 2015), and with the transcription factor SIP1, which is also required for nodulation and AM and induces expression of downstream transcription factor genes such as *NIN* (Kevei *et al.,* 2007; Zhu *et al.,* 2008; Wang *et al.,* 2013). *SYMRK* is necessary for both nodulation (actinorhizal and rhizobial) in the NFC and AM colonization across angiosperms, and is generally considered the starting point of the CSSP, meaning that this single-copy ortholog gained a new function to accommodate nodulation (Table 2.1; Stracke *et al.,* 2002; Radutiou *et al.,* 2003; Kistner *et al.,* 2005; Markmann *et al.,* 2008; Gherbi *et al.,* 2008).

MtMCA8, a calcium pump in the SERCA-type family, calcium channel MtCNGC15 and potassium channels LjCASTOR (MtDMI1) and LjPOLLUX are all required for calcium spiking (Table 1, Fig. 2; Ané *et al.,* 2004; Chen *et al.,* 2009; Capoen *et al.,* 2011; Charpentier *et al.,* 2016). Nuclear pore proteins NUP85, NUP133 and NENA form a complex required for symbiotic calcium spiking in a temperature-dependent fashion, possibly by helping locate ion channels and pumps on the inner nuclear membrane (Kanamori *et al.,* 2006; Saito *et al.,* 2007; Groth *et al.,* 2010; Oldroyd, 2013).

Nuclear calcium spiking is perceived by the calcium- and calmodulin-dependent serine/threonine protein kinase LjCCAMK (MtDMI3) (Table 1; Fig. 2; Lévy *et al.,* 2004; Mitra *et al.,* 2004), through complex conformational changes (Shimoda *et al.,* 2012; Miller *et al.,* 2013; Poovaiah *et al.,* 2013). Activated CCAMK interacts with and phosphorylates the LjCYCLOPS (MtIPD3) transcription factor, and this interaction is required for both nodulation and AM formation, as part of the CSSP (Table 1; Messinese *et al.,* 2007; Yano *et al.,* 2008; Madsen *et al.,* 2010; Horvath *et al.,* 2011). *L. japonicus* mutants with autoactive CCAMK spontaneously form nodules in the absence of rhizobia; double mutants with autoactive CCAMK and *cyclops* loss-of-function fail to form IT, suggesting that *CYCLOPS* is required for cross-signaling between the organogenesis and IT formation pathways (Madsen *et al.,* 2010; Singh *et al.,* 2014).

Downstream of nuclear calcium spiking, several transcription factor genes are induced in a complex cascade that is still not fully understood (Limpens & Bisseling, 2014; Soyano & Hayashi, 2014; Gamas *et al.,* 2017). Phosphorylated CYCLOPS induces the expression of *NIN,* a transcription factor involved in IT and nodule formation (Singh *et al.,* 2014; Soyano & Hayashi, 2014). *NIN* induces the expression of the CCAAT-box

binding transcription factors NF-YA1, NF-YB1 and NF-YC1, which form a heterotrimeric complex with each other, involved in cortical cell divisions and IT formation (Soyano *et al.,* 2013; Battaglia *et al.,* 2014; Singh *et al.,* 2014; Soyano & Hayashi, 2014). The *NF-YA1* (*MtHAP2-1*) is involved in cortical cell divisions and the formation of the nodule meristem (Combier *et al.,* 2006; Laloum *et al.,* 2013). The GRAS domain transcription factor genes *NSP1* and *NSP2* are both required for nodulation (Oldroyd & Long, 2003; Kalo *et al.,* 2005; Smit *et al.,* 2005; Singh *et al.,* 2014). These genes form a heterocomplex during nodulation to associate with Nod-factor inducible promoters (Hirsch *et al.,* 2009), including NIN and the additional transcription factor ERN1 (Cerri *et al.,* 2012) that are essential for nodulation (Marsh *et al.,* 2007; Middleton *et al.,* 2007). The activation of CCAMK and its phosphorylation of the CYCLOPS transcription factor is required for IT and nodule initiation, as is cytokinin signaling mediated through the cytokinin receptor *LHK1* (Madsen *et al.,* 2010).

For the most part, CSSP genes are inherited as single-copy orthologs that mediate both AM and nodulation (Duc *et al.,* 1989; Messinese *et al.,* 2007; Banba *et al.,* 2008; Wang *et al.,* 2010; Delaux *et al.,* 2013b; Delaux *et al.,* 2015). The CSSP is required for the AM symbioses across land plants; *MtDMI1* (*LjCASTOR*), *MtDMI3* (*LjCCAMK*) and *MtIPD3* (*LjCYCLOPS*) have been found in nearly all major land plant lineages (Wang *et al.,* 2010). The recruitment of CSSP genes for nodulation did not involve gene duplication (with exceptions like the paralogous *CASTOR* and *POLLUX*), so recruitment for nodulation involved expansion of function (Markmann *et al.,* 2008), instead of neofunctionalization or subfunctionalization of paralogs. As a result, CSSP mutants are deficient in both AM and nodulation (Duc *et al.,* 1989; Parniske, 2008). *L. japonicus*

*castor, symrk, nup133,* mutants are impaired in their ability to form nodules or arbuscules (Kistner *et al.,* 2005). In *M. truncatula, ccamk* mutant are impaired in nodulation and AM, and RNAi silencing of *MtMCA8* and *HMGR1* blocks nuclear calcium spiking in response to AM and rhizobial LCOs (Lévy *et al.,* 2004; Capoen *et al.,* 2011; Venkateshwaran *et al.,* 2015).

This raises two related questions (Markmann *et al.,* 2008; Bonfante & Requena, 2011; Genre & Russo, 2016). First, if the CSSP is present across land plants, how is it that nodulation only evolved in the NFC? Second, how can a single signaling pathway (the CSSP) discriminate between rhizobia and glomeromycetes to create different downstream phenotypes (IT and nodule for nod factors, or pre-penetration apparatus and arbuscule for myc factors)? These questions are related through signal transduction "bandwidth." In order for the single-copy orthologs of the CSSP to be successfully recruited for nodulation while maintaining AM functionality (as most nodulators do, with a few exceptions like *Lupinus*), the pathway must be able to discriminate between the two signals to induce different downstream genes associated with nodulation or AM. The first question has been approached most thoroughly using *SYMRK* as a case study and the second using *CCAMK*. Both genes have been the focus of much work on the evolutionary origin of nodulation, and are perhaps the most extensively studied CSSP genes in actinorhizal lineages (Markmann *et al.,* 2008; Gherbi *et al.,* 2008; Wang *et al.,* 2010; Svistoonoff *et al.,* 2013).

Like most CSSP genes, *SYMRK* (*MtDMI2*) is required for the AM symbiosis and is present across land plants, and is present as a single-copy ortholog required for both nodulation and AM in nodulating lineages (Kistner *et al.,* 2005; Delaux *et al.,* 2013b;

Delaux *et al.,* 2015). Unlike the upstream LysM-RKs, SYMRK is not involved in symbiont specificity, as shown by complementation between *M. truncatula* and *L. japonicus,* but is required for signal transduction from LCO signal to nuclear calcium spiking (Kistner *et al.,* 2005; Markmann *et al.,* 2008). Domain gains in *SYMRK* have been proposed to contribute to a genetic predisposition to evolve nodulation; the single-copy *SYMRK* ortholog varies substantially in domain architecture in different angiosperm lineages, having three leucine-rich repeats in eurosids and two outside the eurosids (Markmann *et al.,* 2008). The "full-length" (three leucine-rich repeats) SYMRK ortholog from the non-nodulating *Tropaeolum* can rescue nodulation in mutant *L. japonicus,* whereas the "reduced-length" (two LRR's) copy of *O. sativa* and *Solanum lycopersicum* rescues AM but not nodulation (Markmann *et al.,* 2008). However, these gains are specific to all eurosids, and not only the NFC, so could not completely explain the NFC-specific predisposition to nodulate (Markmann *et al.,* 2008). A similar pattern is seen in *DMI3/CCAMK; DMI3* homologs from some bryophytes can rescue AM, but not nodulation, in *M. truncatula dmi3* mutants (Wang *et al.,* 2010). However, other CSSP orthologs show complete functional equivalence; *Oryza sativa cyclops* mutants cannot form AM, and functional *CYCLOPS* from rice can restore nodulation in *L. japonicus*, suggesting functional conservation of *CYCLOPS* (Yano *et al.,* 2008).

There are no detectable differences in nuclear calcium spiking pattern between AM and nodulation symbioses (Sieberer *et al.,* 2012; Oldroyd, 2013; Singh *et al.,* 2014). However, the downstream transcription factor genes induced by nuclear calcium spiking are different, and this induction requires the interaction of *CCAMK* and *CYCLOPS* (Madsen *et al.,* 2010; Soyano & Hayashi, 2014; Singh *et al.,* 2014). Shimoda *et al.* (2012)

showed that a mutation in the calmodulin-binding domain of *CCAMK* in *L. japonicus* suppressed CCAMK activation in rhizobial infection but not in mycorrhization, suggesting that differential calmodulin binding might partially account for discrimination between nodulation and AM. However, the lily ortholog of the *CCAMK* can restore the nodulation phenotype in nod-deficient *M. truncatula ccamk* mutants, and the rice ortholog of *CCAMK* (*OsCCAMK*) can rescue nodulation in nod-deficient *L. japonicus ccamk* mutants (Gleason *et al.,* 2006; Banba *et al.,* 2008). The GRAS transcription factor NSP2 in interacting with RAM1 during mycorrhization and NSP1 during nodulation may also play a role in differentiating the two symbioses (Hirsch *et al.,* 2009; Maillet *et al.,* 2011; Gobbato *et al.,* 2012).

The differential action of transcription factors downstream of the CSSP, such as nodulation-specific *NIN* and the *NF-Y* complex, likely plays a role in signal differentiation (Oldroyd, 2013). It's also possible that, in addition to or instead of sequence evolution in CSSP genes, as-yet unidentified proteins (or other molecules) act to differentiate the two signals. The origin of genes underlying such a mechanism, possibly through duplication and neofunctionalization, could account for the NFC predisposition to nodulate, though no evidence for this exists yet. Finally, rhizobia and glomeromycetes have different infection mechanisms, with rhizobia generally entering through root hairs and glomeromycetes entering through epidermal tissue (atrichoblasts); Genre & Russo (2016) propose that these spatial differences may contribute to signal differentiation. However, this would not explain signal differentiation in lineages in which nodulating bacteria infect through epidermal cells, as in genistoid and dalbergioid legumes, or actinorhizal Rosales or Cucurbitales.

*IT and Symbiosome Formation*

Downstream of induction of *NIN*, NF-Y complex and GRAS transcription factors, *NAP1* and *PIR1* (both members of the *SCAR/WAVE* complex) are required for actin cytoskeleton rearrangements during IT formation (Yokota *et al.,* 2009). Even with constitutively activated CCAMK, *PIR1, CERBERUS* and *NAP1* are all required for proper infection thread progression (Madsen *et al.,* 2010). *MtLIN* (*LjCERBERUS*) and *VAPYRIN* are required for proper IT formation (Kuppusamy *et al.,* 2004; Kiss *et al.,* 2009; Yano *et al.,* 2009; Pumplin *et al.,* 2010; Murray *et al.,* 2011). Both *M. truncatula lin-4* (*cerberus*) mutants and *vapyrin* mutants fail to form IT and, interestingly, form nodules with a central vasculature more similar to non-legume nodules (Guan *et al.,* 2013). Subtilases, a class of protease, are also induced during both AM and nodulation; In *L. japonicus,* the subtilases *LjSBTS* and *LjSBTM4* are required for proper IT formation and nodulation as well as arbuscule formation (Kistner *et al.,* 2005; Takeda *et al.,* 2009). These genes localize to the apoplastic space surrounding growing ITs and nodules, and likely play a role in cell wall restructuring in the IT and symbiosome (Takeda *et al.,* 2009).

In legumes in which rhizobia are endocytosed into intracellular bacterioids (including *M. truncatula* and *L. japonicus*), the formation of the infection droplet is mediated by exocytotic pathway proteins called SNAREs (soluble *N*-ethylmaleimide sensitive factor attachment protein receptor); t-SNAREs (such as syntaxins) attach to the target membrane while v-SNAREs (such as the VAMP72 family in plants) attach to the vesicle membrane, forming a complex (Kwon *et al.,* 2008; Ivanov *et al.,* 2012). Interestingly, these genes are also required for the formation of the periarbuscular

membrane in AM (Ivanov *et al.,* 2012). The syntaxin MtSYP132 is a t-SNARE that localizes specifically to the membrane surrounding ITs, infection droplets and symbiosomes in *M. truncatula* during infection by *Sinorhizobium meliloti* (Catalano *et al.,* 2007; Limpens *et al.,* 2009). The v-SNARE *VAMP721e/VAMP721d* is required for nodule and arbuscule formation in *M. truncatula* (Ivanov *et al.,* 2012). RNAi silencing of *VAMP721e* and *VAMP721d* blocks symbiosome and arbuscule formation, and infection droplets are greatly reduced and deformed (Ivanov *et al.,* 2012).

The genes involved in IT formation were recruited from both the AM pathway and other pathways such as lateral root development, creating a novel developmental program for IT formation with a mosaic evolutionary history (Yokota & Hayashi, 2011; Soyano & Hayashi, 2014). *NIN* is not involved in AM or a part of the CSSP; *nin* mutants of *L. japonicus, M. truncatula,* and *Pisum sativum* have disrupted IT and nodule formation, but no disruption of AM infection (Table 1; Schauser *et al.,* 1999; Borisov *et al.,* 2003; Marsh *et al.,* 2007). *NIN* and the related *NIN*-like proteins (NLPs) are widespread in angiosperms; *NLP7* mediates gene expression changes due to nitrate signaling in *A. thaliana* (Schauser *et al.,* 2005; Soyano & Hayashi, 2014). Recently it was shown that *NIN* is pseudogenized in multiple non-nodulating lineages in the Rosales, including *Malus domestica, Morus notabilis, Trema levigata,* and *Prunus persica* (van Velzen *et al.,* 2018).

Overexpression of the NF-Y complex induces lateral root formation, and NF-YC1 interacts with the GRAS protein SIN1 during both nodulation and lateral root development (Battaglia *et al.,* 2014). These data suggest that NF-Y genes were recruited from a lateral root pathway, and indeed the large NF-Y (CCAAT-box binding)

transcription factor family includes several homologs identified as being involved in lateral root formation (Soyano *et al.,* 2013; Laloum *et al.,* 2013; Soyano & Hayashi, 2014). The v-SNAREs and t-SNAREs involved in bacterial endocytosis are related to genes involved in microbial pathogen defense (Kwon *et al.,* 2008; Ivanov *et al.,* 2012). The genes *NAP1* and *PIR1,* involved in cytoskeletal rearrangement during IT formation, are also not involved in AM; IT-deficient *L. japonicus* mutants such as *pir1* and *nap1* form AM associations (Albrecht *et al.,* 1999; Yokota *et al.,* 2009; Yokota & Hayashi, 2011).

The evolutionary origin of the PIT and IT did also involve the recruitment of genes involved in the AM symbiosis (Journet *et al.,* 2001; Brewin, 2004; Balestrini & Bonfante, 2005; Takeda *et al.,* 2009). AM fungi (glomeromycetes) that infect intracellularly (Parish type) are directed to the site of arbuscule formation by transcellular bridges of the "pre-penetration apparatus," similarly to the PIT (Harrison 1997; Genre *et al.,* 2008). The formation of the IT and the pre-penetration apparatus both direct microsymbionts to involve the directed deposition of the plant cell wall-derived materials around the microsymbiont, and there are a few common genes underlying the two. For example, the GRAS transcription factor *NSP1* and *NSP2* are required for nodulation and also involved in AM, though the heterocomplexes they form are different in the two symbioses (Maillet *et al.,* 2011; Oldroyd 2013; Shtark *et al.,* 2016). The orthologs of these genes from *Oryza sativa* can rescue nodulation in *L. japonicus nsp1* and *nsp2* mutants, showing functional conservation in these genes between AM (Yokota *et al.,* 2009).

*LjSBTM4,* a subtilase required for AM and also induced during nodulation in *L. japonicus*, is in a monophyletic gene lineage with subtilases involved in pathogen defense in the asterid *Solanum lycopersicum* (Taylor & Qiu, 2017). Successive rounds of both whole-genome and tandem duplication in the legumes produced paralogs that are differentially expressed during AM and nodulation in *L. japonicus*. Several different orthologous gene lineages that diverged during the origin of angiosperms have been differentially recruited for nodulation in actinorhizal and legume nodulation. *CG12*, a nodulation-specific subtilase, is expressed only in IT-containing cells in *Casuarina glauca*, and also in transgenic legumes, indicating conservation in transcriptional regulation (Svistoonoff *et al.,* 2003; Svistoonoff *et al.,* 2004). However, *CG12* and its ortholog *AG12* from *Alnus glutinosa* are in a different orthologous lineage from *LjSBTM4* and *LjSBTS*, the subtilases induced during nodulation in *L. japonicus* (Takeda *et al.,* 2009; Taylor & Qiu, 2017).

The IT and pre-penetration apparatus of AM are both derived from primary cell wall materials, including b-1,4-glucans, non-esterified homogalacturonans, xyloglucans, hydroxyproline-rich glycoproteins, and arabinogalactans (Bonfante, 2001; Balestrini & Bonfante, 2005). Many of the "early nodulin" genes induced by LCOs encode arabinogalactans and hydroxyproline-rich glycoproteins that constitute primary cell wall material laid down in the lumen of the IT during infection (Cassab, 1998; Rathbun *et al.,* 2002; Brewin, 2004). In *M. truncatula, MtENOD11*, which encodes a cell wall repetitive proline-rich protein, is induced during both early nodulation and pre-penetration apparatus formation and cortical cell colonization during AM, though this gene is also expressed in a variety of cell types involved in metabolite exchange or cell structural

changes (Journet *et al.,* 2001; Chabaud *et al.,* 2002). It is likely that the particular genes recruited vary across different independent origins of nodulation. In actinorhizal nodules, *Frankia* is never fully released into the cytoplasm of nodule cells, instead being encapsulated in plant-cell derived material made primarily of pectins and non-sulfated, de-esterified polygalacturonic acids (Lalonde & Knowles, 1975; Franche *et al.,* 1998a).

## 2.4 The Evolutionary Origins of Nodulation

*Phylogenetic Distribution of Nodulation Genes in Nodulating Lineages*

The recruitment of genes from the ancestral CSSP for nodulation has been most thoroughly examined in the legumes, and specifically the model papilionoids *M. truncatula* and *L. japonicus* (Catoira *et al.,* 2000; Kistner *et al.,* 2002; Kistner *et al.,* 2005)**.** However, several lines of evidence show multiple nonlegume nodulating lineages also employ genes from the CSSP to form their nodules (Table 1; Gherbi et al., 2008; Streng et al., 2011; Pawlowski et al., 2011; Svistoonoff *et al.,* 2013; Svistoonoff *et al.,* 2014). Nuclear calcium spiking in response to nod factors from *Sinorhizobium fredii* has been observed in *P. andersonii* (Granqvist *et al.,* 2015; Chabaud *et al.,* 2016). In *A. glutinosa* and *C. glauca,* calcium spiking in root hairs is induced by exudates from *Frankia* (Granqvist *et al.,* 2015; Chabaud *et al.,* 2016).

Genetic characterization through RNAi knockdown or mutant complementation studies have shown that CSSP genes are required for nodulation in the actinorhizal species *Datisca glomerata* (Markmann *et al.,* 2008), *A. glutinosa* and *C. glauca* (Franche *et al.,* 2016). In *C. glauca, CgSYMRK* is necessary for nodulation, and can rescue both nodulation and AM in root symbiosis-deficient *L. japonicus symrk* mutants (Gherbi *et al.,*

2008). *DgSYMRK* is also required for nodulation in *D. glomerata,* and can rescue both nodulation and AM in root symbiosis-deficient *L. japonicus symrk* mutants (Markmann *et al.,* 2008). In *C. glauca, CgCCAMK* is necessary for nodulation with *Frankia*, and the *CgCCAMK* from *C. glauca* can restore both nodulation and AM in *M. truncatula ccamk* mutant lines (Svistoonoff *et al.,* 2013). Further, auto-active *CgCCAMK* induces nodulation in both *C. glauca* (which is infected intracellularly through root hair curling) and *Discaria trinervis* (Rhamnaceae), a distantly related actinorhizal species infected intercellularly (Svistoonoff *et al.,* 2013). In the rhizobial *P. andersonii,* a LysM-RK and CCAMK mediates both nodulation and AM (Op den Camp *et al.,* 2011; Streng *et al.,* 2011).

Transcriptomic studies have shown that several genes induced during nodulation in actinorhizal species are homologous to legume nodulation genes. These include both CSSP genes recruited from the AM symbiosis, and non-CSSP genes recruited from other pathways (Table 1), for example, *NIN, SYMREM1, VAPYRIN, CERBERUS,* and *PUB1* in *D. glomerata* (Demina *et al.,* 2013)*, SYMREM1, CASTOR, NUP133, CCAMK, CYCLOPS* and *HMGR1* in *C. glauca* and *SYMREM1, SYMRK, CCAMK, NSP1,* and *HMGR1* in *A. glutinosa* (Hocher *et al.,* 2011; Tromas *et al.,* 2012).

Some nodulatio genes outside of the CSSP have also been characterized in non-legume lineages. *NIN* has also been recruited for nodulation in actinorhizal lineages; RNAi knockdown of *NIN* expression in *C. glauca* showed reduced root hair deformation and nodulation (but not mycorrhization) (Clavijo *et al.,* 2015). Hemoglobins are found in the nodules of legumes as well as actinorhizal nodulators, though the concentrations vary

and these two groups use hemoglobins from different gene lineages (Tjepkema 1983; Roberts *et al.,* 1985; Pawlowski & Bisseling, 1996; Vázquez-Limón *et al.,* 2012).

*Paralogy in Nodulation Genes*

Nodulation is a deeply homologous trait, and the predominant evolutionary story in its origins is the repeated independent recruitment of AM genes, and particularly of the orthologous genes of the CSSP that render mutants both myc- and nod- deficient (Duc *et al.,* 1989; Parniske, 2008; Svistoonoff *et al.,* 2014). For this recruitment of orthologs to allow for effective nodulation while still accommodating AM symbioses (as most nodulators do, with some exceptions such as *Lupinus*), the CSSP had to have increased its "bandwidth" in the NFC, to be able to discriminate between the two symbionts and induce the specific downstream genes required for each symbiosis. Some CSSP orthologs from outside the NFC, such as *CYCLOPS,* can restore nodulation in legume mutants (Yano *et al.,* 2008). In other cases non-NFC orthologs, such as *SYMRK* (Markmann *et al.,* 2008), or *CCAMK/DMI3* (Wang *et al.,* 2010) can recover an AM phenotype but not nodulation in the corresponding legume mutants.

In cases of deep homology, the independent recruitment of orthologous genes for a convergent function would be favored over paralogous genes, since these genes are more likely than paralogs to have similar function (Kondrashov *et al.,* 2002; Gabaldón & Koonin, 2013). However, similarly to the repeated parallel evolution of C4 photosynthesis by recruitment of the C4 genetic toolkit (Christin *et al.,* 2013), some of the homologous genes recruited for nodulation in different lineages are paralogous to one another (Vázquez-Limón *et al.,* 2012; Taylor & Qiu, 2017). In the origins of nodulation,

these differentially recruited paralogs include LysM-RKs which mediate symbiont specificity (Op den Camp *et al.,* 2011; Streng *et al.,* 2011; De Mita *et al.,* 2014), subtilases that are involved in restructuring the symbiont membrane interface (Laplaze *et al.,* 2000; Takeda *et al.,* 2009; Taylor & Qiu, 2017), and hemoglobins that mediate oxygen regulation in the nodules (Vázquez-Limón *et al.,* 2012).

If genes involved in nodulation in different lineages are orthologous, that does not necessarily constitute evidence of homologous inheritance of the trait, since these orthologs may have been independently recruited; however, the recruitment of paralogs does constitute evidence of independent evolution (Abouheif *et al.,* 1997; Doyle, 2016). Indeed, Doyle noted this as early as 1994:

"Only if paralogous, and not orthologous genes were recruited to be nodulins in two different plant lineages…could gene trees suggest independent recruitment and hence independent origins of nodulation in the two groups"

The independent recruitment of paralogous genes has implications not only for the assessment of nodule homology, but also for the phylogenetic distribution of any evolutionary "predisposition" to nodulate and the nature of the resulting nodules. The phylogenetic depth (and lineage-specific retention) of different paralogous genes could determine what precise type of nodule a given plant lineage is predisposed to evolve. The hemoglobins expressed in nodules represent an instance of differential recruitment of derived paralogs in the independent origins of nodulation. Actinorhizal lineages express the (confusingly named) "non-symbiotic" hemoglobins in the *nsHb-1* lineage in nodules

(Franche *et al.,* 1999b; Vázquez-Limón *et al.,* 2012). The leghemoglobins expressed in legume nodules are a distinct paralogous gene clade specific to the legumes, differentiated from non-symbiotic hemoglobins (*nsHb-1* and *nsHb-2*) by decreases in the size of the N- and C-terminal regions and a hexacoordinate to pentacoordinate transition at the heme-Fe (Garrocho-Villegas & Arredondo-Peter, 2008; Gopalasubramaniam *et al.,* 2008). Both *nsHb-1* and leghemoglobins are present in papilionoids, and a hemoglobin that is intermediate between leghemoglobin and *nsHb-2* is present in the basal legume *Chamaecrista fasciculata* (Gopalasubramaniam *et al.,* 2008; Vázquez-Limón *et al.,* 2012)**.**

There is some regulatory convergence between the two lineages of hemoglobins; promoter regions of the *C. glauca nsHb-1* gene can functionally direct leghemoglobin expression in *L. japonicus* nodules, and likewise the promoter regions of *Glycine max* and *P. andersonii* can functionally direct nsHb-1 expression in *C. glauca* and *Allocasuarina verticillata* (Casuarinaceae) (Jacobsen-Lyon *et al.,* 1995; Franche *et al.,* 1998b). However, the pattern of expression in these transgenic roots is not exactly the same as their native hemoglobins; for example, *L. japonicus* expresses *P. andersonii* nsHB-1 in uninfected cells in the nodule (Franche *et al.,* 1998b; Gualtieri & Bisseling, 2000). Taken together, these data suggest that *cis*-regulatory regions remained relatively conserved following the duplication of leghemoglobins from the *nsHb-1* lineage, and that similar *trans*-acting regulatory genes were recruited for nodulation independently in different nodulating lineages.

This differential recruitment of paralogs could affect how legume and non-legume nodules regulate oxygen levels, and may help explain structural differences such

as the peripheral vasculature of legume nodules and the central vasculature of non-legume nodules, or the "nodule roots" of some actinorhizal lineages (Pawloski & Sprent, 2008). It may also help explain differences in symbionts, as *Frankia* forms vesicles that help regulate oxygen for nitrogen fixation while rhizobia do not; however, the rhizobial nodulation of *Parasponia* complicates this possibility, since it is nodulated by *Bradyrhizobium* but expresses *nsHb-1* hemoglobins and has a central vasculature (Lancelle & Torrey, 1984; Franche *et al.,* 1998b). Additionally, the subtilases shown to be required for nodulation in *C. glauca* and *A. glutinosa* are paralogous to those required for nodulation in *L. japonicus* (Taylor & Qiu, 2017). Subtilases are involved in protein turnover during IT progression (Laplaze *et al.,* 2000; Takeda *et al.,* 2009). This could be related to structural differences in ITs in legumes, in which rhizobia are in direct contact with the IT matrix, and those in actinorhizal Fagales, in which there is no IT matrix and *Frankia* is in direct contact with the IT lumen wall (Pawlowski & Demchenko, 2012).

Paralogs that arose in specific clades after they evolved nodulation might also be responsible for derived nodule characteristics in the clades. For example, LysM domain receptor kinases mediating Nod-factor reception in *L. japonicus* (LjNFR1a, LjNFR1b, and LjNFR1c) arose by tandem duplication in the papilionoids (Limpens et al., 2003; Streng *et al.,* 2011; De Mita et al., 2014). This pattern allowed these paralogous receptors to coevolve with different symbionts with less constraint, and could play a role in the sophisticated discrimination between symbionts in the papilionoid clade (Michelmore & Meyers, 1998; Radutoiu et al., 2007; De Mita *et al.,* 2014). This in turn could help explain the widespread persistence of nodulation in papilionoid legumes.

*Evolutionary Tinkering*

Traits are often assembled by the recruitment of different genes from different pathways in "modules" (Jacob, 1977), resulting in novel pathways with mosaic evolutionary histories. Therefore, it can be appropriate to talk about *features* of a trait that are homologous, or degrees of homology in an organ (Roth, 1984). Discussions of the structural homologs of nodules have been ongoing for the better part of a century, because of their similarity to many different plant structures (Hirsch & LaRue, 1997). Nodules have been variously compared to lateral roots (Nutman, 1948), shoots (Sprent, 1989), arbuscule of AM (Duc *et al.,* 1989), induction of a wound meristem (Baron & Zambryski, 1995), or an organ *sui generis* (Libbenga & Boyers, 1974). There are morphological similarities in root hair response to rhizobial LCOs and invasion by root-knot nematodes (Weerasinghe *et al.,* 2005). Now, as the genetic basis of nodulation is being revealed, we do find that different elements of the nodulation pathway have different evolutionary histories (Yokota & Hayashi, 2011; Soyano & Hayashi, 2014).

While most identified genes in the nodulation pathway were recruited from the AM symbiosis (Parniske, 2008), some were recruited from innate immunity, lateral root formation, or even pollen tube development (Yokota & Hayashi, 2011; Soyano & Hayashi, 2014). The NF-Y group of transcription factors appears to have been recruited from lateral root development (Soyano & Hayashi, 2014). Plant innate immunity involves many processes similar to the CSSP, including detecting bacterial and fungal factors and remodeling cell wall and cell membrane structures, so it is perhaps unsurprising that many genes involved in nodulation, such as LysM-RKs, v-SNAREs, t-SNAREs, and

subtilases are related to genes involved in microbial pathogen defense (Kwon *et al.,* 2008; Nakagawa *et al.,* 2011; Ivanov *et al.,* 2012; Taylor & Qiu, 2017).

The resulting nodulation pathway has a mosaic evolutionary history, reflecting the evolutionary tinkering that built it. As different instances of nodulation are more fully genetically characterized, it would not be surprising to find more instances of differential recruitment of orthologous, paralogous, and perhaps even unrelated genes mediating convergent processes in different lineages. However, common selective pressures arising from constraints imposed by the symbiosis bias recruitment towards homologous genes, and particularly orthologous genes.

*To what degree are nodules homologous?*

Some of the similarities (and differences) seen in different nodulating lineages likely reflect physical constraints imposed by the biology of the plant lineage (Wake *et al.,* 2011). As a facile example, infection cannot proceed by root hair curling in the millettioid legume *Lonchocarpus muehlbergianus,* which is infected through non-root hair epidermal cells, because this species lacks root hairs (Cordeiro *et al.,* 1996). This applies on a genetic level as well; the different paralogs recruited for nodulation in legumes and non-legumes reflects the genetic material available in different nodulating lineages, since the leghemoglobin lineage arose in legumes (Vázquez-Limón *et al.,* 2012). While most aspects of nodule morphology and developmental appear to be determined by the host plant (Racette & Torrey, 1989), the effect of microsymbiont metabolism, signaling and effectors also cannot yet be ruled out as causative agents of some differences and similarities in nodules. For example, different transporters can be

expected to be expressed in the membrane interface in amide- and ureide-exporting nodules, though this has not yet been demonstrated (Tajima *et al.,* 2004; Pélissier *et al.,* 2004).

Many of the convergences seen in different nodulating lineages likely reflect selective constraints imposed by the nature of nodulation. For example, the independent recruitment of paralogous hemoglobins in nodulating lineages is likely a response to selection posed by the oxygen dilemma of nitrogen fixation. At a certain level, the existence of homologous genes specifying and patterning convergent structures under such constraints becomes so abstract as to be effectively irrelevant; for example, nodules are created by cell division, so homologous genes involved in cell division are highly likely to be recruited in parallel in nodule development. Perhaps a reasonable line to draw here is whether the knocking out these genes results in the loss of wider functionality (or at an extreme, lethality) or whether only nodules themselves and associated structures such as arbuscules are disrupted.

Some elements of nodulation in different lineages are quite similar on a developmental level, such as root hair deformation, IT formation and cortical cell divisions for the prenodule and nodule primordium, respectively, in the actinorhizal Fagales and rhizobial Fabales (Gualtieri & Bisseling, 2000). There are also genetic similarities between different nodulating lineages, such as the employment of orthologous CSSP genes in all examined nodulating lineages (Svistoonoff *et al.,* 2014). However, different instances of nodulation are not simply homologous with one another because they are not phylogenetically contiguous, and because there are substantial differences in their development (e.g, cortical vs. pericycle origin of nodule primordia)

and morphology (e.g, peripheral vs. central vasculature). Thus, these similarities must be considered convergences (or parallelisms) in these lineages based on a deeply homologous genetic endowment, a process between convergence and common inheritance that Abouheif (2008) called "mesoevolution."

Homology has been described as "correspondence by the continuity of information" (van Valen, 1982). The major evolutionary narrative stemming from the genetic characterization of nodulation has been recruitment of the CSSP and other genes from the AM symbiosis, and the sequence information in these genes is indeed homologous in that it is similar by continuous descent from a common ancestor (Parniske, 2008; Markmann *et al.,* 2009). CSSP genes recruited for nodulation in different NFC lineages serve the same function in structure or development, but have not served that function continuously throughout the phylogeny, so their recruitment represents deep homology (Shubin *et al.,* 1997; Abouheif *et al.,* 1997; Doyle, 2013). For example, *SYMRK* in *P. andersonii* (Rosales), *C. glauca* (Fagales) and *M. truncatula* (Fabales) is simply homologous as a gene with similar sequence due to continuous inheritance, and functioning in signal transduction of LCO signals in root endosymbioses in AM (Markmann & Parniske, 2009), but it is deeply homologous, and not simply homologous, as a signal transducer of LCO signals *from nodulating bacteria*, because it was independently recruited for this function in these three lineages.

The origins of nodulation have a complex evolutionary history, showing both deep homology and evolutionary tinkering in gene recruitment, and successive rounds of lineage-specific gene duplications. The resulting nodule structures have a complex phylogenetic distribution of morphologies, developmental programs, and the genes

mediating them, frustrating homology analysis. It is in these complexities, though, that we can understand the evolution of a complex trait, and the effects of constraining selective pressures operating in parallel on different lineages with different evolutionary histories. Nodulation represents not only an important symbiosis for global nutrient cycling and human agriculture and nutrition, but a fascinating case study in the repeated evolution of a complex, symbiotic organ.

**References:**


**Abouheif, E., Akam, M., Dickinson, W.J., Holland, P.W., Meyer, A., Patel, N.H., Raff, R.A., Roth, V.L. and Wray, G.A.,** 1997. Homology and developmental genes. *Trends in genetics*, *13*(11), pp.432-433.

**Abouheif, E.**, 2008. Parallelism as the pattern and process of mesoevolution. *Evolution and Development*, *10*(1), pp.3-5.

**Akkermans, A.D.L. and Dijk, C.V.,** 1981. Non-leguminous root-nodule symbioses with actinomycetes and Rhizobium.

**Ané, J.M., Kiss, G.B., Riely, B.K., Penmetsa, R.V., Oldroyd, G.E., Ayax, C., Lévy, J., Debellé, F., Baek, J.M., Kalo, P. and Rosenberg, C.,** 2004. Medicago truncatula DMI1 required for bacterial and fungal symbioses in legumes. *Science*, *303*(5662), pp.1364-1367.

**Arrighi, J.F., Barre, A., Amor, B.B., Bersoult, A., Soriano, L.C., Mirabella, R., de Carvalho-Niebel, F., Journet, E.P., Ghérardi, M., Huguet, T. and Geurts, R.,** 2006. The Medicago truncatula lysine motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. *Plant physiology*, *142*(1), pp.265-279.

**Banba, M., Gutjahr, C., Miyao, A., Hirochika, H., Paszkowski, U., Kouchi, H. and Imaizumi-Anraku, H.,** 2008. Divergence of evolutionary ways among common sym genes: CASTOR and CCaMK show functional conservation between two symbiosis systems and constitute the root of a common signaling pathway. *Plant and cell physiology*, *49*(11), pp.1659-1671.

**Baron, C. and Zambryski, P.C.,** 1995. Notes from the underground: highlights from plant-microbe interactions. *Trends in Biotechnology*, *13*(9), pp.356-362.

**Balestrini, R. and Bonfante, P.,** 2005. The interface compartment in arbuscular mycorrhizae: a special type of plant cell wall?. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology*, *139*(1), pp.8-15.

**Battaglia, M., Rípodas, C., Clúa, J., Baudin, M., Aguilar, O.M., Niebel, A., Zanetti, M.E. and Blanco, F.A.,** 2014. A nuclear factor Y interacting protein of the GRAS family is required for nodule organogenesis, infection thread progression, and lateral root growth. *Plant physiology*, *164*(3), pp.1430-1442.

**Bensmihen, S., de Billy, F. and Gough, C.,** 2011. Contribution of NFP LysM domains to the recognition of Nod factors during the Medicago truncatula/Sinorhizobium meliloti symbiosis. *PLoS One*, *6*(11), p.e26114.

**Berg, R. H.** 1999a. Frankia forms infection threads. Can. J. Bot., 77, 1327-1333.

**Berg, R.H.**, 1999b. Cytoplasmic bridge formation in the nodule apex of actinorhizal root nodules. *Canadian Journal of Botany*, *77*(9), pp.1351-1357.

**Berry, A. M., and Sunell, L. A.** 1990. The infection process and nodule development. Pages 61-81 in: The Biology of Frankia and Actinorhizal Plants. C. R. Schwintzer and J. D. Tjepkema, eds. Academic Press, New York.

**Bonfante, P.,** 2001. At the interface between mycorrhizal fungi and plants: the structural organization of cell wall, plasma membrane and cytoskeleton. In *Fungal Associations* (pp. 45-61). Springer Berlin Heidelberg.

**Bonfante, P. and Requena, N.,** 2011. Dating in the dark: how roots respond to fungal signals to establish arbuscular mycorrhizal symbiosis. *Current opinion in plant biology*, *14*(4), pp.451-457.

**Borisov, A.Y., Madsen, L.H., Tsyganov, V.E., Umehara, Y., Voroshilova, V.A., Batagov, A.O., Sandal, N., Mortensen, A., Schauser, L., Ellis, N. and Tikhonovich, I.A.,** 2003. The Sym35 gene required for root nodule development in pea is an ortholog of Nin from Lotus japonicus. *Plant Physiology*, *131*(3), pp.1009-1017.

**Brewin, N.J.,** 2004. Plant cell wall remodelling in the Rhizobium–legume symbiosis. *Critical Reviews in Plant Sciences*, *23*(4), pp.293-316.

**Broghammer, A., Krusell, L., Blaise, M., Sauer, J., Sullivan, J.T., Maolanon, N., Vinther, M., Lorentzen, A., Madsen, E.B., Jensen, K.J. and Roepstorff, P.** 2012. Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding. *Proceedings of the National Academy of Sciences*, *109*(34), pp.13859-13864.

**Calvert, H.E., Chaudhary, A.H. and Lalonde, M.,** 1979. Structure of an unusual root nodule symbiosis in a non-leguminous herbaceous dicotyledon. *Symbiotic Nitrogen fixation in the Management of Temperate Forests. Eds. JC Gordon, CT Wheeler and DA Perry, Corvallis, Oregon, USA*, pp.474-475.

**Calvert, H.E., Pence, M.K., Pierce, M., Malik, N.S. and Bauer, W.D.,** 1984. Anatomical analysis of the development and distribution of Rhizobium infections in soybean roots. *Canadian Journal of Botany*, *62*(11), pp.2375-2384.

**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*(7), p.e11630.

**Capoen, W., Goormachtig, S., De Rycke, R., Schroeyers, K. and Holsters, M.** 2005. SrSymRK, a plant receptor essential for symbiosome formation. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), pp.10369-10374.

**Capoen, W., Sun, J., Wysham, D., Otegui, M.S., Venkateshwaran, M., Hirsch, S., Miwa, H., Downie, J.A., Morris, R.J., Ané, J.M. and Oldroyd, G.E.,** 2011. Nuclear membranes control symbiotic calcium signaling of legumes. *Proceedings of the National Academy of Sciences*, *108*(34), pp.14348-14353.

**Cassab, G. I.** 1998. Plant cell wall proteins. Ann. Rev. Plant Physiol. Plant Mol. Biol. 49: 281–309.

**Catalano, C.M., Czymmek, K.J., Gann, J.G. and Sherrier, D.J.,** 2007. Medicago truncatula syntaxin SYP132 defines the symbiosome membrane and infection droplet membrane in root nodules. *Planta*, *225*(3), pp.541-550.

**Catoira, R., Galera, C., de Billy, F., Penmetsa, R.V., Journet, E.P., Maillet, F., Rosenberg, C., Cook, D., Gough, C. and Dénarié, J.** 2000. Four genes of Medicago truncatula controlling components of a Nod factor transduction pathway. *The Plant Cell*, *12*(9), pp.1647-1665.

**Cerri, M.R., Frances, L., Laloum, T., Auriac, M.C., Niebel, A., Oldroyd, G.E., Barker, D.G., Fournier, J. and de Carvalho-Niebel, F.,** 2012. Medicago truncatula ERN transcription factors: regulatory interplay with NSP1/NSP2 GRAS factors and expression dynamics throughout rhizobial infection. *Plant physiology*, *160*(4), pp.2155-2172.

**Chabaud, M., Venard, C., Defaux‐Petras, A., Bécard, G. and Barker, D.G.,** 2002. Targeted inoculation of Medicago truncatula in vitro root cultures reveals MtENOD11 expression during early stages of infection by arbuscular mycorrhizal fungi. *New Phytologist*, *156*(2), pp.265-273.

**Chabaud, M., Gherbi, H., Pirolles, E., Vaissayre, V., Fournier, J., Moukouanga, D., Franche, C., Bogusz, D., Tisa, L.S., Barker, D.G. and Svistoonoff, S.,** 2016. Chitinase‐resistant hydrophilic symbiotic factors secreted by Frankia activate both Ca2+ spiking and NIN gene expression in the actinorhizal plant Casuarina glauca. *New Phytologist*, *209*(1), pp.86-93.

**Charpentier, M., Sun, J., Martins, T.V., Radhakrishnan, G.V., Findlay, K., Soumpourou, E., Thouin, J., Véry, A.A., Sanders, D., Morris, R.J. and Oldroyd, G.E.,** 2016. Nuclear-localized cyclic nucleotide–gated channels mediate symbiotic calcium oscillations. *Science*, *352*(6289), pp.1102-1105.

**Chen, C., Fan, C., Gao, M. and Zhu, H.,** 2009. Antiquity and function of CASTOR and POLLUX, the twin ion channel-encoding genes key to the evolution of root symbioses in plants. *Plant Physiology*, *149*(1), pp.306-317.

**Clavijo, F., Diedhiou, I., Vaissayre, V., Brottier, L., Acolatse, J., Moukouanga, D., Crabos, A., Auguy, F., Franche, C., Gherbi, H. and Champion, A.,** 2015. The Casuarina NIN gene is transcriptionally activated throughout Frankia root infection as well as in response to bacterial diffusible signals. *New Phytologist*, *208*(3), pp.887-903.

**Combier, J.P., Frugier, F., De Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., Vernié, T., Ott, T., Gamas, P., Crespi, M. and Niebel, A.**, 2006. MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. *Genes & development*, *20*(22), pp.3084-3088.

**Cordeiro, L., Sprent, J.I. and McInroy, S.G.,** 1996. Some developmental and structural aspects of nodules of Lonchocarpus muelhbergian us Hassl. *NATURALIA-SAO PAULO-*, *21*, pp.9-22.

**Dart, P.J.,** 1977. Infection and development of leguminous nodules. *A treatise on dinitrogen fixation*, *3*, pp.367-472.

**De Mita S, Streng A, Bisseling T, Geurts R. 2014.** Evolution of a symbiotic receptor through gene duplications in the legume–rhizobium mutualism. *New Phytologist* **201(3):** 961-972.

**Delaux, P.M., Séjalon-Delmas, N., Bécard, G. and Ané, J.M.,** 2013b. Evolution of the plant–microbe symbiotic 'toolkit'. *Trends in plant science*, *18*(6), pp.298-304.

**Delaux, P.M., Radhakrishnan, G.V., Jayaraman, D., Cheema, J., Malbreil, M., Volkening, J.D., Sekimoto, H., Nishiyama, T., Melkonian, M., Pokorny, L. and Rothfels, C.J.,** 2015. Algal ancestor of land plants was preadapted for symbiosis. *Proceedings of the National Academy of Sciences*, *112*(43), pp.13390-13395.

**Demina IV, Persson T, Santos P, Plaszczyca M, Pawlowski K. 2013.** Comparison of the nodule vs. root transcriptome of the actinorhizal plant *Datisca glomerata*: actinorhizal nodules contain a specific class of defensins. *PloS one* **8(8):** e72442.

**Doyle JJ. 1994.** Phylogeny of the legume family: an approach to understanding the origins of nodulation. *Annual Review of Ecology and Systematics*, 325-349.

**Doyle, J.J., 1998.** Phylogenetic perspectives on nodulation: evolving views of plants and symbiotic bacteria. *Trends in Plant Science*, *3*(12), pp.473-478.

**Doyle JJ. 2011.** Phylogenetic perspectives on the origins of nodulation. *Molecular Plant-Microbe Interactions* **24:** 1289–129.

**Duc G, Trouvelot A, Gianinazzi-Pearson V, Gianinazzi S. 1989.** First report of non-mycorrhizal plant mutants (Myc-) obtained in pea (*Pisum sativum* L.) and fababean (*Vicia faba* L.). *Plant Science* **60:** 215–222.

**Ehrhardt, D., Wais, R., and Long, S.** 1996. Calcium spiking in plant root hairs responding to Rhizobium nodulation signals. Cell 85, 673–681.

**Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kaló, P. and Kiss, G.B.,** 2002. A receptor kinase gene regulating symbiotic nodule development. *Nature*, *417*(6892), p.962.

**Felle, H.H., Kondorosi, É., Kondorosi, Á. and Schultze, M.,** 1999. Elevation of the cytosolic free [Ca2+] is indispensable for the transduction of the Nod factor signal in alfalfa. *Plant Physiology*, *121*(1), pp.273-280.

**Fliegmann, J., Jauneau, A., Pichereaux, C., Rosenberg, C., Gasciolli, V., Timmers, A.C., Burlet‐Schiltz, O., Cullimore, J. and Bono, J.J.,** 2016. LYR3, a high-affinity LCO‐binding protein of Medicago truncatula, interacts with LYK3, a key symbiotic receptor. *FEBS letters*, *590*(10), pp.1477-1487.

**Franche, C., Laplaze, L., Duhoux, E. and Bogusz, D.,** 1998a. Actinorhizal symbioses: recent advances in plant molecular and genetic transformation studies. *Critical Reviews in Plant Sciences*, *17*(1), pp.1-28.

**Franche, C., Diouf, D., Laplaze, L., Auguy, F., Frutz, T., Rio, M., Duhoux, E. and Bogusz, D.,** 1998b. Soybean (lbc3), Parasponia, and Trema hemoglobin gene promoters retain symbiotic and nonsymbiotic specificity in transgenic Casuarinaceae: implications for hemoglobin gene evolution and root nodule symbioses. *Molecular plant-microbe interactions*, *11*(9), pp.887-894.

**Franche, C., Normand, P., Pawlowski, K., Tisa, L.S. and Bogusz, D.,** 2016. An update on research on Frankia and actinorhizal plants on the occasion of the 18th meeting of the Frankia-actinorhizal plants symbiosis. *Symbiosis*, *70*(1-3), pp.1-4.

**Gabaldón, T. , and E. V. Koonin.** 2013 . Functional and evolutionary implications of gene orthology. *Nature Reviews. Genetics* 14 : 360 – 366.

**Gamas, P., Brault, M., Jardinaud, M.F. and Frugier, F.,** 2017. Cytokinins in Symbiotic Nodulation: When, Where, What For? *Trends in plant science*, *22*(9), pp.792-802.

**Garrocho-Villegas, V. and Arredondo-Peter, R.,** 2008. Molecular cloning and characterization of a moss (Ceratodon purpureus) nonsymbiotic hemoglobin provides insight into the early evolution of plant nonsymbiotic hemoglobins. *Molecular biology and evolution*, *25*(7), pp.1482-1487.

**Genre, A., Chabaud, M., Faccio, A., Barker, D.G. and Bonfante, P.,** 2008. Prepenetration apparatus assembly precedes and predicts the colonization patterns of arbuscular mycorrhizal fungi within the root cortex of both Medicago truncatula and Daucus carota. *The Plant Cell*, *20*(5), pp.1407-1420.

**Genre, A. and Russo, G.**, 2016. Does a common pathway transduce symbiotic signals in plant–microbe interactions? *Frontiers in plant science*, *7*.

**Gherbi H, Markmann K, Svistoonoff S, Estevan J, Autran D, Giczey G, Auguy F, *et al.*** 2008. SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and Frankia bacteria. *Proceedings of the National Academy of Science USA* **105:** 4928–4932.

**Gleason, C., Chaudhuri, S., Yang, T., Munoz, A., Poovaiah, B.W. and Oldroyd, G.E.,** 2006. Nodulation independent of rhizobia induced by a calcium-activated kinase lacking autoinhibition. *Nature*, *441*(7097), p.1149.

**Gobbato, E., Marsh, J.F., Vernié, T., Wang, E., Maillet, F., Kim, J., Miller, J.B., Sun, J., Bano, S.A., Ratet, P. and Mysore, K.S.,** 2012. A GRAS-type transcription factor with a specific function in mycorrhizal signaling. *Current Biology*, *22*(23), pp.2236-2241.

**Gomez, S.K., Javot, H., Deewatthanawong, P., Torres-Jerez, I., Tang, Y., Blancaflor, E.B., Udvardi, M.K. and Harrison, M.J.,** 2009. Medicago truncatula and Glomus intraradices gene expression in cortical cells harboring arbuscules in the arbuscular mycorrhizal symbiosis. *BMC Plant Biology*, *9*(1), p.10.

**Gopalasubramaniam, 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: Structure, Function, and Bioinformatics*, *72*(1), pp.252-260.

**Granqvist, E., Sun, J., Op den Camp, R., Pujic, P., Hill, L., Normand, P., Morris, R.J., Downie, J.A., Geurts, R. and Oldroyd, G.E.,** 2015. Bacterial‐induced calcium oscillations are common to nitrogen‐fixing associations of nodulating legumes and non‐legumes. *New Phytologist,* 207(3), pp.551-558.

**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. *The Plant Cell*, *22*(7), pp.2509-2526.

**Gualtieri, G. and Bisseling, T.** 2000. The evolution of nodulation. *Plant Mol. Biol*. 42: 181–194.

**Guan, D., Stacey, N., Liu, C., Wen, J., Mysore, K.S., Torres-Jerez, I., Vernié, T., Tadege, M., Zhou, C., Wang, Z.Y. and Udvardi, M.K.,** 2013. Rhizobial infection is associated with the development of peripheral vasculature in nodules of Medicago truncatula. *Plant physiology*, *162*(1), pp.107-115.

**Guinel, F.C. and Geil, R.D.,** 2002. A model for the development of the rhizobial and arbuscular mycorrhizal symbioses in legumes and its use to understand the roles of ethylene in the establishment of these two symbioses. *Canadian Journal of Botany*, *80*(7), pp.695-720.

**Haney, C.H. and Long, S.R.,** 2010. Plant flotillins are required for infection by nitrogen-fixing bacteria. *Proceedings of the National Academy of Sciences,* 107(1), pp.478-483.

**Harrison, M.J.,** 1997. The arbuscular mycorrhizal symbiosis: an underground association. *Trends in Plant Science*, *2*(2), pp.54-60.

**Hirsch, A.M., Larue, T.A.,** 1997. Is the legume nodule a modified root or stem or an organ sui generis?. *Critical Reviews in Plant Sciences*, *16*(4), pp.361-392.

**Hirsch, S., Kim, J., Muñoz, A., Heckmann, A.B., Downie, J.A. and Oldroyd, G.E.,** 2009. GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in Medicago truncatula. *The Plant Cell*, *21*(2), pp.545-557.

**Hocher V, Alloisio N, Auguy F, Founier P, Doumas P, Pujic P, Gherbi H.** 2011. Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant Physiology* Online publication.

**Horváth, B., Yeun, L.H., Domonkos, Á., Halász, G., Gobbato, E., Ayaydin, F., Miró, K., Hirsch, S., Sun, J., Tadege, M. and Ratet, P.,** 2011. Medicago truncatula IPD3 is a member of the common symbiotic signaling pathway required for rhizobial and mycorrhizal symbioses. *Molecular plant-microbe interactions*, *24*(11), pp.1345-1358.

**Ivanov, S., Fedorova, E.E., Limpens, E., De Mita, S., Genre, A., Bonfante, P. and Bisseling, T.,** 2012. Rhizobium–legume symbiosis shares an exocytotic pathway required for arbuscule formation. *Proceedings of the National Academy of Sciences*, *109*(21), pp.8316-8321.

**Jacob, F.,** 1977. Evolution and tinkering. *Science (New York, NY)*, *196*(4295), pp.1161-1166.

**Jacobsen-Lyon, K., Jensen, E.O., Jørgensen, J.E., Marcker, K.A., Peacock, W.J. and Dennis, E.S.,** 1995. Symbiotic and nonsymbiotic hemoglobin genes of Casuarina glauca. *The Plant Cell*, *7*(2), pp.213-223.

**Journet, E.P., El-Gachtouli, N., Vernoud, V., de Billy, F., Pichon, M., Dedieu, A., Arnould, C., Morandi, D., Barker, D.G. and Gianinazzi-Pearson, V.,** 2001. Medicago truncatula ENOD11: a novel RPRP-encoding early nodulin gene expressed during mycorrhization in arbuscule-containing cells. *Molecular Plant-Microbe Interactions*, *14*(6), pp.737-748.

**Kaku, H., Nishizawa, Y., Ishii-Minami, N., Akimoto-Tomiyama, C., Dohmae, N., Takio, K., Minami, E. and Shibuya, N.,** 2006. Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. *Proceedings of the National Academy of Sciences*, *103*(29), pp.11086-11091.

**Kaló, P., Gleason, C., Edwards, A., Marsh, J., Mitra, R.M., Hirsch, S., Jakab, J., Sims, S., Long, S.R., Rogers, J. and Kiss, G.B.,** 2005. Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science*, *308*(5729), pp.1786-1789.

**Kanamori, N., Madsen, L.H., Radutoiu, S., Frantescu, M., Quistgaard, E.M., Miwa, H., Downie, J.A., James, E.K., Felle, H.H., Haaning, L.L. and Jensen, T.H.,** 2006. A nucleoporin is required for induction of Ca2+ spiking in legume nodule development and essential for rhizobial and fungal symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), pp.359-364.

**Kevei, Z., Lougnon, G., Mergaert, P., Horváth, G.V., Kereszt, A., Jayaraman, D., Zaman, N., Marcel, F., Regulski, K., Kiss, G.B. and Kondorosi, A.,** 2007. 3-Hydroxy-3-methylglutaryl coenzyme A reductase1 interacts with NORK and is crucial for nodulation in Medicago truncatula. *The Plant Cell*, *19*(12), pp.3974-3989.

**Kiss, E., Olah, B., Kalo, P. *et al.*** 2009. LIN, a novel type of U-box/WD40 protein, controls early infection by rhizobia in legumes. *Plant Physiol.*151, 1239–1249.

**Kistner C, Parniske, M. 2002.** Evolution of signal transduction in intracellular symbiosis. *Trends in Plant Science* **7:** 511–518.

**Kistner C, Winzer T, Pitzschke A, Mulder L, Sato S, Kaneko T, Tabata S. 2005.** Seven Lotus japonicus genes required for transcriptional reprogramming of the root during fungal and bacterial symbiosis. *The Plant Cell **17*(8):** 2217-2229.

**Kondrashov, F.A., Rogozin, I.B., Wolf, Y.I. and Koonin, E.V.,** 2002. Selection in the evolution of gene duplications. *Genome biology*, *3*(2), pp.research0008-1.

**Kuppusamy, K.T., Endre, G., Prabhu, R., Penmetsa, R.V., Veereshlingam, H., Cook, D.R., Dickstein, R. and VandenBosch, K.A.,** 2004. LIN, a Medicago truncatula gene required for nodule differentiation and persistence of rhizobial infections. *Plant physiology*, *136*(3), pp.3682-3691.

**Kwon, C., Neu, C., Pajonk, S., Yun, H.S., Lipka, U., Humphry, M., Bau, S., Straus, M., Kwaaitaal, M., Rampelt, H. and El Kasmi, F.,** 2008. Co-option of a default secretory pathway for plant immune responses. *Nature*, *451*(7180), pp.835-840.

**Lalonde, M. and Knowles, R.,** 1975. Ultrastructure, composition, and biogenesis of the encapsulation material surrounding the endophyte in Alnus crispa var. mollis root nodules. *Canadian journal of botany*, *53*(18), pp.1951-1971.

**Laloum, T., De Mita, S., Gamas, P., Baudin, M. and Niebel, A.,** 2013. CCAAT-box binding transcription factors in plants: Y so many? *Trends in plant science*, *18*(3), pp.157-166.

**Lancelle, S.A. and Torrey, J.G.,** 1984. Early development of Rhizobium-induced root nodules of *Parasponia rigida*. I. Infection and early nodule initiation. *Protoplasma*, *123*(1), pp.26-37.

**Lancelle, S.A. and Torrey, J.G.,** 1985. Early development of Rhizobium-induced root nodules of Parasponia rigida. II. Nodule morphogenesis and symbiotic development. *Canadian journal of botany*, *63*(1), pp.25-35.

**Laplaze, L., Duhoux, E., Franche, C., Frutz, T., Svistoonoff, S., Bisseling, T., Bogusz, D. and Pawlowski, K.,** 2000. Casuarina glauca prenodule cells display the same differentiation as the corresponding nodule cells. *Molecular plant-microbe interactions*, *13*(1), pp.107-112.

**Lavin, M., Pennington, R.T., Klitgaard, B.B., Sprent, J.I., de Lima, H.C. and Gasson, P.E.,** 2001. The dalbergioid legumes (Fabaceae): delimitation of a pantropical monophyletic clade. *American Journal of Botany*, *88*(3), pp.503-533.

**Lefebvre, B., Timmers, T., Mbengue, M., Moreau, S., Hervé, C., Tóth, K., Bittencourt-Silvestre, J., Klaus, D., Deslandes, L., Godiard, L. and Murray, J.D.,** 2010. A remorin protein interacts with symbiotic receptors and regulates bacterial infection. *Proceedings of the National Academy of Sciences*, *107*(5), pp.2343-2348.

**Lévy, J., Bres, C., Geurts, R., Chalhoub, B., Kulikova, O., Duc, G., Journet, E.P., Ané, J.M., Lauber, E., Bisseling, T. and Dénarié, J.,** 2004. A putative Ca2+ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science*, *303*(5662), pp.1361-1364.

**Libbenga, K.R. and Bogers, R.J.,** 1974. Development of root-nodule symbioses. Root-nodule morphogenesis. *Biology of Nitrogen Fixation. A. Quispel, ed*.

**Limpens E, Franken C, Smit P, Willemse J, Bisseling T, Geurts R. 2003.** LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. *Science* **302(5645):** 630-633.

**Limpens, E., Mirabella, R., Fedorova, E., Franken, C., Franssen, H., Bisseling, T. and Geurts, R.,** 2005. Formation of organelle-like N2-fixing symbiosomes in legume root nodules is controlled by DMI2. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), pp.10375-10380.

**Limpens, E., Ivanov, S., van Esse, W., Voets, G., Fedorova, E. and Bisseling, T.,** 2009. Medicago N2-fixing symbiosomes acquire the endocytic identity marker Rab7 but delay the acquisition of vacuolar identity. *The Plant Cell*, *21*(9), pp.2811-2828.

**Limpens, E. and Bisseling, T.,** 2014. CYCLOPS: a new vision on rhizobium-induced nodule organogenesis. *Cell host & microbe*, *15*(2), pp.127-129.

**Liu, Q. and Berry, A.M.,** 1991. The infection process and nodule initiation in the Frankia-Ceanothus root nodule symbiosis. *Protoplasma*, *163*(2), pp.82-92.

**Lohmann, G.V., Shimoda, Y., Nielsen, M.W., Jørgensen, F.G., Grossmann, C., Sandal, N., Sørensen, K., Thirup, S., Madsen, L.H., Tabata, S. and Sato, S.,** 2010. Evolution and regulation of the Lotus japonicus LysM receptor gene family. *Molecular plant-microbe interactions*, *23*(4), pp.510-521.

**Madsen, E.B., Madsen, L.H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., Sato, S., Kaneko, T., Tabata, S., Sandal, N. and Stougaard, J.,** 2003. A receptor kinase gene of the LysM type is involved in legumeperception of rhizobial signals. *Nature*, *425*(6958), p.637.

**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. *Nature communications*, *1*, p.10.

**Maillet F, Poinsot V, André O, Puech-Pagès V, Haouy A, Gueunier M, Cromer L *et al.*** 2011. Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature*, **469(7328):** 58-63.

**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 biology*, *6*(3), p.e68.

**Markmann, K. and Parniske, M.,** 2009. Evolution of root endosymbiosis with bacteria: How novel are nodules?. *Trends in plant science*, *14*(2), pp.77-86.

**Marsh, J.F., Rakocevic, A., Mitra, R.M., Brocard, L., Sun, J., Eschstruth, A., Long, S.R., Schultze, M., Ratet, P. and Oldroyd, G.E.,** 2007. Medicago truncatula NIN is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. *Plant Physiology*, *144*(1), pp.324-335.

**Mbengue, M., Camut, S., de Carvalho-Niebel, F., Deslandes, L., Froidure, S., Klaus-Heisen, D., Moreau, S., Rivas, S., Timmers, T., Hervé, C. and Cullimore, J.,** 2010. The Medicago truncatula E3 ubiquitin ligase PUB1 interacts with the LYK3 symbiotic receptor and negatively regulates infection and nodulation. *The Plant Cell*, *22*(10), pp.3474-3488.

**Messinese, E., Mun, J.H., Yeun, L.H., Jayaraman, D., Rougé, P., Barre, A., Lougnon, G., Schornack, S., Bono, J.J., Cook, D.R. and Ané, J.M.,** 2007. A novel nuclear protein interacts with the symbiotic DMI3 calcium-and calmodulin-dependent protein kinase of Medicago truncatula. *Molecular Plant-Microbe Interactions*, *20*(8), pp.912-921.

**Michelmore RW, Meyers BC. 1998.** Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research* **8.11:** 1113-1130.

**Middleton PH, Jakab J, Penmetsa RV, Starker CG, Doll J, Kaló P, Prabhu R. 2007.** An ERF transcription factor in Medicago truncatula that is essential for Nod factor signal transduction. *The Plant Cell* **19(4):** 1221-1234.

**Miller, J.B., Pratap, A., Miyahara, A., Zhou, L., Bornemann, S., Morris, R.J. and Oldroyd, G.E.,** 2013. Calcium/Calmodulin-dependent protein kinase is negatively and positively regulated by calcium, providing a mechanism for

decoding calcium responses during symbiosis signaling. *The Plant Cell*, *25*(12), pp.5053-5066.

**Mitra, R.M., Gleason, C.A., Edwards, A., Hadfield, J., Downie, J.A., Oldroyd, G.E. and Long, S.R.,** 2004. A Ca2+/calmodulin-dependent protein kinase required for symbiotic nodule development: gene identification by transcript-based cloning. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(13), pp.4701-4705.

**Miya, A., Albert, P., Shinya, T., Desaki, Y., Ichimura, K., Shirasu, K., Narusaka, Y., Kawakami, N., Kaku, H. and Shibuya, N.,** 2007. CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. *Proceedings of the National Academy of Sciences*, *104*(49), pp.19613-19618.

**Miyata K, Kozaki T, Kouzai Y, Ozawa K, Ishii K, et al. 2014**. Bifunctional plant receptor, OsCERK1, regulates both chitin-triggered immunity and arbuscular mycorrhizal symbiosis in rice. *Plant Cell Physiol*. 55:1864–72.

**Murray, J.D., Karas, B.J., Sato, S., Tabata, S., Amyot, L. and Szczyglowski, K.,** 2007. A cytokinin perception mutant colonized by Rhizobium in the absence of nodule organogenesis. *Science*, *315*(5808), pp.101-104.

**Murray, J.D., Muni, R.R.D., Torres-Jerez, I., Tang, Y., Allen, S., Andriankaja, M., Li, G., Laxmi, A., Cheng, X., Wen, J. and Vaughan, D.,** 2011. Vapyrin, a gene essential for intracellular progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of Medicago truncatula. *The Plant Journal*, *65*(2), pp.244-252.

**Naisbitt, T., James, E.K. and Sprent, J.I.,** 1992. The evolutionary significance of the legume genus Chamaecrista, as determined by nodule structure. *New phytologist*, *122*(3), pp.487-492.

**Nakagawa, T., Kaku, H., Shimoda, Y., Sugiyama, A., Shimamura, M., Takanashi, K., Yazaki, K., Aoki, T., Shibuya, N. and Kouchi, H.,** 2011. From defense to symbiosis: limited alterations in the kinase domain of LysM receptor-like kinases are crucial for evolution of legume–Rhizobium symbiosis. *The Plant Journal*, *65*(2), pp.169-180.

**Nutman, P.S.,** 1948. Physiological studies on nodule formation: I. The relation between nodulation and lateral root formation in red clover. *Annals of Botany*, *12*(46), pp.81-96.

**Oldroyd, G.E. and Downie, J.A.,** 2008. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu. Rev. Plant Biol.*, *59*, pp.519-546.

**Oldroyd, G.E. and Long, S.R.,** 2003. Identification and characterization of nodulation-signaling pathway 2, a gene of Medicago truncatula involved in Nod factor signaling. *Plant Physiology*, *131*(3), pp.1027-1032.

**Oldroyd GE. 2013.** Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews Microbiology* **11(4):** 252-263.

**Op den Camp R, Streng A, De Mita S, Cao Q, Polone E, Lie W, Ammiraju JSS *et al.* 2011.** LysM-type mycorrhizal receptor recruited for Rhizobium symbiosis in nonlegume Parasponia. *Science* **331:** 909–912.

**Parniske M. 2008.** Arbuscular Mycorrhiza: the Mother of Plant Root Endosymbioses. *Nature Reviews of Microbiology* **6:** 763–775.

**Pawlowski, K. and Bisseling, T.,** 1996. Rhizobial and actinorhizal symbioses: what are the shared features?. *The Plant Cell*, *8*(10), p.1899.

**Pawlowski, K. and Demchenko, K.N.**, 2012. The diversity of actinorhizal symbiosis. *Protoplasma*, *249*(4), pp.967-979.

**Pawlowski K, Sprent JI. 2008.** Comparison between actinorhizal and legume symbiosis. In *Nitrogen-fixing actinorhizal symbioses* (pp. 261-288). Springer Netherlands.

**Pawlowski K, Bogusz D, Ribeiro A, Berry AM. 2011.** Progress on research on actinorhizal plants. *Functional Plant Biology* **38(9):** 633-638.

**Pélissier, H.C., Frerich, A., Desimone, M., Schumacher, K. and Tegeder, M.,** 2004. PvUPS1, an allantoin transporter in nodulated roots of French bean. *Plant physiology*, *134*(2), pp.664-675.

**Poovaiah, B.W., Du, L., Wang, H. and Yang, T.,** 2013. Recent advances in calcium/calmodulin-mediated signaling with an emphasis on plant-microbe interactions. *Plant physiology*, *163*(2), pp.531-542.

**Pumplin, N., Mondo, S.J., Topp, S., Starker, C.G., Gantt, J.S. and Harrison, M.J.,** 2010. Medicago truncatula Vapyrin is a novel protein required for arbuscular mycorrhizal symbiosis. *The Plant Journal*, *61*(3), pp.482-494.

**Racette, S. and Torrey, J.G.,** 1989. Root nodule initiation in Gymnostoma (Casuarinaceae) and Shepherdia (Elaeagnaceae) induced by Frankia strain HFPGpI1. *Canadian Journal of Botany*, *67*(10), pp.2873-2879.

**Radutoiu, S., Madsen, L.H., Madsen, E.B., Felle, H.H., Umehara, Y., Grønlund, M., Sato, S., Nakamura, Y., Tabata, S., Sandal, N. and Stougaard, J.**, 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature*, *425*(6958), p.585.

**Radutoiu S, Madsen LH, Madsen EB, Jurkiewicz A, Fukai E, Quistgaard EMH, Albrektsen AS, *et al.* 2007.** LysM domains mediate lipochitin–oligosaccharide recognition and *Nfr* genes extend the symbiotic host range. *EMBO Journal* **26:** 3923–3935.

**Raffaele, S., Mongrand, S., Gamas, P., Niebel, A. and Ott, T.,** 2007. Genome-wide annotation of remorins, a plant-specific protein family: evolutionary and functional perspectives. *Plant physiology*, *145*(3), pp.593-600.

**Rathbun, E.A., Naldrett, M.J. and Brewin, N.J.,** 2002. Identification of a family of extensin-like glycoproteins in the lumen of Rhizobium-induced infection threads in pea root nodules. *Molecular plant-microbe interactions*, *15*(4), pp.350-359.

**Roberts, M.P., Jafar, S. and Mullin, B.C.,** 1985. Leghemoglobin-like sequences in the DNA of four actinorhizal plants. *Plant molecular biology*, *5*(6), pp.333-337.

**Roth, V.L.,** 1984. On homology. *Biological Journal of the Linnean Society*, *22*(1), pp.13-29.

**Saito, K., Yoshikawa, M., Yano, K., Miwa, H., Uchida, H., Asamizu, E., Sato, S., Tabata, S., Imaizumi-Anraku, H., Umehara, Y. and Kouchi, H.,** 2007. NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in Lotus japonicus. *The Plant Cell*, *19*(2), pp.610-624.

**Schauser, L., Roussis, A., Stiller, J. and Stougaard, J.,** 1999. A plant regulator controlling development of symbiotic root nodules. Nature, 402(6758), p.191.

**Schauser, L., Wieloch, W. and Stougaard, J.,** 2005. Evolution of NIN-like proteins in Arabidopsis, rice, and Lotus japonicus. *Journal of molecular evolution*, *60*(2), pp.229-237.

**Shah, V.K. and Brill, W.J.,** 1977. Isolation of an iron-molybdenum cofactor from nitrogenase. *Proceedings of the National Academy of Sciences*, *74*(8), pp.3249-3253.

**Shimizu, T., Nakano, T., Takamizawa, D., Desaki, Y., Ishii‐Minami, N., Nishizawa, Y., Minami, E., Okada, K., Yamane, H., Kaku, H. and Shibuya, N.,** 2010. Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. *The Plant Journal*, *64*(2), pp.204-214.

**Shimoda, Y., Han, L., Yamazaki, T., Suzuki, R., Hayashi, M. and Imaizumi-Anraku, H.,** 2012. Rhizobial and fungal symbioses show different requirements for calmodulin binding to calcium calmodulin–dependent protein kinase in *Lotus japonicus*. *The Plant Cell*, *24*(1), pp.304-321.

**Shubin, N., Tabin, C. and Carroll, S.**, 1997. Fossils, genes and the evolution of animal limbs. *Nature*, *388*(6643), p.639.

**Shtark, O.Y., Sulima, A.S., Zhernakov, A.I., Kliukova, M.S., Fedorina, J.V., Pinaev, A.G., Kryukov, A.A., Akhtemova, G.A., Tikhonovich, I.A. and Zhukov, V.A.,** 2016. Arbuscular mycorrhiza development in pea (Pisum sativum L.) mutants impaired in five early nodulation genes including putative orthologs of NSP1 and NSP2. *Symbiosis*, *68*(1-3), pp.129-144.

**Sieberer, B.J., Chabaud, M., Fournier, J., Timmers, A.C. and Barker, D.G.,** 2012. A switch in Ca2+ spiking signature is concomitant with endosymbiotic microbe entry into cortical root cells of Medicago truncatula. *The Plant Journal*, *69*(5), pp.822-830.

**Silvester, W.B., Harris, S.L. and Tjepkema, J.D.,** 1990. Oxygen regulation and hemoglobin. In *The biology of Frankia and actinorhizal plants* (pp. 157-176).

**Singh, S., Katzer, K., Lambert, J., Cerri, M. and Parniske, M., 2014.** CYCLOPS, a DNA-binding transcriptional activator, orchestrates symbiotic root nodule development. *Cell Host & Microbe*, *15*(2), pp.139-152.

**Smit, P., Raedts, J., Portyanko, V., Debellé, F., Gough, C., Bisseling, T. and Geurts, R.,** 2005. NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. *Science*, *308*(5729), pp.1789-1791.

**Soyano, T., Kouchi, H., Hirota, A. and Hayashi, M.,** 2013. Nodule inception directly targets NF-Y subunit genes to regulate essential processes of root nodule development in Lotus japonicus. *PLoS Genetics*, *9*(3), p.e1003352.

**Soyano , T. , and M. Hayashi.** 2014. Transcriptional networks leading to symbiotic nodule organogenesis. Current Opinion in Plant Biology 20: 146 – 154.

**Sprent JI.** 2001. Nodulation in legumes. London: Royal Botanic Gardens Kew.

**Sprent, J.I.**, 2007. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. *New Phytologist*, *174*(1), pp.11-25.

**Sprent, J.I. and James, E.K.,** 2007. Legume evolution: where do nodules and mycorrhizas fit in?. *Plant Physiology*, *144*(2), pp.575-581.

**Stracke, S., Kistner, C., Yoshida, S., Mulder, L., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J., Szczyglowski, K. and Parniske, M.,** 2002. A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature*, *417*(6892), p.959.

**Streng, A., op den Camp, R., Bisseling, T. and Geurts, R.,** 2011. Evolutionary origin of rhizobium Nod factor signaling. Plant signaling & behavior, 6(10), pp.1510-1514.

**Sturms, R., Kakar, S., Trent III, J. and Hargrove, M.S.,** 2010. *Trema* and *Parasponia* hemoglobins reveal convergent evolution of oxygen transport in plants. *Biochemistry, 49*(19), pp. 4085-4093

**Svistoonoff, S., Laplaze, L., Auguy, F., Runions, J., Duponnois, R., Haseloff, J., Franche, C. and Bogusz, D.,** 2003. cg12 expression is specifically linked to infection of root hairs and cortical cells during Casuarina glauca and Allocasuarina verticillata actinorhizal nodule development. *Molecular plant-microbe interactions*, *16*(7), pp.600-607.

**Svistoonoff, S., Laplaze, L., Liang, J., Ribeiro, A., Gouveia, M.C., Auguy, F., Fevereiro, P., Franche, C. and Bogusz, D.,** 2004. Infection-related activation of the cg12 promoter is conserved between actinorhizal and legume-rhizobia root nodule symbiosis. *Plant physiology*, *136*(2), pp.3191-3197.

**Svistoonoff, S., Benabdoun, F.M., Nambiar-Veetil, M., Imanishi, L., Vaissayre, V., Cesari, S., Diagne, N., Hocher, V., De Billy, F., Bonneau, J. and Wall, L.,** 2013. The independent acquisition of plant root nitrogen-fixing symbiosis in Fabids recruited the same genetic pathway for nodule organogenesis. *PLoS One*, *8*(5), p.e64515.

**Svistoonoff, S., Hocher, V. and Gherbi, H.,** 2014. Actinorhizal root nodule symbioses: what is signalling telling on the origins of nodulation?. *Current opinion in plant biology*, *20*, pp.11-18.

**Swensen, S.M.,** 1996. The evolution of actinorhizal symbioses: evidence for multiple origins of the symbiotic association. *American Journal of Botany*, pp.1503-1512.

**Szczyglowski, K., Shaw, R.S., Wopereis, J., Copeland, S., Hamburger, D., Kasiborski, B., Dazzo, F.B. and de Bruijn, F.J.,** 1998. Nodule organogenesis and symbiotic mutants of the model legume Lotus japonicus. *Molecular Plant-Microbe Interactions*, *11*(7), pp.684-697.

**Tajima, S., Nomura, M. and Kouchi, H.**, 2004. Ureide biosynthesis in legume nodules. *Frontiers in Bioscience*, *9*, pp.1374-1381.

**Takeda, N., Sato, S., Asamizu, E., Tabata, S. and Parniske, M.,** 2009. Apoplastic plant subtilases support arbuscular mycorrhiza development in Lotus japonicus. *The Plant Journal*, *58*(5), pp.766-777.

**Taté, R., Patriarca, E.J., Riccio, A., Defez, R. and Iaccarino, M.,** 1994. Development of Phaseolus vulgaris root nodules. *MPMI-Molecular Plant Microbe Interactions*, *7*(5), pp.582-589.

**Taylor, A. and Qiu, Y.L.,** 2017. Evolutionary history of subtilases in land plants and their involvement in symbiotic interactions. *Molecular Plant-Microbe Interactions*, *30*(6), pp.489-501.

**Timmers, A.C., Auriac, M.C. and Truchet, G.,** 1999. Refined analysis of early symbiotic steps of the Rhizobium-Medicago interaction in relationship with microtubular cytoskeleton rearrangements. *Development*, *126*(16), pp.3617-3628.

**Tirichine, L., Sandal, N., Madsen, L.H., Radutoiu, S., Albrektsen, A.S., Sato, S., Asamizu, E., Tabata, S. and Stougaard, J.**, 2007. A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science*, *315*(5808), pp.104-107.

**Tjepkema, J.**, 1978. The role of oxygen diffusion from the shoots and nodule roots in nitrogen fixation by root nodules of Myrica gale. *Canadian Journal of Botany*, *56*(11), pp.1365-1371.

**Tjepkema, J.D.** 1983. Hemoglobins in the nitrogen-fixing root nodules of actinorhizal plants. *Canadian Journal of Botany* 61, no. 11: 2924-2929.

**Torrey, J.G.,** 1976. Initiation and development of root nodules of Casuarina (Casuarinaceae). *American journal of botany*, pp.335-344.

**Tóth, K., Stratil, T.F., Madsen, E.B., Ye, J., Popp, C., Antolín-Llovera, M., Grossmann, C., Jensen, O.N., Schüßler, A., Parniske, M. and Ott, T.,** 2012. Functional domain analysis of the remorin protein LjSYMREM1 in Lotus japonicus. *PLoS One*, *7*(1), p.e30817.

**Tromas, A., Parizot, B., Diagne, N., Champion, A., Hocher, V., Cissoko, M., Crabos, A., Prodjinoto, H., Lahouze, B., Bogusz, D. and Laplaze, L., 2012.** Heart of endosymbioses: transcriptomics reveals a conserved genetic program among arbuscular mycorrhizal, actinorhizal and legume-rhizobial symbioses. *PLoS One*, *7*(9), p.e44742.

**Truchet, G. and Roche, P.,** 1991. Sulphated lipo-oligosaccharide signals of Rhizobium meliloti elicit root nodule organogenesis in alfalfa. *Nature*, *351*(6328), p.670.

**van Brussel, A.A., Bakhuizen, R., van Spronsen, P.C., Spaink, H.P., Tak, T., Lugtenberg, B.J. and Kijne, J.W.,** 1992. Induction of pre-infection thread structures in the leguminous host plant by mitogenic lipo-oligosaccharides of Rhizobium. *Science(Washington)*, *257*(5066), pp.70-72.

**Van Valen, L.M.,** 1982. Homology and causes. *Journal of Morphology*, *173*(3), pp.305-312.

**van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., Santuari, L., Cao, Q., Sharma, T., Shen, D. and Roswanjaya, Y.,** 2018. Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proceedings of the National Academy of Sciences*, p.201721395.

**Vázquez-Limón, C., Hoogewijs, D., Vinogradov, S.N. and Arredondo-Peter, R.,** 2012. The evolution of land plant hemoglobins. *Plant science*, *191*, pp.71-81.

**Venkateshwaran, M., Jayaraman, D., Chabaud, M., Genre, A., Balloon, A.J., Maeda, J., Forshey, K., den Os, D., Kwiecien, N.W., Coon, J.J. and Barker, D.G.,** 2015. A role for the mevalonate pathway in early plant symbiotic signaling. *Proceedings of the National Academy of Sciences*, *112*(31), pp.9781-9786.

**Wake, D.B., Wake, M.H. and Specht, C.D.,** 2011. Homoplasy: from detecting pattern to determining process and mechanism of evolution. *science*, *331*(6020), pp.1032-1035.

**Wang B, Yeun LH, Xue JY, Yang L, Ane JM, Qiu YL. 2010.** Presence of three mycorrhizal genes in the common ancestor of land plants suggests a key role of mycorrhizas in the colonization of land by plants. *New Phytologist* **186:** 514–525.

**Wang, C., Zhu, H., Jin, L., Chen, T., Wang, L., Kang, H., Hong, Z. and Zhang, Z.,** 2013. Splice variants of the SIP1 transcripts play a role in nodule organogenesis in Lotus japonicus. *Plant molecular biology*, *82*(1-2), pp.97-111.

**Weerasinghe, R.R., Bird, D.M. and Allen, N.S.,** 2005. Root-knot nematodes and bacterial Nod factors elicit common signal transduction events in Lotus japonicus. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(8), pp.3147-3152.

**Yang, W.C., de Blank, C., Meskiene, I., Hirt, H., Bakker, J., van Kammen, A., Franssen, H. and Bisseling, T.,** 1994. Rhizobium nod factors reactivate the cell cycle during infection and nodule primordium formation, but the cycle is only completed in primordium formation. *The plant cell*, *6*(10), pp.1415-1426.

**Yano, K., Yoshida, S., Müller, J., Singh, S., Banba, M., Vickers, K., Markmann, K., White, C., Schuller, B., Sato, S. and Asamizu, E.,** 2008. CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proceedings of the National Academy of Sciences*, *105*(51), pp.20540-20545.

**Yano, K., Shibata, S., Chen, W.L., Sato, S., Kaneko, T., Jurkiewicz, A., Sandal, N., Banba, M., Imaizumi‐Anraku, H., Kojima, T. and Ohtomo, R.,** 2009. CERBERUS, a novel U‐box protein containing WD‐40 repeats, is required for formation of the infection thread and nodule development in the legume–Rhizobium symbiosis. *The Plant Journal,* 60(1), pp.168-180.

**Yokota, K., Fukai, E., Madsen, L.H., Jurkiewicz, A., Rueda, P., Radutoiu, S., Held, M., Hossain, M.S., Szczyglowski, K., Morieri, G. and Oldroyd, G.E.,** 2009. Rearrangement of actin cytoskeleton mediates invasion of Lotus japonicus roots by Mesorhizobium loti. *The Plant Cell*, *21*(1), pp.267-284.

**Yokota, K. and Hayashi, M.,** 2011. Function and evolution of nodulation genes in legumes. *Cellular and molecular life sciences*, *68*(8), pp.1341-1351.

**Zhang XC, Wu X, Findley S, Wan J, Libault M, Nguyen HT, Cannon SB, *et al*. 2007.** Molecular evolution of lysin motif-type receptor-like kinases in plants. *Plant Physiology* **144:** 623-636.

**Zhang, X.C., Cannon, S.B. and Stacey, G., 2009.** Evolutionary genomics of LysM genes in land plants. *BMC evolutionary biology*, *9*(1), p.183.

**Zhu, H., Riely, B.K., Burns, N.J. and Ané, J.M.,** 2006. Tracing nonlegume orthologs of legume genes required for nodulation and arbuscular mycorrhizal symbioses. *Genetics*, *172*(4), pp.2491-2499.

**Zhu, H., Chen, T., Zhu, M., Fang, Q., Kang, H., Hong, Z. and Zhang, Z.,** 2008. A novel ARID DNA-binding protein interacts with SymRK and is expressed during early nodule development in Lotus japonicus. *Plant Physiology*, *148*(1), pp.337-347.

Table 2.1: Overview of Genes Required for Nodulation. Abbreviations: RNS = Rhizobial Nodulation Symbiosis, ANS = Actinorhizal Nodulation Symbiosis, AM = Arbuscular Mycorrhizae.
Citations: 1. Limpens *et al.,* 2003; 2. Radutoiu *et al.,* 2003; 3. Madsen et al., 2003; 4. Nakagawa *et al.,* 2011; 5. Stracke *et al.,* 2002; 6. Op den Camp *et al.,* 2011; 7. Chen *et al.,* 2009; 8. Kistner *et al.,* 2005; 9. Zhang *et al.,* 2009; 10. Streng *et al.,* 2011; 12. Ané *et al.,* 2004; 13. Capoen *et al.,* 2011; 14. Markmann *et al.,* 2008; 15. Gherbi *et al.,* 2008; 16. Saito *et al.,* 2007; 17. Kanamori *et al., 2006;* 18. Groth *et al.,* 2010; 19. Lévy *et al.,* 2004; 20. Mitra *et al.,* 2004; 21. Svistoonoff *et al.,* 2013; 22. Endre *et al.,* 2002; 23. Messinese *et al.,* 2007; 24. Horvath *et al.,* 2011 25. Pumplin *et al.,* 2010 26. Murray, 2011 27. Venkateshwaran *et al.,* 2015; 28. Marsh *et al.,* 2007; 29. Schauser *et al.,* 1999; 30. Borisov *et al.,* 2003; 31. Clavijo *et al.,* 2015; 32. Hirsch *et al.,* 2009; 33. Smit *et al.,* 2005; 34. Kaló *et al.,* 2005; 35. Maillet *et al.,* 2011; 36. Yano *et al.,* 2008 37. Laloum *et al.,* 2013; 38. Ivanov *et al.,* 2012; 39. Laplaze *et al.,* 2000; 40. Yokota *et al.,* 2009; 41. Wang *et al.,* 2013; 42. Zhu *et al.,* 2008; 43. Taylor & Qiu, 2017; 44. Shtark *et al.,* 2016; 45. Raffaele *et al.,* 2007; 46. Soyano *et al.,* 2013

| **Gene** | **Predicted structure** | **Function** | **Symbioses** | **Species** | **Evolutionary History** |
|---|---|---|---|---|---|
| *LjNFR1/ MtLYK3* | LysM-RK | LCO Signal reception | Legume RNS (1,2) | *L. japonicus* (2)*, M. truncatula* (1) | Paralog of CERK1, recruited from defense to mutualism (9, 4) |
| *LjNFR5/ MtNFP* | LysM-RK | LCO Signal reception | Legume RNS (2,3), Rosales RNS (6) | *L. japonicus* (2, 3), *M. truncatula* (1), *P. andersonii (6)* | Possibly recruited from AM LCO receptor function following duplication (10) |
| *LjSYMRK/ MtDMI2/ GmNORK* | Leucine-rich-repeat receptor-like kinase | LCO Signal transduction | AM (5), Legume RNS (5), Cucurbitales ANS (14), Fagales ANS (15) | *L. japonicus* (5), *M. truncatula* (22), *D. glomerata* (14), *C. glauca* (15) | CSSP (8,5) |
| *MtHMGR1* | Mevalonate synthesis - 3-Hydroxy-3-Methylglutaryl CoA Reductase 1 | LCO Signal transduction | AM (27), Legume RNS (27) | *M. truncatula* (27) | CSSP (27) |
| *LjPOLLUX/ MtDMI1* | Cation channel | Nuclear Calcium Spiking | AM (12), Legume RNS (12) | *L. japonicus* (8), *M.truncatula* (12) | CSSP (8) |
| *LjCASTOR* | Cation channel | Nuclear Calcium Spiking | AM (8), Legume RNS (8) | *L. japonicus* (8) | CSSP (8) |

| | | | | | |
|---|---|---|---|---|---|
| *MtMCA8* | SERCA-type calcium pump | Nuclear Calcium Spiking | AM (13), Legume RNS (13) | *M. truncatula* (13) | CSSP (8) |
| *LjNUP133* | Nuclear pore component | Nuclear Calcium Spiking | AM (17), Legume RNS (17) | *L. japonicus* (17) | CSSP (17) |
| *LjNUP85* | Nuclear pore component | Nuclear Calcium Spiking | AM (16), Legume RNS (16) | *L. japonicus* (16) | CSSP (16) |
| *LjNENA* | Nuclear pore component | Nuclear Calcium Spiking | AM (18), Legume RNS (18) | *L. japonicus* (18) | CSSP (18) |
| *LjCCAMK/ MtDMI3* | Calcium and calmodulin-dependent protein kinase | Nuclear Calcium Spiking Transduction | AM (19), Legume RNS (19), Fagales ANS (21), Rosales RNS (6) | *M. truncatula* (19, 20), *C. glauca* (21), *P. andersonii* (6) | CSSP (19) |
| *LjCYCLOPS/ MtIPD3* | Transcription Factor | Nuclear Calcium Spiking Transduction | AM (24, 36), Legume RNS (23, 24) | *M. truncatula* (23), *L. japonicus* (36) | CSSP (24) |
| *NIN* | "Nodule Inception" Transcription Factor | IT and Nodule Formation | Legume RNS (28, 29, 30), Fagales ANS (31) | *L. japonicus* (29), *M. truncatula* (28), *P. sativum* (30), *C. glauca* (31) | Recruited from other pathway, possibly nitrate signaling (37) |
| *NF-YA,B,C* | CCAAT-box Transcription Factor | Nodule Formation | Legume RNS (46) | *L. japonicus* (46) | Likely recruited from lateral root development (37) |
| SIP1 | ARID Transcription Factor | IT and Nodule Formation | AM (41), Legume RNS (42) | *L. japonicus* (42) | CSSP (41, 42) |
| *NSP1* | GRAS Transcription factor | IT and Nodule Formation | Legume RNS (33), AM (44) | *M. truncatula (33)* | Common to Nodulation and AM (33, 40) |
| *NSP2* | GRAS Transcription factor | IT and Nodule Formation | Legume RNS (34), AM (35) | *M. truncatula (34)* | Common to Nodulation and AM (34, 35, 40) |

| | | | | | |
|---|---|---|---|---|---|
| *VAPYRIN* | Novel protein with major sperm domain and ankyrin repeats | IT and Nodule Formation | AM (25), Legume RNS (26) | *M. truncatula (25,26)* | Common to Nodulation and AM (25,26) |
| *LjSbtM4/ CG12* | Subtilase (Subtilisin-like Serine Protease) | IT and Nodule Formation | AM (8), Legume RNS (8), Fagales ANS (39) | *L. japonicus* (8), *C. glauca* (39) | Legume RNS orthologous with AM, Fagales ANS paralogous (43) |
| *VAMP721e/ VAMP721d* | Vesicle-associated membrane protein | Infection droplet formation | Legume RNS (38), AM (38) | *L. japonicus* (38) | Common to Nodulation and AM (38) |

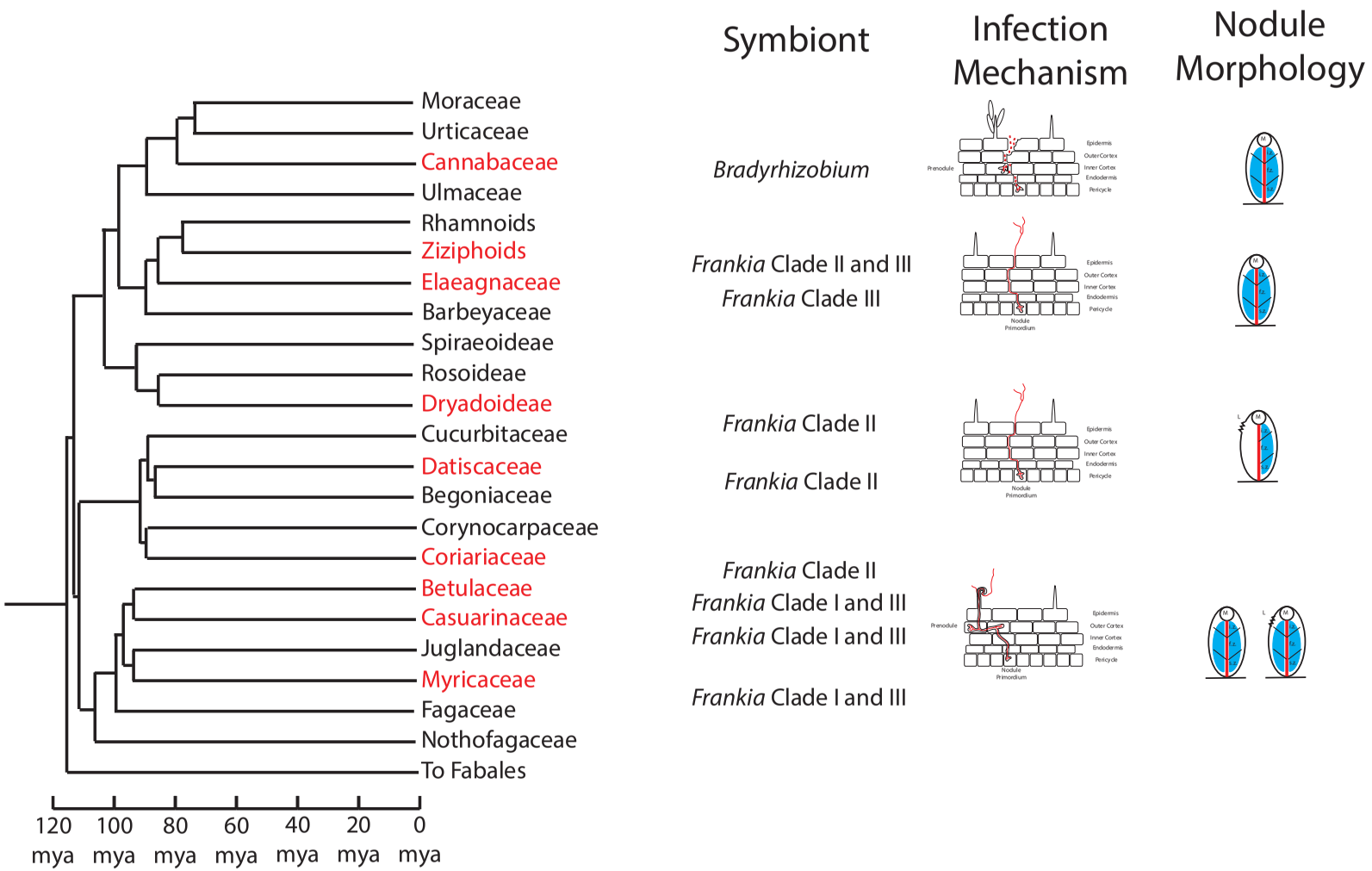



**Figure 2.1:** Distribution of nodulation among Fagales, Cucurbitales and Rosales; lineages with nodulating species in red, lineages with no nodulating species in black. Phylogenetic relationships and dates modified from Li et al. (2015). Betulaceae (includes Ticodendraceae), Cucurbitaceae (Anisophylleaceae), Datiscaceae (Tetrameleaceae). Juglandaceae (Rhoipteleaceae) found to be sister to Myricaceae/Casuarinaceae/Betulaceae in Li et al., 2015 with 62% BS support. Nodule morphology abbreviations: M = meristem, L = lenticel, i.z. = infection zone, f.z. = fixation zone. S.z. = senescence zone Host-Symbiont Specificities from Benson & Clawson (2000). Infection mechanisms and nodule morphologies from Torrey, 1976; Lancelle & Torrey, 1984; Racette & Torrey, 1989; Pawlowski & Sprent, 2008, Pawlowski & Demchenko, 2012.

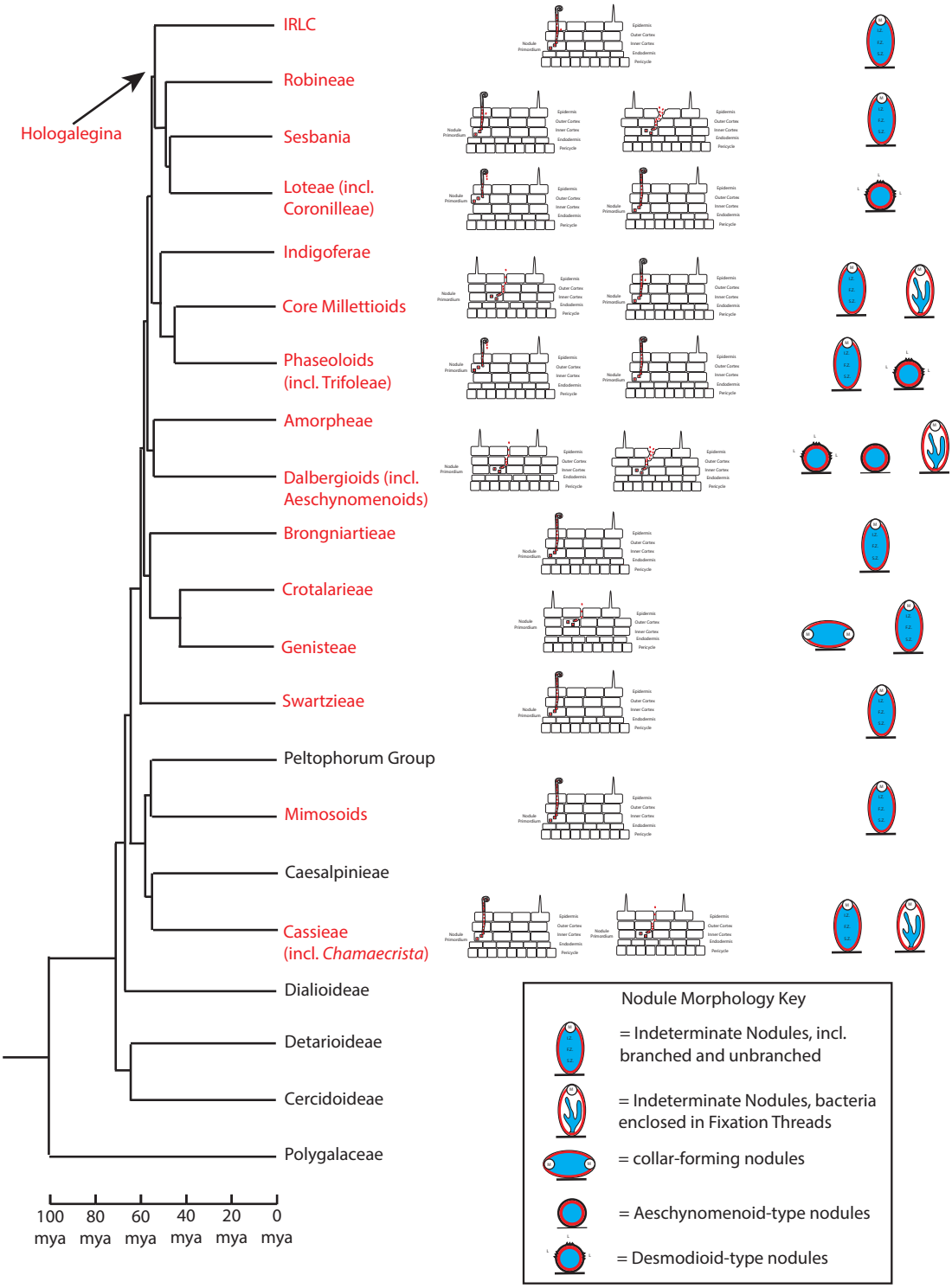



**Figure 2.2:** Distribution of nodulation among Fabales; lineages with nodulating species in red, lineages with no nodulating species in black. Phylogenetic relationships and dates modified from LPWG (2017), nodule morphologies and infection mechanisms from Sprent (2001). Polygalaceae (includes Suriancaceae, Quillajaceae), Peltophorum (Dimorphandra Group A, Tachigali), Phaseoloids (Trifoleae, Demodieae, Psoraleae), Loteae (Coronilleae). Some clades excluded from the figure for clarity. IRLC = Inverted Repeat Loss Clade. Clade names in black do not nodulate, clade names in red do nodulate. Fixation threads in dalbergioids restricted to Andira and Hymenolobium genera, fixation threads in millettioids restricted to Cyclolobium, Dahlstedtia and Poecilanthe genera. Nodule morphology abbreviations: M = meristem, L = lenticel, i.z. = infection zone, f.z. = fixation zone, s.z. = senescence zone

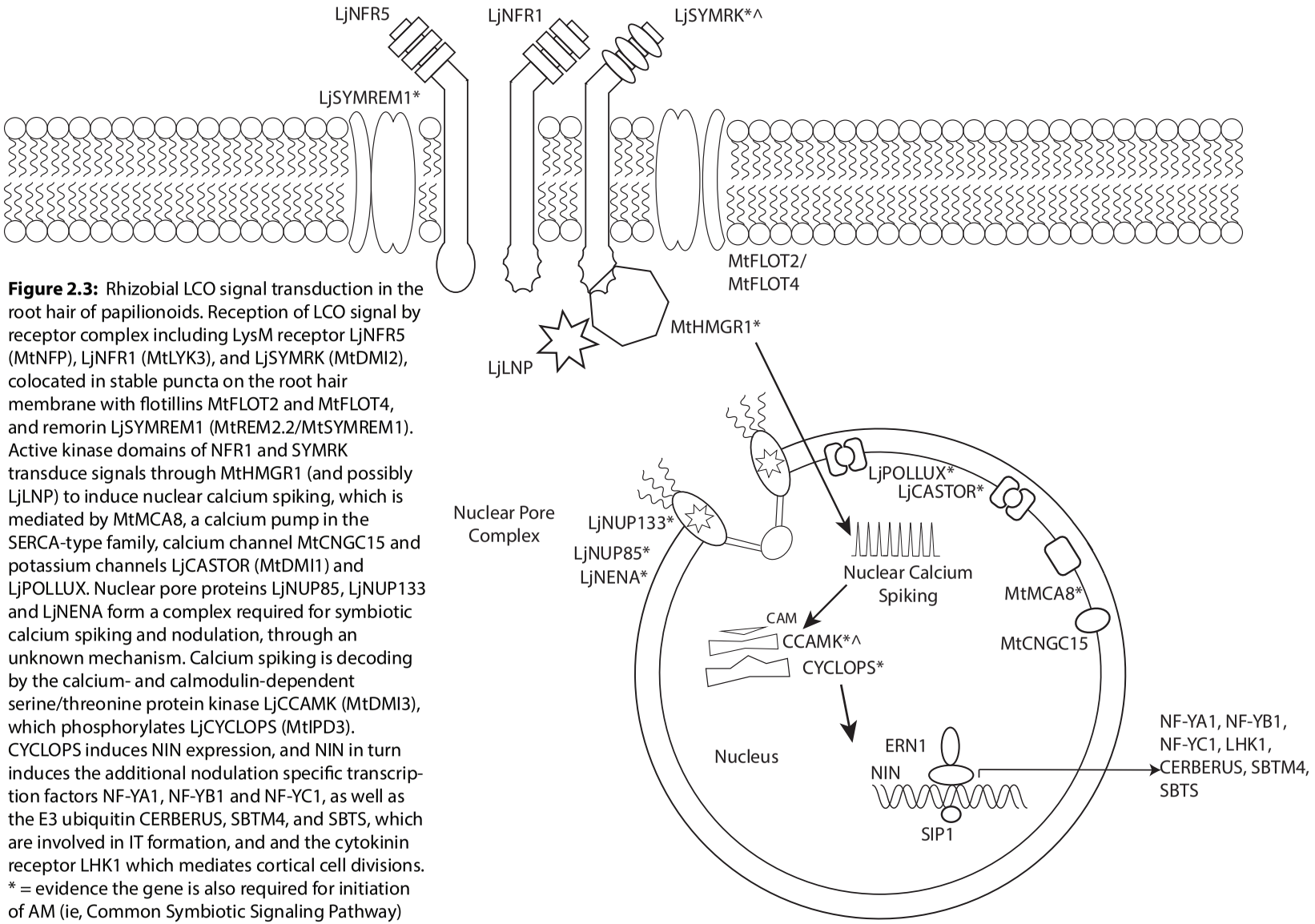



**Figure 2.3:** Rhizobial LCO signal transduction in the root hair of papilionoids. Reception of LCO signal by receptor complex including LysM receptor LjNFR5 (MtNFP), LjNFR1 (MtLYK3), and LjSYMRK (MtDMI2), colocated in stable puncta on the root hair membrane with flotillins MtFLOT2 and MtFLOT4, and remorin LjSYMREM1 (MtREM2.2/MtSYMREM1). Active kinase domains of NFR1 and SYMRK transduce signals through MtHMGR1 (and possibly LjLNP) to induce nuclear calcium spiking, which is mediated by MtMCA8, a calcium pump in the SERCA-type family, calcium channel MtCNGC15 and potassium channels LjCASTOR (MtDMI1) and LjPOLLUX. Nuclear pore proteins LjNUP85, LjNUP133 and LjNENA form a complex required for symbiotic calcium spiking and nodulation, through an unknown mechanism. Calcium spiking is decoding by the calcium- and calmodulin-dependent serine/threonine protein kinase LjCCAMK (MtDMI3), which phosphorylates LjCYCLOPS (MtIPD3). CYCLOPS induces NIN expression, and NIN in turn induces the additional nodulation specific transcription factors NF-YA1, NF-YB1 and NF-YC1, as well as the E3 ubiquitin CERBERUS, SBTM4, and SBTS, which are involved in IT formation, and and the cytokinin receptor LHK1 which mediates cortical cell divisions.
* = evidence the gene is also required for initiation of AM (ie, Common Symbiotic Signaling Pathway)
^ = evidence the gene is also required for initiation of nodulation in actinorhizal lineages (see Table 2.1)

# CHAPTER 3

# Evolutionary History of Subtilases in Land Plants and Their Involvement in Symbiotic Interactions

## 3.1 Abstract and Introduction

Subtilases a family of proteases involved in a variety of developmental processes in land plants, are also involved in both mutualistic symbiosis and host-pathogen interactions in different angiosperm lineages. We examined the evolutionary history of subtilase genes across land plants through a phylogenetic analysis integrating amino acid sequence data from full genomes, transcriptomes, and characterized subtilases of 341 species of diverse green algae and land plants, along with subtilases from 12 species of other eukaryotes, archaea and bacteria. Our analysis reconstructs the subtilase gene phylogeny, and identifies eleven new gene lineages, six of which have no previously characterized members. Two large, previously unnamed subtilase gene lineages that diverged before the origin of angiosperms accounted for the majority of subtilases shown to be associated with symbiotic interactions. These lineages expanded through both whole genome and tandem duplication, with differential neofunctionalization and subfunctionalization creating paralogs associated with different symbioses, including nodulation with nitrogen-fixing bacteria, arbuscular mycorrhizae, and pathogenesis, in different plant clades. This study for the first time demonstrates that a key gene family

involved in plant-microbe interactions proliferated in size and functional diversity before the explosive radiation of angiosperms.

*Introduction*

Subtilisin-like serine proteases (subtilases) of the S8A family are a large group of proteases with broad substrate specificity generally involved in protein turnover and organ development in land plants (Siezen & Leunissen, 1997; Svistoonoff *et al.,* 2003b; Schaller *et al.*, 2012). Subtilases have roles in developmental processes such as lateral root development, epidermal differentiation, cuticle formation, and xylem differentiation (Neuteboom *et al.,* 1999; Tanaka *et al.,* 2001; Zhao *et al.,* 2000). In addition, subtilases have been shown to be involved in a variety of symbiotic interactions, including Arbuscular Mycorrhization (AM), nodulation, and pathogenesis (Table 1).
Despite the wide range of important functions of subtilases in plants, phylogenetic relationships in this gene family have not been updated since a phylogenetic study of subtilases in the single model organism *Arabidopsis thaliana by* Rautengarten *et al.* (2005). The availability of sequence data from a wide variety of fully sequenced genomes and transcriptomes across the tree of life now makes a much more comprehensive phylogenetic reconstruction possible. This not only allows for a more accurate analysis of relationships among gene clades but also permit assessment of evolutionary depth of each clade by comparison with the species tree. Having these two types of information can facilitate evolutionary interpretation of evidence on gene functions obtained from genetic and developmental studies. This study uses a phylogeny of 2,441 subtilase sequences from across 341 green plants, along with 19 subtilase sequences from 12 species of other

eukaryotes, archaea and bacteria, to examine the evolutionary history of this important gene family. Because the genetic basis for the evolutionary origin of nodulation is an important question with implications for agriculture as well as our understanding of convergence in complex symbiotic traits, our interpretation of the phylogeny focuses on the recruitment of subtilases to mediate AM and nodulation, two mutualistic symbiotic interactions plants form with microbes in their roots.

*Subtilases in Root Mutualisms*

Expression of certain subtilases is induced by activation of the "Common Symbiotic" (*sym*) pathway that mediates the formation of two ecologically and economically important mutualisms that plants form with microbes in their roots, AM and nodulation with nitrogen-fixing bacteria (Kistner & Parniske, 2002; Kistner *et al.,* 2005; Parniske, 2008; Gherbi *et al.,* 2008; Hocher *et al.,* 2011; Pawlowski *et al.,* 2011). The *sym* pathway involves the plant perception of lipo-chito-oligosaccharide (LCO) signals convergently produced by rhizobial bacteria and AM fungi, engendering intracellular calcium spiking in the plant root epidermis (Stracke *et al.,* 2002; Maillet *et al.,* 2011). Subsequently, the plant forms a cytoplasmic bridge, called the "infection thread" for nodulation and "pre-penetration apparatus" for AM, that guides the microbial symbiont to the site of symbiosome or arbuscule formation (Szczyglowski *et al.,* 1998; Genre *et al.,* 2005; Kistner *et al*., 2005; Parniske, 2008). Subtilase genes in the *sym* pathway are expressed in infected root hair or epidermal cells forming the infection thread or pre-penetration apparatus, with proteins secreted in the apoplastic space in the symbiotic membrane-membrane interface of the symbiosome or arbsucule, where they

are thought to be involved in restructuring the plant membrane surface (Table 1; Laplaze *et al*., 2000; Svistoonoff *et al.,* 2003; Kistner *et al.,* 2005; Takeda *et al.,* 2009).

Several lines of evidence link subtilases to AM (Table 1). In early expression analyses using cDNA probe arrays, expression of *OsAM21* (renamed *OsSBTM1* by Takeda *et al.,* 2011) in *Oryza sativa* (Güimil *et al*., 2005) and AW584611 (*MtSBTM1*) in *Medicago truncatula* (Liu *et al*., 2003) was found to be induced in root tissues following inoculation with AM fungi (Table 1). Histochemical localization showed that the subtilases LjSBTM1, LjSBTM3, LjSBTM4, and LjSBTS are localized to the apoplastic space in cells forming arbuscules in *Lotus japonicus* following inoculation with the AM fungus *Glomus intraradices* (Takeda *et al.,* 2009). Suppression of *LjSBTM1* and *LjSBTM3* expression by RNAi decreased arbuscule formation in *L. japonicus*, demonstrating that these subtilases play a role in AM (Table 1, Takeda *et al*., 2009).

Spatial expression data in legumes also implicates subtilases in the formation of nodules with rhizobial bacteria (Table 1; Kistner *et al.,* 2005; Takeda *et al.,* 2009). Expression of *LjSBTM4* and *LjSBTS* is induced by inoculation with the rhizobial bacterium *Mesorhizobium loti* (Takeda *et al.,* 2009). Histochemical localization showed that LjSBTS expression is transiently induced in epidermal cells during early rhizobial infection, while LjSBTM4 is induced in both epidermal and cortical cells around the infection thread and nodule primordia, as well as in mature nodules (Takeda *et al*., 2009). A recent RNA-seq study conducted as part of the Expression Atlas project (Kapushesky *et al.,* 2009) showed that exposure to LCO signals from *Sinorhizobium meliloti* induces expression of several subtilases in *M. truncatula* roots (van Zeijl *et al.,* 2015).

Additionally, subtilase expression is induced during nodulation with *Frankia* bacteria in nodulating species in the Fagales and Cucurbitales (Table 1; Ribeiro *et al.,* 1995; Laplaze *et al.,* 2000; Svistoonoff *et al.,* 2003; Pawloswki *et al.,* 2011). Using GUS and GFP reporter gene constructs, Svistoonoff *et al.* (2003) showed that the subtilase gene *CG12* is expressed in root hair cells in *Casuarina glauca,* during infection thread formation and in infected cortical nodule and prenodule cells during early nodulation, but not during the period of nitrogen fixation as predicted by bacterial *nifH* expression (Svistoonoff *et al*., 2003). Expression of a subtilase that is a close homolog to *CG12* is induced in *Datisca glomerata* during the initiation of nodulation with *Frankia* based on EST data (Demina *et al.,* 2013).

Nodulation is restricted to the Nitrogen-Fixing Clade (NFC) of rosids (Fabales, Fagales, Cucurbitales, and Rosales), but has evolved multiple times independently within that clade (Soltis *et al*., 1995; Swensen & Mullin, 1997; Doyle, 2011). The *sym* pathway, which mediates AM across land plants, was recruited during each evolutionary origin of nodulation for which the genetic basis has been examined (Kistner & Parniske, 2002; Kistner *et al.,* 2005; Gherbi *et al.,* 2008; Hocher *et al.,* 2011; Op den Camp *et al*., 2011; Pawlowski *et al.,* 2011; Demina *et al.,* 2013). AM is the ancestral condition in land plants, and elements of the *sym* pathway are functionally conserved from legumes to liverworts (Wang & Qiu, 2006; Wang *et al*., 2010), and perhaps even to charophytes (Delaux *et al.,* 2015). The evolutionary origins of nodulation are thus an example of "deep homology" (Shubin *et al*., 1997; Shubin *et al*., 2009; Doyle, 2011), meaning that phylogenetically distinct nodulation symbioses originated by the repeated, independent recruitment of homologous genes from a plesiomorphic (ancestral) pathway for novel,

homoplastic functions in different lineages. It is currently unknown how subtilases involved in nodulation and AM are related to each other, though one small gene phylogeny of 13 subtilases indicated that those mediating nodulation and AM in different plant lineages are of independent origins, in contrast to the pattern observed for *sym* pathway genes (Takeda *et al.,* 2007).

*Subtilases in Pathogenesis*

In addition to their roles in nodulation and AM, subtilases are expressed during pathogenesis. The *Arabidopsis thaliana* subtilase gene *AtSBT3.3* is involved in immune priming in response to *Pseudomonas syringae* and *Hyaloperonospora arabidopsidis*, as demonstrated by weakened immune responses in *sbt3.3* mutants (Ramirez *et al.,* 2013). *SlSBT3* is expressed in response to herbivory by the insect *Manduca sexta* in *Solanum lycopersicum* (Meyer *et al.,* 2016). Expression of *SlP69* subtilase paralogs in *S. lycopersicum* is induced in leaf and stem tissues in response to a variety of pathogens, including *P. syringae*, *Phytophthora infestans*, and the citrus exocortis viroid (Tornero *et al*., 1996; Tornero *et al*., 1997; Jordá *et al*., 1999; Jordá *et al*., 2000). More recently, as part of the Expression Atlas project (Kapushesky *et al.,* 2009), transcriptome analysis showed differential regulation of subtilases during *P. infestans* infection in tuber tissue of *Solanum tuberosum* (Gao *et al.,* 2013). Strengthening the case for a role in pathogenesis, the subtilase *SlP69B* is induced in *S. lycopersicum* during infection with *P. infestans,* and the pathogen expresses a Kazal-like protease inhibitor that inhibits its activity, indicating pathogen coevolution in response to this subtilase (Tian *et al*., 2004; Tian *et al.,* 2005).

*Study aims*

This study aimed to determine the phylogenetic relationships of subtilases in Viridiplantae, to enhance understanding of the origin and evolution of different lineages in this important gene family (Rautengarten *et al.,* 2005; Schaller *et al.,* 2012). Our analysis and interpretation of this phylogeny focused on subtilase lineages associated with symbiotic interactions in angiosperms. This study also aimed to elucidate the pattern of duplication that led to the multiple paralogous symbiosis-induced subtilases *LjSBTM1, LjSBTM3,* and *LjSBTM4* in *L. japonicus* and *SlP69* paralogs in *S. lycopersicum,* and particularly to determine whether the duplication of symbiosis-induced subtilases occurred via tandem or whole genome duplication (WGD). Tandem duplication and subsequent neofunctionalization is common in genes mediating pathogenesis (Michelmore & Meyers, 1998) while WGD has been implicated in the origin of nodulation in legumes (Vanneste *et al.,* 2014; Cannon *et al.,* 2015).

## 3.2 Results and Discussion

*Relationships Among Plant Subtilases*

Our phylogenetic analysis of 2,460 subtilase amino acid sequences for the first time produces a gene phylogeny that reconstructs evolutionary history of this important gene family across green plants (Fig. 1). We greatly expanded taxonomic sampling compared to the gene phylogeny of Rautengarten *et al.* (2005)„ incorporating subtilase homologs from 341 species of diverse viridiplantae, 7 species of non-viridiplantae eukaryotes, 1 species of archaea and 4 species of bacteria (See Supp. Tables 1, 2 and 4). This expanded taxonomic coverage reveals the approximate age of each gene lineage

through comparison of the gene phylogeny with the corresponding organismal phylogeny (Qiu *et al*., 2006, 2010; Finet *et al*., 2010; Soltis *et al*., 2011). There are two important caveats to this approach. One is the uneven sampling of genomic data – amongst non-angiosperm land plants, this study has only one bryophyte, *Physcomitrella patens* and one lycophyte, *Selaginella moellendorffii,* and lacks any genome from gymnosperms or monilophytes (Supp. Table 2). The second is the fact that transcriptomic data likely contain only a small subset of the homologs in a given species' genome, those that happen to be expressed at the time of tissue sampling (Supp. Table 4).

Despite these caveats, this expanded gene phylogeny allows for several insights into the evolution of subtilases in green plants. First, there is a general pattern of increased gene diversification in angiosperms, resulting in 21 of the 33 named gene lineages restricted to angiosperms (Fig. 1). Additionally, we identified eleven new gene lineages: *SBT2.7*, *SBT3.19, SBT4.16, SBT4.17, SBT4.18, SBT4.20*, *SBT4.21, SBT1.10, SBT1.11*, *SBT1.12* and *SBT1.13*, six of which have no previously characterized members (Fig. 1). Some of these lineages were quite ancient, dating to the origin of seed plants (*SBT3.19*) or vascular plants (*SBT4.16, SBT4.17*), but were missing in the analysis of Rautengarten *et al.* (2005) because *A. thaliana* had no homologs in these lineages. This effect was especially notable in gene lineages containing subtilases mediating symbiotic interactions, such as *SBT4.16, SBT1.10, and SBT1.13,* all of which contain subtilase genes expressed during nodulation and/or AM interactions, and which contain no homologs from *A. thaliana*. This is in keeping a more general pattern of loss of AM-related genes in *A. thaliana*, a non-nodulating and non-mycorrhizal plant (Delaux *et al.,* 2014; Bravo *et al.,* 2016). Finally, while the topology of our gene phylogeny largely

agreed with that of Rautengarten *et al.* (2005), there were major revisions to the gene clades *SBT5*and *SBT6,* (see below).

We found strong bootstrap (BS) support for *SBT1* (96% BS), *SBT2* (100% BS), and *SBT3* (93% BS), three gene clades named by Rautengarten *et al.* (2005) (Fig 3.1). However, *SBT4* was weakly supported (46% BS), though there was strong support for named lineages within this group (Fig. 1). Several clades at this high level would have to be recognized if *SBT4* was more narrowly defined to a clade with at least 70% BS support. While Rautengartern *et al.* (2005) found *SBT5* to be paraphyletic, we found the *AtSBT5* sequences to be polyphyletic, with *AtSBT5.1* and *AtSBT5.2* falling into the weakly supported *SBT4*, while the rest form a monophyletic group (99% BS). Finally, while Rautengarten *et al.* (2005) recognized *SBT6* as a monophyletic group, with two sequences in *A. thaliana*, *AtSBT6.1* and *AtSBT6.2*, our analysis found that these two sequences fell into two large gene clades that together with a small clade formed a paraphyletic group at the base of the entire subtilase tree of Viridiplantae (Fig. 3.1), which are rooted with prokaryotic subtilase sequence (Fig. 3.1).

A series of two consecutive deep divergences of subtilase sequences right above the outgroup prokaryotic sequences, with moderate to strong bootstrap support, indicates that this large gene family underwent two rounds of duplication during early stages of eukaryotic evolution, as the clades that define the two splits contain animal and green plant sequences (Fig. 3.1). The first divergence in eukaryotes is between a small ancient clade (100% BS) containing the subtilase homolog *PC9* from the animal *Mus musculus* and two homologs from the charophyte green alga *Klebsormidium flaccidum*, and all other eukaryotic subtilase clades. The second divergence results in *SBT7* and a un-named

super-clade that includes *SBT1, 2, 3, 4, 5,* and *6* and a small ancient clade containing sequences from animals, fungi, and stramenopiles, in addition to those from green plants. This result is similar to what Rautengarten *et al.* (2005) found, but the improved taxon sampling in this study significantly increased robustness of the conclusion.

*SBT7*, a newly named subfamily in this study, contains *AtSBT6.1*, which previously was placed in *SBT6* by Rautengarten *et al.* (2005). It clearly originated early in eukaryote evolution, before the divergence of Viridiplantae from Metazoa (Adl *et al.,* 2012), as evidenced by presence of the subtilase homolog *SIP* from *Homo sapiens,* and a subtilase homolog from the phaeophyte *Ectocarpus siliculosus* in this clade. The clade containing *AtSBT6.1* also has subtilase homologs from across Viridiplantae, including the chlorophyte alga *Uronema belkae* (Chaetophorales), the prasinophycean green alga *Ostreococcus tauri,* the charophyte algae *Klebsormidium flaccidum, Chlorokybus atmophyticus, Cylindrocystis cushleckae,* and *Coleochaete scutata* (Fig. 3.1; Adl *et al.,* 2012). The clade containing *AtSBT6.2* is restricted to streptophytes, containing a subtilase homolog from the charophyte alga *Klebsormidium flaccidum* as well as land plants.

*AtSBT6.1* and *AtSBT6.2* are characterized by a stronger similarity to the mammalian kexins and pyrolysins than to plant subtilases, whereas all other *Arabidopsis thaliana SBT* subfamilies do not partition with any of the known human prohormone convertases (Fig. 3.1, Rautengarten *et al*., 2005). In a phylogenetic analysis of all members of *A. thaliana SBT1, 2,* and *6* subfamilies with yeast *Kex2p* and the human prohormone convertases (*PCs*, Furin, *SK1*), *AtSBT6.1* and *AtSBT6.2* were embedded among the yeast and human sequences (Rautengarten *et al*., 2005). We found the clades containing *AtSBT6.1* and *AtSBT6.2* to form a grade, with a clade between containing

subtilase homologs from diverse eukaryotes, including animals (*Mus, Homo*), fungi (*Saccharomyces, Kluyveromyces*), diatom (*Fragilariopsis*), prasinophyte (*Nephroselmis),* and a prymnesiophyte (*Emiliana*) (Fig 3.1). Because the divergence of the clades containing *AtSBT6.1* and *AtSBT6.2* predates the Viridiplantae, it is no longer justifiable to place *AtSBT6.1* and *AtSBT6.2* in the same major group as Rautengarten *et al.* (2005) did. The clade containing *AtSBT6.1* is thus re-named *SBT7* in this study, with *SBT6* conserved for the clade containing *AtSBT6.2*. Both *SBT7* and *SBT6* seem to have low copy numbers throughout their phylogenetic distribution ranges, green plants and land plants, respectively (Fig. 3.1, Supp. Table 2), and are well represented in the 1KP dataset, suggesting that they are expressed in the young leaf and shoot tissue typically collected for the 1KP project (Supp. Table 4).

The *SBT2* lineage originated early in the land plants, containing homologs in the liverworts *Pallavicinia lyelli* and *Frullania sp,* as well as the moss *Physcomitrella patens* (Fig 1). The *SBT2* lineage likely underwent one round of duplication early in land plant evolution, as supported by the presence of bryophyte paralogs in the *SBT2.5* (*AtSLP3*) gene lineage and two parts at the base of the *SBT2* gene lineage (Fig. 1). The *SBT2.4* clade contains only angiosperm subtilases, including *AtALE1* (abnormal leaf shape 1), which is involved in cuticle formation and epidermal cell differentiation in embryos and juvenile *A. thaliana* plants (Tanaka *et al*., 2001). *LlLIM9*, which is induced during meiotic prophase in *Lilium longiflorum* (Kobayashi *et al*., 1994) during microspore development (Taylor *et al.,* 1997), is monophyletic with subtilase sequences from several monocots and eudicots*,* suggesting that this gene lineage, now named *SBT2.7,* probably arose during the origin of mesangiosperms, a clade that includes all angiosperms except

two or three species-poor lineages at the base of angiosperm phylogeny, namely, Amborellales, Nymphaeales, and Austrobaileyales (Qiu *et al.* 2010, Soltis *et al.* 2011). Three characterized members of the *SBT2* lineage, *AtSBT2.5* (*AtSLP3*), *LlLIM9,* and *AtSBT2.4* (*AtALE1*) are all involved in tissue differentiation and organogenesis (Table 1), while *StSBT2.2a* and *StSBT2.2b* are found to be down- and up-regulated following infection with *P. infestans* respectively (Table 3.1, Gao *et al.,* 2013). This lineage has retained relatively steady and low copy numbers through land plant evolution (Supp. Table 2).

All other plant subtilases (*SBT1, SBT3, SBT4* and *SBT5*) form a large, poorly supported clade (33% BS). If this lineage is real, it originated before the divergence of Embryophyta and Zygnematophyceae (Adl *et al.,* 2012) as evidenced by the presence of subtilase homologs from three zygnematophycean green algae *Cylindrocystis cushleckae, Spirogyra sp.,* and *Roya obtusa* in a basal position.

The *SBT5* gene clade contains two functionally characterized members. *AtAIR3,* involved in lateral root formation (Neuteboom *et al*., 1999), and *MtSBT5.3*, which is expressed in response to LCO signals from the nodulating bacteria *Sinorhiz*obium meliloti (van Zeijl *et al*., 2015). *AtSBT5.3, AtSBT5.4, AtSBT5.5* and *AtSBT5.6* form a well supported monophyletic group (99% BS) that we call *SBT5*. This clade does not include *AtSBT5.1* and *AtSBT5.2,* which are in the *SBT4.19* clade. *SBT5* contains homologs mosses and liverworts, indicating an origin of this clade near the origin of land plants.

The *SBT3* clade is the only major clade of the subtilase gene family that has members exclusively in seed plants in our dataset, with the most basal taxon represented being the cycad *Encephalartos barteri* in the *SBT3.19* clade (Fig. 3.1). The only

functionally characterized subtilase in this family is *AtSBT3.3*, which is involved in immune priming in *A. thaliana* (Ramirez *et al.,* 2013). In *A. thaliana*, *SBT3* is the largest clade with 18 members (Rautengarten *et al*., 2005), but these genes seem to be derived from recent duplications that occurred within the *Arabidopsis* genus, the Brassicaceae, or the Brassicales, as most of the 18 copies formed a moderately supported clade by themselves, not with any sequences from related species that are in our matrix, e.g., *Gossypium raimondii*, *Populus trichocarpa*, and *Ricinus communis,* which are in different orders (Fig 3.1).

In the poorly-supported "*SBT4"* clade (46% BS), the earliest diverging subtilase lineage, *SBT4.16/SBT4.17,* appears to have originated in tracheophytes, with homologs in *S. moellendorffii, Azolla caroliniana,* and *Isoetes* sp. (Fig. 3.1). Several subtilases from the bryophyte grade are recovered as being nested in *"SBT4*," but with low bootstrap support for specific placement, including homologs from the moss *Physcomitrella patens,* recovered as sister to *SBT4.18/SBT4.19/SBT4.7/SBT4.6/SBT4.3/SBT4.20/SBT4.*21 (19% BS), and a homolog from the hornwort *Nothoceros vincentianus* recovered to be sister to *SBT4.16/SBT4.17,* (63% BS). *SBT4.16* includes *LjSBTS*, which is expressed during nodulation and AM in *L. japonicus* (Takeda *et al.,* 2009), as well as a subtilase in *M. truncatula* that is upregulated in response to LCO signals from *Sinorhizobium meliloti* (van Zeijl *et al.,* 2015), which we named *MtSBTS* to reflect its similarity to to *LjSBTS* in expression profile and phylogenetic position. The *SBT4* clade also contains *AtXSP1* (*AtSBT4.14*), involved in xylem differentiation (Zhao *et al.,* 2000), the subtilase gene encoding Cucumisin, which is involved in fruit ripening in *Cucumis melo* (Yamagata *et al*., 1994), and *GmSLP1* and *GmSLP2*, involved in seed coat development in *G. max*

(Beilinson *et al.,* 2002).

The *SBT1* clade, according to our analysis, is by far the largest of the subtilase subfamilies, and clearly dates back to the beginning of land plant evolution (Supp. Table 2). Two clades containing subtilases associated with symbiosis, *SBT1.10* and *SBT1.13,* account for almost one third of that expansion (Supp. Table 2), which likely happened in the common ancestor of mesangiosperms (Fig. 3.1). Using only homolog counts from whole genomes, there are, on average 21.69 *SBT1* homologs per species across land plants (compared to 11.78 *SBT4* homologs, the next largest clade), of which an average of 9.17 are in either the *SBT1.10* or *SBT1.13* lineages (Supp. Table 2). The majority of previously characterized subtilases involved in symbiotic interactions are in the *SBT1* clade of the S8A subtilases, in the *SBT1.10* or *SBT1.13* clade, which are involved in symbiotic interactions (Table 1). Many of these subtilases exist as paralogous genes clustered in the genome, indicating tandem duplication (see below).

In addition to subtilases mediating biotic interactions, the *SBT1* clade contains several other characterized subtilases, mostly involved in developmental processes (Schaller *et al*., 2012). *SlP69D-F* are expressed during development in various tissues (Table 3.1, Jordá *et al.,* 1999; Jordá *et al.,* 2000). *AtSBT1.5* and *AtSBT1.6* are expressed constitutively in all cell types and are thought to be involved in non-specific protein degradation and turnover (Rautengarten *et al.,* 2005). In the *SBT1.4* lineage, *TaSSP1* is involved in protein degradation in senescing leaves of *Triticum aestivum* (Roberts *et al.,* 2003). In the *SBT1.7* lineage, *AtARA12 (AtSLP1, AtSCS1)* is found in the intercellular spaces in *A. thaliana* stems and degrades proteins with little specificity (Hamilton *et al.,* 2003), while in *G. max* the ortholog of this gene is involved in seed coat development

(Rautengarten *et al.,* 2008). *SBT1.1* processes preproAtPSK4, a precursor for phytosulfokines, involved in mitogenic activity (Srivastava *et al.,* 2008). *AtSBT1.2 (SDD1)* mediates stomatal density and distribution in *A. thaliana* (Berger & Altman, 2000).

*Relationship Among Subtilases Associated with Symbiosis*

The majority of subtilase genes associated with symbiotic interactions fall into two newly identified distinct gene clades, *SBT1.10* (100% BS) and *SBT1.13* (100% BS) (Fig 3.1, Table 1). Both of these clades contain subtilase genes associated with both nodulation and AM, as well as pathogenesis. While both clades have members from across the mesangiosperms, *SBT1.13* is embedded in a large, strongly supported clade with *SBT1.11, SBT1.12,* and *SBT1.9* (100% BS), containing a gene from the basal angiosperm *Nuphar advena* (Fig 3.1), while *SBT1.10* is embedded in a clade with *SBT1.1* and *SBT1.2*, containing a gene from the basal angiosperm *Amborella trichopoda* (Fig. 3.1), suggesting a divergence between *SBT1.10* and *SBT1.13* early in angiosperm evolution. The two major symbiotic lineages are both in the *SBT1* subfamily, and their large size is likely due to many rounds of whole genome and tandem duplication (see below). Neither of these symbiotic gene lineages was present or named in the gene phylogeny of Rautengarten *et al.* (2005) due to their absence in *A. thaliana*, a non-mycorrhizal and non-nodulating species.

The *SBT1.13* clade includes subtilases expressed during pathogenesis in *Solanum*, AM in *O. sativa,* nodulation in *Alnus glutinosa* and *Casuarina glauca*, and subtilases upregulated in response to rhizobial LCO signals in *M. truncatula,* as well as a subtilase

involved in wound response to insect herbivory in *S. lycopersicum* (Fig 3.1, Table 1). The *SBT1.10* clade includes subtilases expressed during pathogenesis in *Solanum*, AM in *L. japonicus,* and nodulation in the legumes *L. japonicus* and *M. truncatula* (Fig 3.2, Table 1). Due to the absence of genomic and transcriptomic sequence data, it is unclear whether nodulating species in the Fagales have subtilase genes in the *SBT1.10* clade, or whether these genes are expressed during nodulation.

With the ages of gene clades assessed from the corresponding plant clades, it is clear that appearance of both major symbiotic gene lineages significantly predates the origin of the NFC and, by extension, nodulation. This line of evidence in turn suggests that the genes recruited for nodulation were involved in different functions before plants acquired nodulating capability. Consistent with this interpretation, both of the major gene lineages containing nodulation-induced subtilases also include genes induced during AM development. *OsSBTM1,* which was weakly supported as being in the *SBT1.13* clade (34% BS), is expressed during AM formation in *O. sativa* when the root is inoculated by *G. intraradices* (Table 1, Güimil *et al*., 2005). In the *SBT1.10* clade, *LjSBTM4* is expressed in both AM and root nodule symbioses, suggesting an expansion of symbiont range to include nodulating bacteria (Kistner *et al.,* 2005; Takeda *et al.*, 2007, 2009, 2011).In both of these gene clades, subtilases expanded their expression to be induced by new symbionts; the function of these subtilases in restructuring plant cell walls may make these gene clades functionally labile in symbiotic interactions (Schaller *et al.,* 2012).

Six subtilase genes associated with symbiosis did not fall into these two clades, or into the *SBT1* clade, and instead they were scattered in four major clades that were in less derived positions than *SBT1* (Fig. 1). *LjSBTS*, expressed during both nodulation and AM

(Table 1, Takeda *et al.,* 2009), and *MtSBTS*, upregulated in response to LCO signals (Table 1, van Zeijl *et al.,* 2015), are members of a newly identified gene lineage, *SBT4.16* (100% BS). This gene clade likely arose during the origin of vascular plants, as evidenced by their presence in the lycophyte *Selaginella moellendorffii* (Fig. 3.1). Two *S. tuberosum* subtilases, *StSBT2.2a* and *StSBT2.2b*, found to be down- and up-regulated following infection with *P. infestans* respectively (Table 3.1, Gao *et al.,* 2013), are in the *SBT2.2* clade. *MtSBT5.3*, a subtilase from *M. truncatula* that is upregulated in response to *Rhizobium* LCO signals (van Zeijl *et al.,* 2015), is in the *SBT5.3* clade. *AtSBT3.3* in *A. thaliana* is involved in immune priming in response to *P. syringae* and *Hyaloperonospora arabidopsidis* (Ramirez *et al.,* 2013).

The above relationships of subtilase genes associated with symbiosis show a clear non-random distribution of these genes in the subtilase gene tree of green plants, despite limited numbers of genes and organisms that have been characterized, since more than half gene clades in the entire tree have no characterized members (Fig. 3.1). Subtilases have evolved roles in symbiotic interactions many times independently, but symbiosis-induced subtilases are concentrated in one major clade, *SBT1*, and even in this clade they are found only in two subclades, *SBT1.10* and *SBT1.13*, both of which extend to the beginning of mesangiosperm evolution, and which diverged from one another even earlier. A series of tandem and whole genome duplication events may help explain how such a distribution pattern arose.

*Tandem and Whole Genome Duplication in the SBT1.10 clade*

New paralogs in the *SBT1.10* gene clade have been acquired independently in

different plant lineages primarily via tandem gene duplication, but also through whole genome duplication (Fig. 3.2, Fig. 3.3). Subsequently, these paralogs have been recruited for new but related symbiotic interaction functions in different lineages, likely through neo- and subfunctionalization (Ohno, 1970; Lynch & Force 2000; He & Zhang 2005), resulting in a gene lineage with paralogs involved in a variety of symbioses, including pathogenesis, AM and nodulation (Fig 1).

*SBT1.10* is represented by multiple, lineage-specific paralogs in the Papilionoideae (*SBTM1* and *SBTM3;* Fig. 2, node B), Rosaceae (Fig. 3.2, node C), and Malphigiales (Fig. 3.2, node D) and Solanaceae (*P69A-F,* Fig. 3.2, node F), whereas copy numbers of subtilase homologs in the *SBTM4* clade have remained relatively low during eudicot evolution (Fig. 3.2, node A, Fig. 3.3). The *SBTM1/M3* and *P69* paralogs arose from *SBTM4* early in eudicot evolution, before the divergence of asterids and rosids, as evidenced by the respective monophyly (Fig 2, nodes E and C) of *SBTM1/M3* and *P69* paralogs, and their synteny (Fig 3.3A). In the Papilionoideae, *SBTM1/SBTM3* duplicated in the ancestor of the subfamily at least once to produce *SBTM1* and *SBTM3,* as in the case of *Phaseolus vulgaris,* and many times in the case of the paralogs of *M. truncatula* (Fig. 3.2, node B).

To assess the mode of these duplications, we performed a synteny analysis of *SBT1.10* paralogs in selected taxa with well-annotated genomes (Fig. 3.3). In the genomes of *Fragaria vesca* (Rosales) and *Populus trichocarpa* (Malphigiales), *SBT1.10* homologs underwent independent tandem duplication, resulting in multiple *SBT1.10* paralogs that are monophyletic in each order (Fig. 3.2, nodes C and D), and tandem with *SBTM4* (Fig. 3.3). Synteny is preserved between the regions containing *SBT1.10* and

*SBTM4* paralogs in *F. vesca* and *P. trichocarpa* (Fig. 3B).

In all sampled species in the Papilionoideae, *SBTM1* and *SBTM3* paralogs are found on a separate chromosome from *SBTM4,* as compared to their tandem arrangement in *Fragaria vesca* and *Populus trichocarpa* (Fig. 3). This evidence supports a scenario in which *SBTM1/M3* arose by tandem duplication from *SBTM4* before the divergence of Rosales and Fabales, and *SBTM4* then ended up on a separate chromosome from *SBTM1* and *SBTM3* during a larger segmental or whole genome duplication and subsequent pseudogenization. This derived condition, shared by all sampled papilionoids, likely occurred during the WGD event near the origin of Papilionoideae (Vanneste *et al.,* 2014; Cannon *et al.,* 2015), as is supported by the synteny of large chromosomal regions containing *SBTM4* with the regions containing *SBTM1* and *SBTM3* in the Papilionoideae (Fig. 3C). Unlike other *sym* pathway genes that were retained and neofunctionalized or subfunctionalized after this whole genome duplication (Vanneste *et al.,* 2014), the ancestral tandem copy of *SBTM1/M3* near *SBTM4* in legumes was likely lost due to functional redundancy, and vice-versa. This is supported by the presence of the pseudogenes ψ*SBTM2* tandem to *SBTM1* and *SBTM3* and ψ*SBTM5* tandem to *SBTM4* in *L. japonicus* (Takeda *et al.,* 2009) (Fig. 3.3).

An exception to both the low copy number of *SBTM4* and the solely tandem duplication of *SBTM1* and *SBTM3* is found in *G. max,* in which tandem *SBTM1* and *SBTM3* paralogs are found in syntenic regions of chromosomes 1 and 11 (Fig. 3.3). Likewise, the four *SBTM4* paralogs are arranged as two tandem copies on two syntenic regions of chromosomes 5 and 17 (Fig. 3.3), suggesting one tandem duplication and then a subsequent duplication during the whole genome duplication in the *Glycine* genus

(Shoemaker *et al.,* 2006; Schmutz *et al.,* 2010). This is in keeping with the high copy number of subtilases in *G. max* generally, which has 104 subtilase paralogs against an average of 64 in the rosids, likely due to recent polyploidization in this genus (Shoemaker *et al.,* 2006; Schmutz *et al.,* 2010, Supp. Table 2). *L. japonicus* has three copies of *SBTM4,* two of which are arranged tandemly and one quite distant on the same chromosome (Fig. 3.3). Because *Lotus* and *Glycine* belong to two separate major clades of the subfamily Papilionoideae (The Legume Phylogeny Working Group, 2013), and species in basal lineages of both clades that have been sequenced, *P. vulgaris* and *M. truncatula*, have a single copy of *SBTM4*, the high copy number of the gene in *G. max* and *L. japonicus* is clearly the result of two recent, independent duplication events.

The retention of multiple *P69* paralogs across lamiids suggests that they have an adaptive function, perhaps coevolving with specific symbionts (Michelmore and Meyers, 1998). This interpretation is supported by studies showing the coevolution of those pathogens with *P69* subtilases (Table 3.1, Tian *et al.,* 2004). Genes mediating pathogenesis and mutualism often show different evolutionary patterns, with those mediating mutualism remaining static over time while those involved in pathogenesis showing accelerated rates of evolution, in an "arms race," though we did not recover that pattern here (Kimbrel *et al.,* 2013; Bravo *et al.,* 2016).

The pattern of widespread tandem duplication and neofunctionalization seen in these subtilases reflects a general pattern of evolution for host genes involved in co-evolution with symbionts, particularly in organisms with innate (rather than adaptive) immune systems and is found in the LRR class of R genes (Michelmore & Meyers, 1998; Jones & Dangl, 2006). This pattern has been proposed to create multiple paralogs that can

each co-evolve with specific symbionts, free of constraints of retaining functionality with ancestral symbionts, in a divergent selection regime (Michelmore & Meyers, 1998). Other paralogs in the nodulation pathway, such as LysM-domain receptor kinases mediating Nod-factor reception in *L. japonicus* (*LjNFR1a, LjNFR1b,* and *LjNFR1c*), arose by tandem duplication (Limpens *et al.,* 2003; De Mita *et al.,* 2014). This legume-specific gene duplication has been proposed to account for derived traits of symbiont specificity in legumes, by allowing these paralogous receptors to expand host range and coevolve with different symbionts without constraint (Radutoiu *et al.,* 2007).

Whether duplications in genes recruited for nodulation occured at a whole genome-scale or local scale determines the extent to which a genetic "predisposition" for nodulation can be claimed, for which derived states (e.g., crack entry vs. root hair deformation), and at which nodes of the plant phylogeny. Some paralogs recruited for nodulation, such as the ERF transcription factors *MtERN1* and *MtERN2* (Middleton *et al.,* 2007), arose through whole genome duplication in the papilionoid legumes (Vanneste *et al.*, 2014). This wholesale duplication of all genes has been suggested as a possible genetic mechanism for an evolutionary "predisposition" for nodulation by supplying selectively unconstrained genetic material for neofunctionalization (Soltis *et al*., 1995; Swensen & Mullin, 1997; Li *et al.,* 2013; Werner *et al.,* 2014), though recent work has cast doubt on this hypothesis (Cannon *et al.,* 2015).

## 3.3 Conclusions

Subtilases play a role in a wide variety of developmental processes in land plants through the processing and degradation of proteins in the apoplastic space between cells

(Schaller *et al.,* 2012). By incorporating genomic and transcriptomic data into a large dataset spanning land plants, our analysis shows that the subtilase gene family underwent multiple rounds of duplication and diversification, resulting in many subtilase clades with different functions. This diversification was particularly prominent in the angiosperms, to which 21 of the 33 named subtilase lineages are restricted (Fig. 1).

The ability of subtilases to restructure cell walls may be adaptive for a variety of symbiotic interactions with nodulating bacteria, AM fungi, and pathogens, as subtilases were recruited to mediate these interactions multiple times across land plant evolution (Fig. 1, Table 1). Here we show that the majority of subtilases shown to mediate symbiotic interactions fall into two gene lineages, the *SBT1.10* lineage and the *SBT1.13* lineage. However, in both of these lineages, homologous subtilases mediate at least three different symbiotic interactions in different plant species.

Further, in the *SBT1.10* clade, patterns of duplication, as well as neo- and subfunctionalization, were specific to each nodulating lineage (Fig. 2, nodes B,C,D). Most genes in the *sym* pathway are single-copy orthologs, but in those represented by more than one copy, the specific pattern of gene duplication and recruitment has been shown to have functional consequences for nodulation in different lineages (Op den Camp *et al.,* 2011). Whole genome duplications in the rosids have been proposed as a mechanism for the genetic predisposition to nodulation in the NFC (Swensen & Mullin, 1998; Li *et al.,* 2013; Vanneste *et al.,* 2014). Tandem duplication and neofunctionalization is a common pattern in genes mediating biotic interactions (Michelmore & Meyers, 1998), and has been proposed to account for synapomorphies (clade specific derived traits) in the evolution of nodulation (Radutoiu *et al.,* 2007; Zhang

*et al.,* 2007; De Mita *et al.,* 2014).

Nodulation evolved via the recruitment of the pre-existing *sym* pathway, which mediates AM formation across land plants (Wang *et al.,* 2010), and which has elements that originated before the divergence of charophytes and land plants (Delaux *et al.,* 2015). Here we show that the evolution of the subtilase gene family, which contributed to the evolutionary origin of nodulation, involved multiple rounds of duplication and changes in symbiont specificity across the gene phylogeny.

## 3.4 Materials and Methods

Full proteome data were retrieved for 23 selected taxa across the land plant phylogeny from Phytozome v10.2 (Goodstein *et al.,* 2012, see Supp. Table 2), and the *L. japonicus* proteome was retrieved from the Kazusa DNA Research Institute (see Supp. Table 2). These 24 full proteome sequences were assembled into a local BLAST+ database (Altschul *et al.* 1990), and an additional BLAST search of the fully sequenced *Klebsormidium flaccidum* genome, to increase taxonomic sampling to 25 fully sequenced genomes across the viridiplantae.

Amino acid sequences of 54 of the 56 subtilases from Rautengarten's (2005) *A. thaliana* subtilase gene phylogeny were retrieved from GenBank (NCBI, Supp. Table 1); two sequences found to be pseudogenes in alignment, *AtSBT3.1* and *AtSBT4.2,* were excluded. Eight of the 54 *A. thaliana* subtilases have been further functionally characterized in other studies (Supp. Table 1). A literature review of plant subtilases was performed to catalogue characterized plant subtilases in species other than *A. thaliana*, of which 23 were used in BLAST searches, for a total of 77 subtilases from the literature to

capture the phylogenetic breadth of the subtilase (Supp. Table 1). Additional expression data on subtilases was retrieved from the Expression Atlas project (Kapushesky *et al.,* 2010). The 77 sequences from the literature were used for a local BLASTP query of the 24 full proteomes with an e-value cutoff of 0, to discover the full complement of subtilase homologs in each proteome. These were supplemented by a BLASTP query of transcriptomes from 316 taxa in the 1KP database (Supp. Table 4, Matasci *et al.,* 2014), for a total of 341 taxa sampled. An additional 19 subtilase sequences from 12 non-Viridiplantae species (1 archaea, 4 bacteria, and 7 eukaryotes) were added in order to determine the phylogenetic depth of ancient subtilase clades *SBT6* and *SBT7* (Supp. Table 1).

After removal of duplicates and sequences under 300 AA, these searches yielded, in total, 2,460 subtilase amino acid sequences: 77 subtilases retrieved from the literature, 1,159 subtilases from 316 taxa in the 1KP database, 19 subtilases from outside Viridiplantae retrieved from NCBI, and 1,205 subtilases retrieved from the 25 selected proteomes downloaded from Phytozome v10.2 and the Kazusa DNA Research Institute. A multiple sequence alignment of amino acid sequences was performed using CLUSTALO (Thompson *et al.,* 1997).

Sequence alignments were uploaded onto the CIPRES science gateway (Miller *et al.,* 2010) and a maximum likelihood phylogeny was constructed using RAxML (Stamatakis, 2006), using a Dayhoff substitution model and 100 bootstrap replicates (Dayhoff *et al.,* 1978). The tree was rooted using a subtilase sequences from *Bacillus amyloliquefaciens* as an outgroup. The Generalized Time Reversible (GTR) substitution model was also tested, and topologies relevant to the conclusions presented here

remained stable through multiple phylogenetic analyses with different substitution models (Lanave *et al.,* 1984). The resulting gene trees were visually inspected, custom python scripts counting taxon names in tip names were used to record and count genes in orthologous gene lineages (orthogroups), in order to describe patterns of specific gene lineage expansions in different land plant clades.

The *A. thaliana* subtilase gene nomenclature proposed by Rautengarten *et al.* (2005) was used as a foundation for our nomenclature system for subtilases in the Viridiplantae. In cases when a large, monophyletic, well-supported (BS value >70%) gene lineage did not have a representative in *A. thaliana,* we assigned names based on the number sequence used in Rautengarten *et al.* (2005) – for example, the highest number in Rautengarten's *AtSBT4* clade was *AtSBT4.15*, so we named the first unnamed lineage in *SBT4* "*SBT4.16."* Some small or paraphyletic groups were left unnamed, for example the small lineage sister to *SBT6*. For names of major clades in green plants, we followed Cavalier-Smith (1981) for Viridiplantae, Lewis and McCourt (2004) for Charophyta, Mishler & Qiu (in press) for Embryophyta, and Cantino et al. (2007) for Tracheophyta, Spermatophyta, Angiospermae, Mesangiospermae, and Eudicotyledoneae.

Synteny analysis of a large clade containing the majority of characterized symbiosis-induced subtilases was conducted to investigate patterns of duplication of these paralogs in a phylogenetic framework. The genomic position and strandedness of sequences in these orthogroups were identified in the most recent genome assemblies available from Phytozome v10.2 or the Kazusa DNA Research Institute. Syntenic relationships of these genes were investigated using comparative genomics (CoGe) tools SynMap (for chromosome-scale synteny) and GEvo (for fine-scale synteny).

**References:**


**Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW *et al*. 2012.** The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology* ***59*****(5):** 429-514.

**Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. 1990.** Basic local alignment search tool. *Journal of molecular biology* **215(3):** 403-410.

**Arnegard ME, Zwickl DJ, Lu Y, Zakon HH. 2010.** Old gene duplication facilitates origin and diversification of an innovative communication system—twice. *Proceedings of the National Academy of Sciences* **107(51):** 22172-22177.

**Beilinson V, Moskalenko OV, Livingstone DS, Reverdatto SV, Jung R, Nielsen NC. 2002.** Two subtilisin-like proteases from soybean. *Physiol Plant* **115:** 585–597.

**Berger D, Altmann T. 2000.** A subtilisin-like serine protease involved in the regulation of stomatal density and distribution in *Arabidopsis thaliana*. *Genes Development* **14:** 1119–1131.

**Bravo A, York T, Pumplin N, Mueller LA, Harrison MJ. 2016.** Genes conserved for arbuscular mycorrhizal symbiosis identified through phylogenomics. *Nature plants* **2:** p.15208.

**Cannon SB, McKain MR, Harkess A, Nelson MN, Dash S, Deyholos MK, Peng Y. 2015.** Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Molecular biology and evolution* **32(1):** 193-210.

**Cantino PD, Doyle JA, Graham SW , Judd WS, Olmstead RG, Soltis DE, Soltis PS, & Donoghue DJ. 2007.** Towards a phylogenetic nomenclature of Tracheophyta. *Taxon* **56:** 822-846.

**Cavalier-Smith T. 1981. Eukaryotic kingdoms: seven or nine?** *Biosystems* **14:** 461-481.

**Christin PA, Boxall SF, Gregory R, Edwards EJ, Hartwell J, Osborne CP. 2013.** Parallel recruitment of multiple genes into C4 photosynthesis. *Genome Biology and Evolution* ***5*****(11):** 2174-2187.

**Conte GL, Arnegard ME, Peichel CL, Schluter D. 2012.** The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society of London B: Biological Sciences* **279(1749):** 5039-5047.

**Dayhoff M, Schwartz R, Orcutt B. 1978.** A model of evolutionary change in protein. *Atlas Protein Seq. Struct* **5:**345-352

**De Mita S, Streng A, Bisseling T, Geurts R. 2014.** Evolution of a symbiotic receptor through gene duplications in the legume–rhizobium mutualism. *New Phytologist* **201(3):** 961-972.

**Delaux PM, Radhakrishnan GV, Jayaraman D, Cheema J, Malbreil M, Volkening JD, Sekimoto H *et al.* 2015.** Algal ancestor of land plants was preadapted for symbiosis. *Proceedings of the National Academy of Sciences* ***112*****(43):**13390-13395.

**Demina IV, Persson T, Santos P, Plaszczyca M, Pawlowski K. 2013.** Comparison of the nodule vs. root transcriptome of the actinorhizal plant *Datisca glomerata*: actinorhizal nodules contain a specific class of defensins. *PloS one* **8(8):** e72442.

**Doyle JJ. 1994.** Phylogeny of the legume family: an approach to understanding the origins of nodulation. *Annual Review of Ecology and Systematics*, 325-349.

**Doyle JJ. 2011.** Phylogenetic perspectives on the origins of nodulation. *Molecular Plant-Microbe Interactions* **24:** 1289–129

**Finet C, Timme RE, Delwiche CF, Marletaz F. 2010.** Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Current Biolology* **20:** 2217-2222.

**Gao L, Tu ZJ, Millett BP, Bradeen JM. 2013.** Insights into organ-specific pathogen defense responses in plants: RNA-seq analysis of potato tuber-*Phytophthora infestans* interactions. *BMC genomics* ***14*****(1):** 340.

**Genre A, Chabaud M, Timmers T, Bonfante P, Barker DG. 2005.** Arbuscular mycorrhizal fungi elicit a novel intracellular apparatus in *Medicago truncatula* root epidermal cells before infection. *The Plant Cell* ***17*****(12):** 3489-3499.

**Gherbi H, Markmann K, Svistoonoff S, Estevan J, Autran D, Giczey G, Auguy F, *et al.* 2008.** SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and *Frankia* bacteria. *Proceedings of the National Academy of Science USA* **105:** 4928–4932.

**Ghorbani S, Hoogewijs K, Pečenková T, Fernandez A, Inzé A, Eeckhout D, Kawa D. 2016.** The SBT6. 1 subtilase processes the GOLVEN1 peptide controlling cell elongation. *Journal of Experimental Botany*, p.erw241.

**Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T *et al.* 2012.** Phytozome: a comparative platform for green plant genomics. *Nucleic acids research* **40(D1):** D1178-D1186.

**Güimil S, Chang HS, Zhu T, Sesma A, Osbourn A, Roux C, Ioannidis V. 2005.** Comparative transcriptomics of rice reveals an ancient pattern of response to microbial colonization. *Proceedings of the National Academy of Sciences* ***102*****(22):** 8066-8070.

**Hamilton J, Simpson D, Hyman S, Ndimba B, Slabas A. 2003.** Ara12 subtilisin-like protease from *Arabidopsis thaliana*: purification, substrate specificity and tissue localization. *Biochem. J*, **370:** 57-67.

**He X, Zhang J. 2005.** Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169:** 1157-1164.

**Hocher V, Auguy F, Argout X, Laplaze L, Franche C, Bogusz D. 2006.** Expressed sequence-tag analysis in *Casuarina glauca* actinorhizal nodule and root. *New Phytologist*, *169*(4), pp.681-688.

**Hocher V, Alloisio N, Auguy F, Founier P, Doumas P, Pujic P, Gherbi H. 2011.** Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant Physiology* Online publication.

**Jones JD, Dangl JL. 2006.** The plant immune system. *Nature* ***444*****(7117):** 323-329.

**Jordá L, Coego A, Conejero V, Vera P. 1999.** A genomic cluster containing four differentially regulated subtilisin-like processing protease genes is in tomato plants. *Journal of Biological Chemistry* **274**: 2360–2365.

**Jordá L, Conejero V, Vera P. 2000.** Characterization of P69E and P69F, two differentially regulated genes encoding new members of the subtilisin-like proteinase family from tomato plants. *Plant physiology* **122(1):** 67-74.

**Kapushesky M, Emam I, Holloway E, Kurnosov P, Zorin A, Malone J, Rustici G, Williams E, Parkinson H, Brazma A. 2010.** Gene expression atlas at the European bioinformatics institute. *Nucleic acids research* ***38*****(suppl 1):** D690-D698.

**Kimbrel JA, Thomas WJ, Jiang Y, Creason AL, Thireault CA, Sachs JL, Chang JH. 2013.** Mutualistic co-evolution of type III effector genes in *Sinorhizobium fredii* and *Bradyrhizobium japonicum. PLoS Pathogens* ***9*****(2):** p.e1003204.

**Kistner C, Parniske, M. 2002.** Evolution of signal transduction in intracellular symbiosis. *Trends in Plant Science* **7:** 511–518.

**Kistner C, Winzer T, Pitzschke A, Mulder L, Sato S, Kaneko T, Tabata S. 2005.** Seven *Lotus japonicus* genes required for transcriptional reprogramming of the root during fungal and bacterial symbiosis. *The Plant Cell* ***17*****(8):** 2217-2229.

**Kobayashi T, Kobayashi E, Sato S, Hotta Y, Miyajima N, Tanaka A, Tabata S. 1994.** Characterization of cDNAs induced in meiotic prophase in lily microsporocytes. *DNA Research* **1(1):** 15-26.

**Lanave C, Preparata G, Saccone C, Serio G. 1984.** A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution* **20:** 86–93.

**Laplaze L, Ribeiro A, Franche C, Duhoux E, Auguy F, Bogusz D, Pawlowski K. 2000.** Characterization of a *Casuarina glauca* nodule-specific subtilisin-like protease gene, a homolog of *Alnus glutinosa* ag12. *Molecular Plant-Microbe Interactions* **13:** 113–117.

**Lewis LA, McCourt RM. 2004.** Green algae and the origin of land plants. *American Journal of Botany* **91:** 1535–1556.

**Li QG, Zhang L, Li C, Dunwell JM, Zhang YM. 2013.** Comparative genomics suggests that an ancestral polyploidy event leads to enhanced root nodule symbiosis in the Papilionoideae. *Molecular Biology and Evolution* **30(12):** 2602-2611.

**Limpens E, Franken C, Smit P, Willemse J, Bisseling T, Geurts R. 2003.** LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. *Science* **302(5645):** 630-633.

**Liu J, Blaylock LA, Endre G, Cho J, Town CD, VandenBosch KA, Harrison MJ. 2003.** Transcript profiling coupled with spatial expression analyses reveals genes involved in distinct developmental stages of an arbuscular mycorrhizal symbiosis. *The Plant Cell* **15(9):** 2106-2123.

**Liu J-X, Srivastava R, Che P, Howell SH. 2007.** Salt stress responses in Arabidopsis utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. *Plant Journal* **51:**897-909.

**Lynch M, Force A. 2000.** The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154:** 459-473.

**Maillet F, Poinsot V, André O, Puech-Pagès V, Haouy A, Gueunier M, Cromer L *et al*. 2011.** Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature*, **469(7328):** 58-63.

**Matasci N, Hung LH, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N *et al.* 2014.** Data access for the 1,000 Plants (1KP) project. *GigaScience* **3(1):** 1-10.

**Meyer M, Huttenlocher F, Cedzich A, Procopio S, Stroeder J, Pau-Roblot C, Lequart-Pillon M. 2016.** The subtilisin-like protease SBT3 contributes to insect resistance in tomato. *Journal of experimental botany*, p.erw220.

**Michelmore RW, Meyers BC. 1998.** Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research* **8.11:** 1113-1130.

**Middleton PH, Jakab J, Penmetsa RV, Starker CG, Doll J, Kaló P, Prabhu R. 2007.** An ERF transcription factor in *Medicago truncatula* that is essential for Nod factor signal transduction. *The Plant Cell* **19(4):** 1221-1234.

**Miller MA, Pfeiffer W, Schwartz T. 2010.** "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1 - 8.

**Mishler BD & Y-L Qiu.** In Press. Embryophyta. In K. de Queiroz, P. D. Cantino, and J. Gauthier (eds.), *Phylonyms: a companion to the PhyloCode*. University of California Press, Berkeley.

**Nakagawa T, Kaku H, Shimoda Y, Sugiyama A, Shimamura M, Takanashi K, Yazaki K, Aoki T, Shibuya N, Kouchi H. 2011.** From defense to symbiosis: limited alterations in the kinase domain of LysM receptor-like kinases are crucial for evolution of legume–rhizobium symbiosis. *Plant Journal,* 65, 169–180.

**Neuteboom LW, Veth-Tello LM, Clijdesdale OR, Hooykaas PJ, van der Zaal BJ. 1999.** A novel subtilisin-like protease gene from *Arabidopsis thaliana* is expressed at sites of lateral root emergence. *DNA Research* **6(1):** 13-19.

**Ohno S. 1970.** Evolution by Gene Duplication. New York: Springer-Verlag

**Oldroyd GE, Murray JD, Poole PS, Downie JA. 2011.** The rules of engagement in the legume-rhizobial symbiosis. *Annual Review of Genetics* **45:** 119–144.

**Oldroyd GE. 2013.** Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews Microbiology* **11(4):** 252-263.

**Op den Camp R, Streng A, De Mita S, Cao Q, Polone E, Lie W, Ammiraju JSS *et al.* 2011.** LysM-type mycorrhizal receptor recruited for *Rhizobium* symbiosis in nonlegume Parasponia. *Science* **331:** 909–912.

**Parniske M. 2008.** Arbuscular Mycorrhiza: the Mother of Plant Root Endosymbioses. *Nature Reviews of Microbiology* **6:** 763–775.

**Pawlowski K, Sprent JI. 2008.** Comparison between actinorhizal and legume symbiosis. In *Nitrogen-fixing actinorhizal symbioses* (pp. 261-288). Springer Netherlands.

**Pawlowski K, Bogusz D, Ribeiro A, Berry AM. 2011.** Progress on research on actinorhizal plants. *Functional Plant Biology* **38(9):** 633-638.

**Qiu Y-L, Li L, Wang B, Xue JY, Hendry TA, Li RQ, Chen Z. *et al.* 2010.** Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution* **48(6):** 391-425.

**Qiu Y-L, Li L, Wang B, Chen Z, Knoop V, Groth-Malonek M, Dombrovska O. *et al.* 2006.** The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences* **103(42):** 15511-15516.

**Radutoiu S, Madsen LH, Madsen EB, Jurkiewicz A, Fukai E, Quistgaard EMH, Albrektsen AS, *et al.* 2007.** LysM domains mediate lipochitin–oligosaccharide recognition and *Nfr* genes extend the symbiotic host range. *EMBO Journal* **26:** 3923–3935.

**Ramírez V, López A, Mauch-Mani B, Gil MJ, Vera P. 2013.** An extracellular subtilase switch for immune priming in *Arabidopsis*. *PLoS Pathog*ens ***9*(6):** p.e1003445.

**Rautengarten C, Steinhauser D, Büssis D, Stintzi A, Schaller A, Kopka J, Altmann T. 2005.** Inferring hypotheses on functional relationships of genes: analysis of the *Arabidopsis thaliana* subtilase gene family. *PLoS Computational Biology* **1:** e40.

**Ribeiro, A, Akkermans ADL, van Kammen A, Bisseling T, Pawlowski K. 1995.** A nodule-specific gene encoding a subtilisin-like protease is expressed in early stages of actinorhizal nodule development. *The Plant Cell Online* **7.6:** 785-794.

**Roberts IN, Murray PF, Caputo CP, Passeron S, Barneix AJ. 2003.** Purification and characterization of a subtilisin-like serine protease induced during the senescence of wheat leaves. *Physiologia Plantarum* **118(4):** 483-490.

**Schaller A, Stintzi A, Graff L. 2012.** Subtilases–versatile tools for protein turnover, plant development, and interactions with the environment. *Physiologia plantarum* **145(1):** 52-66.

**Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL *et al.* 2010.** Genome sequence of the palaeopolyploid soybean. *Nature* ***463*(7278):** 178-183.

**Shoemaker RC, Schlueter J, Doyle JJ. 2006.** Paleopolyploidy and gene duplication in soybean and other legumes. *Current opinion in plant biology* **9(2):** 104-109.

**Shubin N, Tabin C, Carroll S. 1997.** Fossils, genes and the evolution of animal limbs. *Nature* **388(6643):** 639-648.

**Shubin N, Tabin C, Carroll S. 2009.** Deep homology and the origins of evolutionary novelty. *Nature* **457(7231):** 818-823.

**Siezen RJ, Leunissen JA. 1997.** Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Science*, **6(3):** 501-523.

**Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG. 1995.** Chloroplast gene sequence data suggest a single origin of predisposition for symbiotic nitrogen fixation in angiosperms. *Proceedings of the National Academy of Sciences, USA* **92**: 2647-2651.

**Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez *et al.* 2011.** Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* **98(4):** 704-730.

**Srivastava R, Liu JX, Howell SH. 2008.** Proteolytic processing of a precursor protein for a growth-promoting peptide by a subtilisin serine protease in *Arabidopsis*. *Plant Journal* **56:** 219–227.

**Stamatakis, A. 2006.** RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

**Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, Kaneko T, Tabata S, *et al.* 2002.** A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature **417*(6892):** 959-962.

**Svistoonoff S, Laplaze L, Auguy F, Runions J, Duponnois R, Haseloff J, Franche C *et al.* 2003.** cg12 expression is specifically linked to infection of root hairs and cortical cells during Casuarina glauca and Allocasuarina verticillata actinorhizal nodule development. *Molecular Plant-Microbe Interactions* **16(7):** 600-607.

**Svistoonoff S, Laplaze L, Liang J, Ribeiro A, Gouveia MC, Auguy F, Fevereiro P, *et al.* 2004.** Infection-related activation of the *cg12* promoter is conserved between actinorhizal and legume- rhizobia root nodule symbiosis. *Plant Physiology* **136:** 3191–3197.

**Svistoonoff S, Hocher V, Gherbi H. 2014.** Actinorhizal root nodule symbioses: what is signalling telling on the origins of nodulation? *Current opinion in plant biology* **20:** 11-18.

**Swensen SM, Mullin BC. 1997.** The impact of molecular systematics on hypotheses for the evolution of root nodule symbioses and implications for expanding symbioses to new host plant genera. *Plant and soil* **194(1-2):** 185-192.

**Szczyglowski K, Shaw RS, Wopereis J, Copeland S, Hamburger D, Kasiborski B, Dazzo FB. 1998.** Nodule organogenesis and symbiotic mutants of the model legume *Lotus japonicus*. *Molecular Plant-Microbe Interactions* ***11*****(7):** 684-697.

**Takeda N, Kistner C, Kosuta S, Winzer T, Pitzschke A, Groth M, Sato S *et al.* 2007.** Proteases in plant root symbiosis. *Phytochemistry* **68(1):** 111-121.

**Takeda N, Sato S, Asamizu E, Tabata S, Parniske M. 2009.** Apoplastic plant subtilases support arbuscular mycorrhiza development in *Lotus japonicus*. *Plant Journal* **58:** 766–777.

**Takeda N, Haage K, Sato S, Tabata S, Parniske M. 2011.** Activation of a *Lotus japonicus* subtilase gene during arbuscular mycorrhiza is dependent on the common symbiosis genes and two cis-active promoter regions. *Molecular Plant-Microbe Interactions* **24.6:** 662-670.

**Tanaka H, Onouchi H, Kondo M, Hara-Nishimura I, Nishimura M, Machida C, Machida Y. 2001.** A subtilisin-like serine protease is required for epidermal surface formation in *Arabidopsis* embryos and juvenile plants. *Development* **128(23):** 4681-4689.

**Taylor AA, Horsch A, Rzepczyk A, Hasenkampf CA, Riggs CD. 1997.** Maturation and secretion of a serine proteinase is associated with events of late microsporogenesis. *The Plant Journal* **12(6):** 1261-1271.

**The Legume Phylogeny Working Group. 2013.** Legume phylogeny and classification in the 21st century: Progress, prospects and lessons for other species-rich clades. *Taxon* **62:** 217-248.

**Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997.** The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic acids research* **25(24):** 4876-4882.

**Tian M, Huitema E, Da Cunha L, Torto-Alalibo T, Kamoun S. 2004.** A Kazal-like extracellular serine protease inhibitor from *Phytophthora infestans* targets the tomato pathogenesis-related protease P69B. *Journal of Biological Chemistry* **279:** 26370–26377.

**Tian M, Benedetti B, Kamoun S. 2005.** A Second Kazal-like protease inhibitor from *Phytophthora infestans*inhibits and interacts with the apoplastic pathogenesis-related protease P69B of tomato. *Plant Physiology* **138:**1785–1793.

**Tornero P, Conejero V, Vera P. 1996.** Primary structure and expression of a pathogen-induced protease (PR-P69) in tomato plants: similarity of functional domains to subtilisin-like endoproteases. *Proceedings of the National Academy of Sciences* **93(13):** 6332-6337.

**Tornero P, Conejero V, Vera P. 1997.** Identification of a new pathogen-induced member of the subtilisin-like processing protease family from plants. *Journal of Biological Chemistry* **272:** 14412–14419.

**True JR, Carroll SB. 2002.** Gene co-option in physiological and morphological evolution. *Annual Review of Cell and Developmental Biology* **18(1):** 53-80.

**Vanneste K, Maere S, Van de Peer Y. 2014.** Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **369(1648):** 20130353.

**Wang B, Yeun LH, Xue JY, Yang L, Ane JM, Qiu YL. 2010.** Presence of three mycorrhizal genes in the common ancestor of land plants suggests a key role of mycorrhizas in the colonization of land by plants. *New Phytologist* **186:** 514–525.

**Wang B, Qiu YL. 2006.** Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16**: 299–363.

**Werner GD, Cornwell WK, Sprent JI, Kattge J, Kiers ET. 2014.** A single evolutionary innovation drives the deep evolution of symbiotic N2-fixation in angiosperms. *Nature Communications 5*:4087 | DOI: 10.1038/

**Yamagata H, Masuzawa T, Nagaoka Y, Ohnishi T, Iwasaki, T. 1994.** Cucumisin, a serine protease from melon fruits, shares structural homology with subtilisin and is generated from a large precursor. *Journal of Biological Chemistry* **269(52):** 32725-32731.

**van Zeijl A, den Camp RHO, Deinum EE, Charnikhova T, Franssen H, den Camp HJO, Bouwmeester H, Kohlen W, Bisseling T, Geurts R. 2015**. *Rhizobium* lipo-chitooligosaccharide signaling triggers accumulation of cytokinins in *Medicago truncatula* roots. *Molecular plant* **8(8):** pp.1213-1226.

**Zhang XC, Wu X, Findley S, Wan J, Libault M, Nguyen HT, Cannon SB, *et al*. 2007.** Molecular evolution of lysin motif-type receptor-like kinases in plants. *Plant Physiology* **144:** 623-636.

**Zhao C, Johnson BJ, Kositsup B, Beers EP. 2000.** Exploiting secondary growth in Arabidopsis. Construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. *Plant Physiology* **123:** 1185–1196.

Table 3.1: Expression and functional evidence for selected characterized subtilases in angiosperms. Citations refer to references list of Chapter 3.

| **Gene** | **Organism** | **Clade** | **Function** | **Evidence** | **Citation** |
|---|---|---|---|---|---|
| *SBT6.1* | *Arabidopsis thaliana* | *SBT7* | Cell elongation, heat stress response | Genetic suppressor screen | Ghorbani *et al.,* 2016 |
| *AtALE1* | *Arabidopsis thaliana* | *SBT2.4* | Cuticle formation and epidermal differentiation | Gene knockout | Tanaka *et al.*, 2001 |
| *LlLIM9* | *Lilium longiflorum* | *SBT2.7* | Microspore development | Immunocytochemical localization | Riggs & Horsch 1995, Taylor *et al*., 1997 |
| *StSBT2.2a* | *Solanum tuberosum* | *SBT2.2* | Pathogenesis - *Phytophthora infestans* | Expression | Gao *et al*., 2013 |
| *StSBT2.2b* | *Solanum tuberosum* | *SBT2.2* | Pathogenesis - *Phytophthora infestans* | Expression | Gao *et al*., 2013 |
| *AtAIR3* | *Arabidopsis thaliana* | *SBT5.3* | Lateral root development, loosening cell walls | Spatial expression of GUS reporter gene fusion | Neuteboom *et al*., 1999 |
| *MtSBT5.3* | *Medicago truncatula* | *SBT5.3* | Expressed in response to LCO signaling molecules from *Sinorhizobium meliloti* | Expression | van Zeijl *et al.,* 2015 |
| *AtSBT3.3* | *Arabidopsis thaliana* | *SBT3.3* | Immune priming in response to *Pseudomonas syringae* and *Hyaloperonospora arabidopsidis* | Expression, gene knockout, chromatin immunoprecipitation | Ramirez *et al.,* 2013 |
| *MtSBTS* | *Medicago truncatula* | *SBT4.16* | Expressed in response to LCO signaling molecules from *Sinorhizobium meliloti* | Expression | van Zeijl *et al.,* 2015 |

| | | | | | |
|---|---|---|---|---|---|
| *LjSBTS* | *Lotus japonicus* | *SBT4.16* | Arbuscular Mycorrhizae - *Glomus intraradices* and Nodulation - *Mesorhizobium loti* | Expression, expression reduced in *sym* pathway mutants | Kistner *et al*., 2005; Takeda *et al*., 2009 |
| *GmSLP1, SSTP2* | *Glycine max* | *SBT4.19* | Seed coat development | Immunocytochemical localization | Beilinson *et al.*, 2002, Rautengarten *et al.*, 2008 |
| *GmSLP2, SSTP1* | *Glycine max* | *SBT4.19* | Cotyledon development | Immunocytochemical localization | Beilinson *et al.*, 2002, Rautengarten *et al.*, 2008 |
| *AtXSP1* | *Arabidopsis thaliana* | *SBT4.6* | Xylem differentiation | Expression | Zhao *et al.,* 2000 |
| Cucumisin | *Cucumis melo* | *SBT4.21* | General protein degradation during fruit maturation | Immunocytochemical localization | Yamagata *et al.*, 1994 |
| *AtSDD1* | *Arabidopsis thaliana* | *SBT1.2* | Regulates stomatal density | Gene knockout, RNA-blot localization | Berger &Altmann 2000 |
| *SlP69A* | *Solanum lycopersicum* | *SBT1.10* | Pathogenesis - Citrus exocortis viroid | Expression | Tornero *et al.*, 1996 |
| *SlP69B* | *Solanum lycopersicum* | *SBT1.10* | Pathogenesis - *Pseudomonas syringae, Phytophthora infestans*, citrus exocortis viroid | Expression, Coevolved pathogen inhibitor | Tornero *et al*., 1997; Jordá *et al*., 1999; Tian *et al*., 2004 |
| *SlP69C* | *Solanum lycopersicum* | *SBT1.10* | Pathogenesis - *Pseudomonas syringae* | Expression | Jordá *et al.*, 1999 |
| *SlP69D* | *Solanum lycopersicum* | *SBT1.10* | Expressed in young leaves | Expression | Jordá *et al.*, 1999 |
| *SlP69E* | *Solanum lycopersicum* | *SBT1.10* | Expressed constitutively in roots | Expression | Jordá *et al*., 2000 |

| | | | | | |
|---|---|---|---|---|---|
| *SlP69F* | *Solanum lycopersicum* | *SBT1.10* | Expressed in hydathodes | Expression | Jordá *et al*., 2000 |
| *LjSBTM1* | *Lotus japonicus* | *SBT1.10* | Arbuscular Mycorrhizae - *Glomus intraradices* | Spatial expression of GUS reporter gene fusion, RNAi inhibition reduces arbuscule formation | Kistner *et al*., 2005; Takeda *et al*., 2009; van Zeijl et al., 2015 |
| *LjSBTM3* | *Lotus japonicus* | *SBT1.10* | Arbuscular Mycorrhizae - *Glomus intraradices* | Expression, RNAi inhibition reduces arbuscule formation | Takeda *et al*., 2009 |
| *LjSBTM4* | *Lotus japonicus* | *SBT1.10* | Arbuscular Mycorrhizae - *Glomus intraradices* and Nodulation - *Mesorhizobium loti* | Spatial expression of GUS reporter gene fusion | Takeda *et al*., 2009; van Zeijl *et al*., 2015 |
| *MtSBTM1* | *Medicago truncatula* | *SBT1.10* | Arbuscular Mycorrhizae - *Glomus versiforme* | Expression | Liu *et al*., 2003 |
| *OsSBTM1* | *Oryza sativa* | *SBT1.13* | Arbuscular Mycorrhizae - *Glomus intraradices* | Expression | Güimil *et al*., 2005 |
| *AG12* | *Alnus glutinosa* | *SBT1.13* | Nodulation - *Frankia sp.* | Expression | Ribeiro *et al*., 1995 |
| *CG12* | *Casuarina glauca* | *SBT1.13* | Nodulation with *Frankia sp*., but not mycorrhization with ECM *Pisolithus alba* or AM *Glomus intraradices* | Spatial expression of GFP reporter gene fusion | Laplaze *et al*., 2000; Svistoonoff *et al*., 2003; Hocher *et al*., 2006 |
| *StCG12a* | *Solanum tuberosum* | *SBT1.13* | Pathogenesis - *Phytophthora infestans* | Expression | Gao *et al*., 2013 |
| *StCG12b* | *Solanum tuberosum* | *SBT1.13* | Pathogenesis - *Phytophthora infestans* | Expression | Gao *et al*., 2013 |

| | | | | | |
|---|---|---|---|---|---|
| *SlSBT3* | *Solanum lycopersicum* | *SBT1.13* | Response to herbivory by *Manduca sexta* | Expression, gene knockout | Meyer *et al.,* 2016 |
| *MtCG12a* | *Medicago truncatula* | *SBT1.13* | Expressed in response to LCO signaling molecules from *Sinorhizobium meliloti* | Expression | van Zeijl *et al.,* 2015 |
| *MtCG12b* | *Medicago truncatula* | *SBT1.13* | Expressed in response to LCO signaling molecules from *Sinorhizobium meliloti* | Expression | van Zeijl *et al.,* 2015 |
| *AtSLP2* | *Arabidopsis thaliana* | *SBT1.6* | General protein turnover and metabolism | Expression | Golldack *et al.*, 2003 |
| *TaSSP1* | *Triticum aestivum* | *SBT1.4* | Peptide degradation in senescing leaves | Expression, Proteolytic assay | Roberts *et al*., 2003 |
| *AtARA12, AtSLP1, AtSCS1* | *Arabidopsis thaliana* | *SBT1.7* | General protein degradation in intercellular space of stem, Expression in leaves in young plants and leaves, roots, and stems in older plants | Expression | Hamilton *et al.,* 2003; Golldack-Brockhausen *et al.*, 2003 |

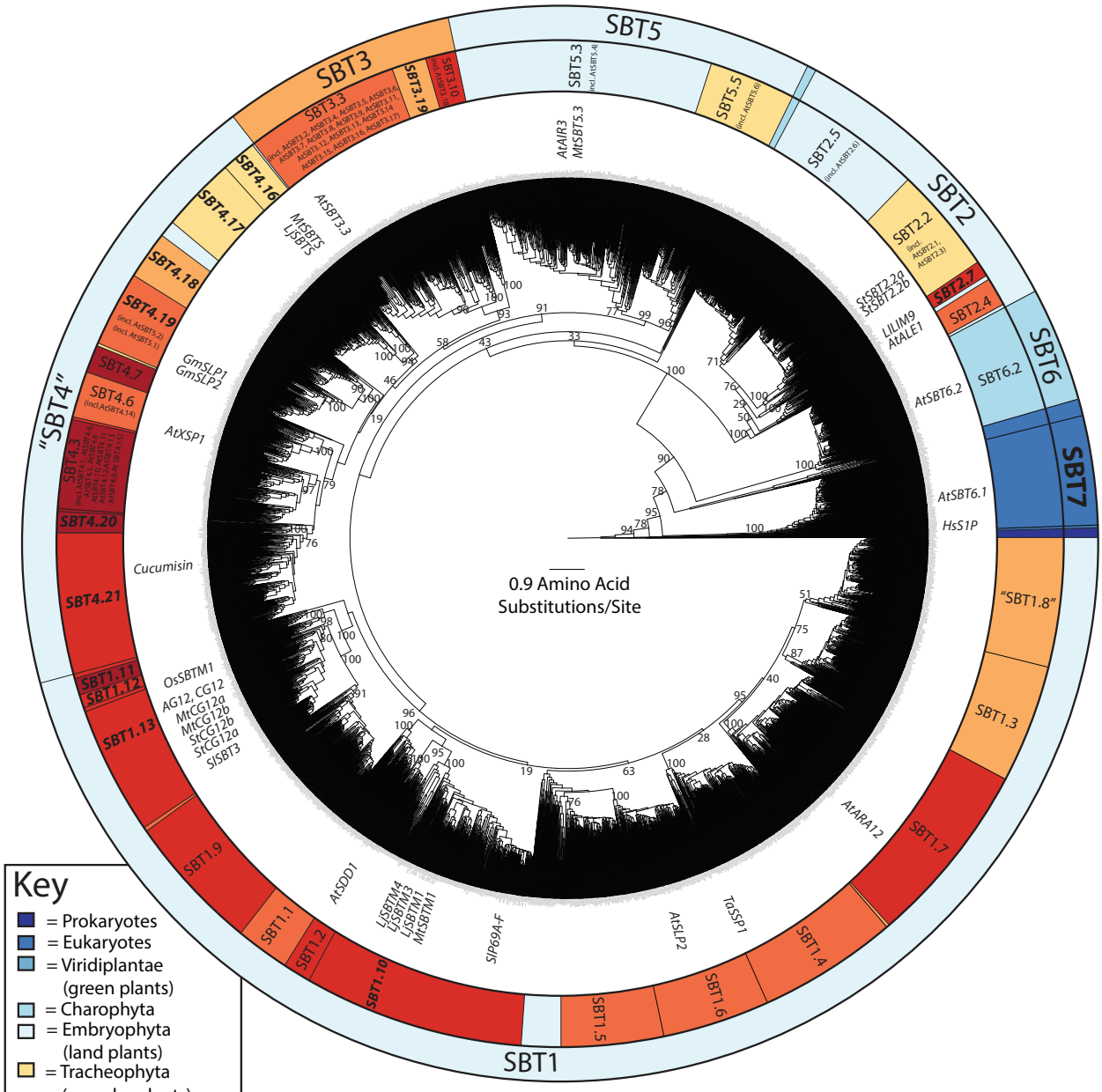



Figure 3.1: Maximum-likelihood phylogenetic tree of 2,460 subtilases. including 2,441 subtilases in the Viridiplantae. Subtilases of interest are noted around the tree, with corresponding functional or expression information in Table 1. Rings around the phylogeny delineate subtilase gene subfamilies and gene lineages, following the nomenclature of Rautengarten et al. (2005) where possible (see Methods for nomenclature guidelines). Gene lineage names introduced in this paper are in bold italic font, names in quotation marks indicate less than 70% BS support, and gene lineages containing many recently duplicated named A. thaliana subtilase paralogs include those paralogs in parentheses. Phylogenetic depth of gene lineages indicated by color in key; criterion is the earliest-diverging plant lineage for which a subtilase homolog is present in that gene lineage. Magnify ~2000X for details of the tree

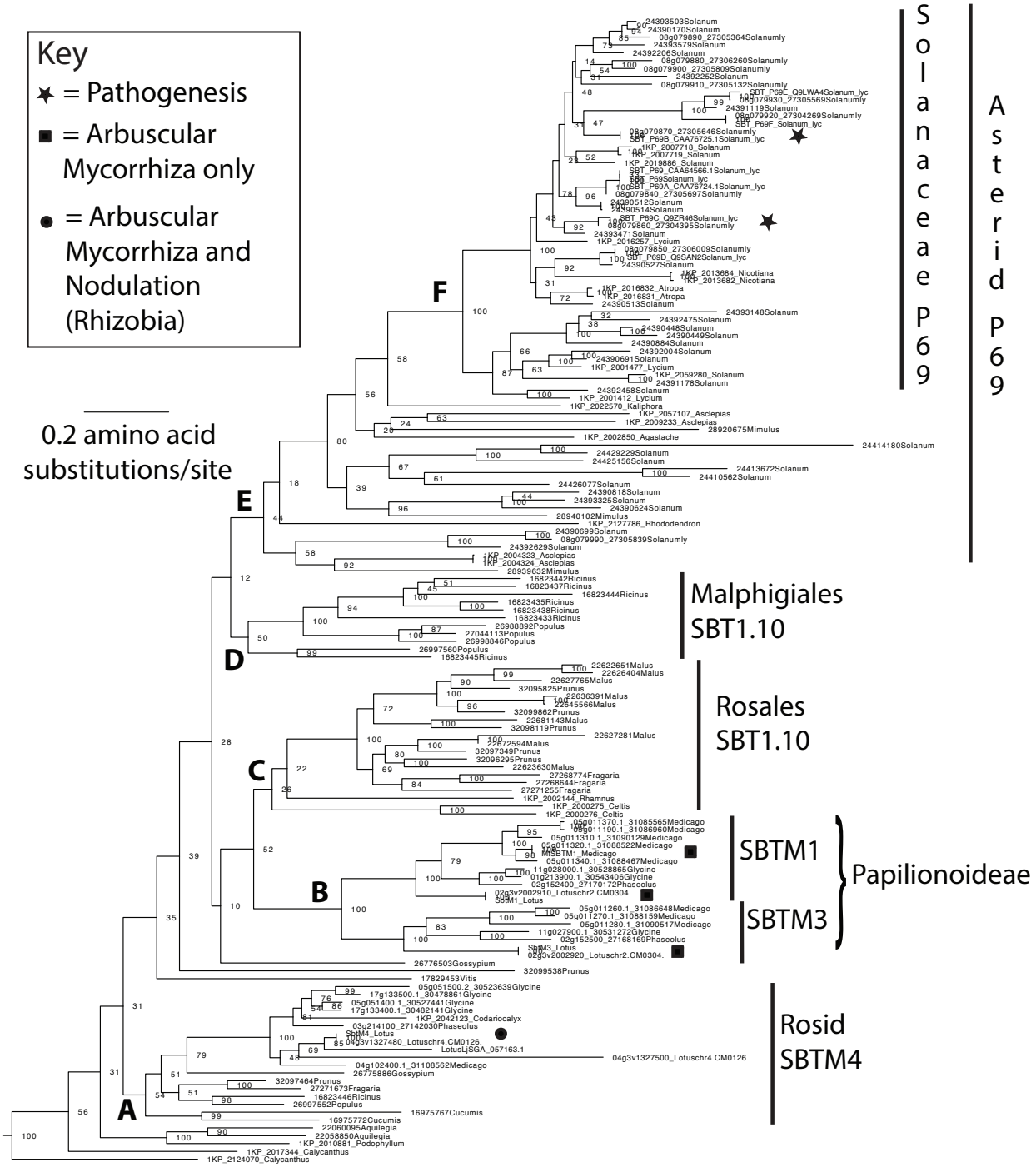



Figure 3.2: A portion of the maximum-likelihood phylogenetic tree shown in Fig. 1, to show details of the SBT1.10 gene clade. Bootstrap values out of 100 replicates are found at each node. Notable function characterizations are provided with symbols in Key. Node A: Orthologous gene lineage containing LjSBTM4. Node B: Gene lineage containing LjSBTM1 and LjSBTM3, duplication leading the LjSBTM1 and LjSBTM3 occurring before origin Papilionoideae, and restricted to this clade. Node C: Orthologous gene lineage of SBT1.10 genes, specific to Rosales. Node D: Orthologous gene lineage of SBT1.10 genes, containing multiple paralogs restricted to Malphigiales. Node E: Gene lineage containing P69 paralogs specific to asterids. Node F: Gene lineage containing all described P69 paralogs, which are restricted to the Solanaceae.
Gene names starting with letters were directly downloaded from NCBI; accession numbers and other information can be found in Supp. Table 1. Gene names starting with "1KP" are from the 1KP project, with 1KP sequence ID number and genus provided (further information provided in Supp. Table 4). All other genes are from full proteome data, contain Phytozome PACID# or Kazusa ID# (for Lotus japonicus), and begin with AGI code if available in annotation. Full species names can be found in Supp. Table 2. Solanum_tub is Solanum tuberosum and Solanum_lyc is Solanum lycopersicum

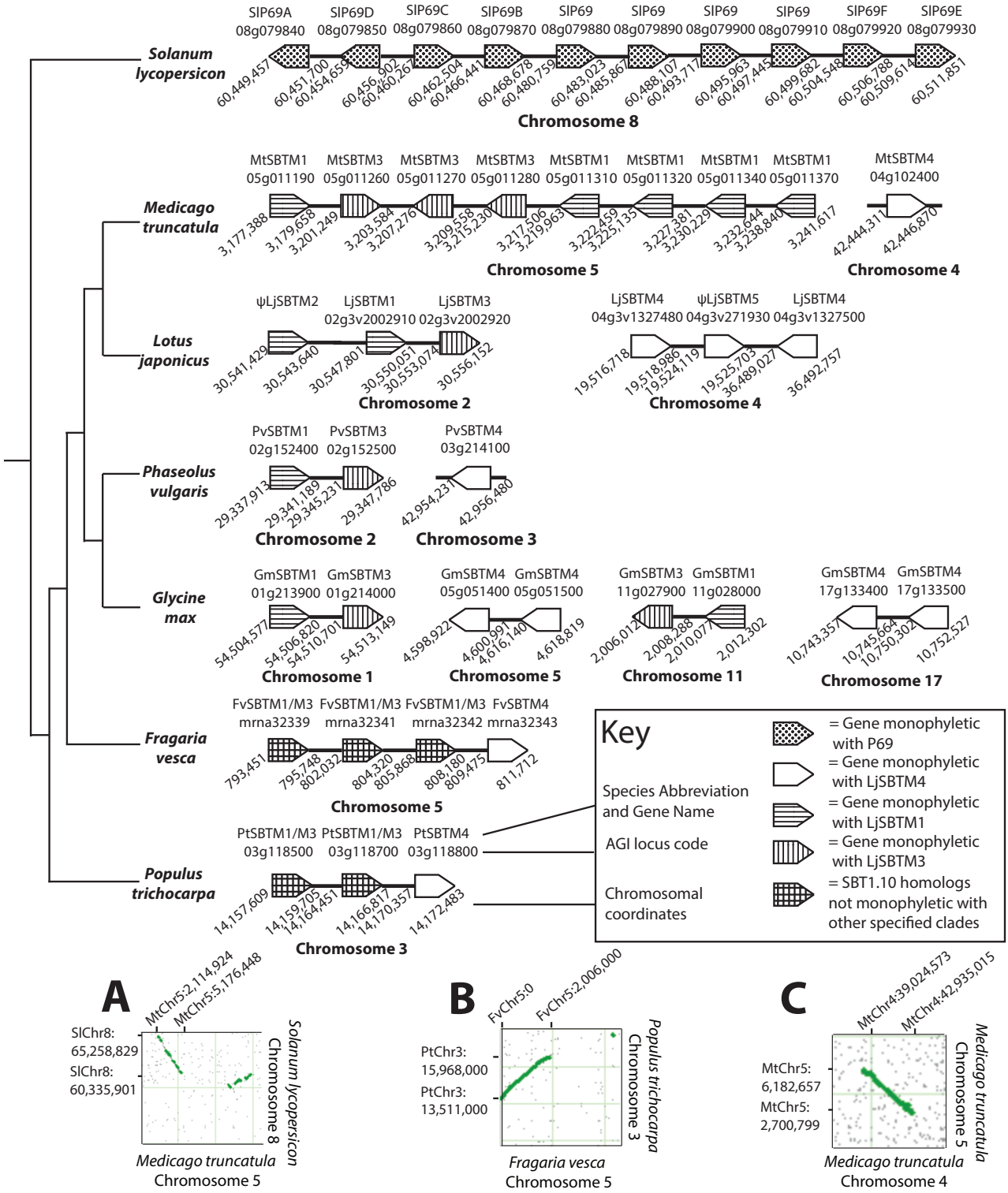



Figure 3.3: Schematic representation (not to scale) showing syntenic relationships of SBT1.10 genes in eudicots, with different paralogs, coordinates and strandedness retrieved from phytozome annotations (further information available in Supp. Table 3). Panels A, B, and C are retrieved from our synteny analyses in CoGe (comparative genomics) tool "SynMap" showing retained synteny. A: Conserved synteny between Solanum lycopersicum chromosome 8,and Medicago truncatula chromosome 5 in the regions containing SBT1.10 paralogs. B: Conserved synteny between Populus trichocarpa chromosome 3 and Fragaria vesca chromosome 5, supporting a conserved state of tandem arrangement between SBTM4 and SBTM1/M3 between malvids and NFC. C: Conserved synteny between Medicago truncatula chromosome 5 and chromosome 4, showing that the chromosomal region containing SBTM4 homolog shares ancestry with the chromosomal region containing SBTM1/M3

# CHAPTER 4

## Transcriptomics of Nodulation in *Elaeagnus umbellata*


### Abstract

Nodulation, the symbiosis in which nitrogen-fixing bacteria are housed in specialized nodule organs on plant roots, evolved multiple times independently in the Nitrogen-Fixing Clade of rosids. Each examined origin of nodulation involved the recruitment of homologous genes, meaning that nodules are deeply homologous structures. However, multiple lineages representing independent origins of nodulation have not been characterized on a genetic level. Here, we report the first transcriptomic study of the roots of *Elaeagnus umbellata,* an actinorhizal shrub for which the genetic basis of nodule development has not been examined, upon exposure to its nodule bacteria *Frankia*. Transcriptome assembly recovered multiple genes orthologous with those mediating nodulation in other lineages. The evolutionary history of these genes is examined.


### 4.1 Introduction

Nodulation, the symbiotic association in which nitrogen-fixing bacteria are housed in nodule organs on plant roots, is an important driver of the global nitrogen cycle in both wild and agricultural ecosystems (Galloway *et al.,* 1995; Smil 1999). This symbiosis is best known in the agriculturally important legume family (Fabaceae), which associate with the nitrogen-fixing bacteria collectively called “rhizobia” (Sprent, 2001).

Outside of the legumes, however, one genus in the Rosales (*Parasponia*, Cannabaceae) nodulates with rhizobia, and about 220 non-legume "actinorhizal" species in 25 genera spanning eight families in the Rosales, Fagales, and Cucurbitales nodulate with actinobacteria in the genus *Frankia* (Swensen *et al.,* 1996; Wall, 2000). Nodulation occurs only in these four orders of rosids, which together constitute a clade called the Nitrogen-Fixing Clade (NFC) (Soltis *et al.,* 1995). Nodules originated multiple times independently in the NFC, as reflected by its incongruous phylogenetic distribution, diverse bacterial symbionts and differences in infection mechanism, as well as nodule morphology and development patterns in different nodulating lineages (Swensen, 1996; Pawlowski & Sprent, 2008; Swensen & Benson, 2008; Doyle, 2011).

Nodulation requires the Common Symbiotic Signaling Pathway (CSSP), which also mediates the ancient arbuscular mycorrhizal (AM) symbiosis present across land plants (Wang *et al.*, 2010). The genetic underpinnings of nodulation have been most fully examined in the model papilionoid legumes *Medicago truncatula* and *Lotus japonicus* (for review, see Oldroyd, 2013), but genes involved in nodulation have also been identified in some nodulating non-legume species, notably the actinorhizal *Casuarina glauca* and *Alnus glutinosa* (Fagales) (Laplaze *et al.,* 2000; Hocher *et al.* 2011; Svistoonoff *et al.,* 2013), *Datisca glomerata* (Cucurbitales) (Demina *et al.,* 2013), and the rhizobial *Parasponia andersonii* (Rosales) (Op den Camp *et al.,* 2011). These studies show that the independent evolution of nodulation involved repeated recruitment of homologous CSSP genes (Doyle, 2011). Additional genes that do not mediate AM, such as hemoglobins and transcription factors such as *NIN* and the *NF-Y* complex, have also been recruited for nodulation multiple times independently in different lineages

(Vázquez-Limón *et al.,* 2012; Soyano *et al.,* 2013; Soyano & Hayashi, 2014; Clavijo *et al.,* 2015).

Nodules are thus "deeply homologous," since they originated by the repeated independent recruitment of homologous genes for a convergent function (Doyle, 2011). However, the precise orthologs recruited for nodulation are not identical in each independent origin of the symbiosis (Gopalasubramaniam *et al.,* 2008; Taylor & Qiu, 2017). The genetic basis of nodulation in convergently nodulating lineages has implications for the predictability of the evolution of nodulation, the feasibility of engineering it in non-nodulating crops, and the phylogenetic distribution of a genetic endowment or "predisposition" to nodulate. Despite over two decades of progress, the genetic basis of nodulation remains understudied in many lineages that represent independent origins of this symbiosis, especially in actinorhizal lineages (Pawlowski *et al.,* 2011; Svistoonoff *et al.,* 2015) and lineages with infection mechanisms that differ from model species (Demina *et al.,* 2013; Fabre *et al.,* 2015).

The model lineages for which nodulation has been characterized most extensively on a cellular and genetic level - the rhizobial legumes *M. truncatula* and *L. japonicus,* and the actinorhizal *A. glutinosa* and *C. glauca* (order Fagales) - are infected through root hair curling and the formation of transcellular infection threads leading to nodule primordia (Timmers *et al.,* 1999; Berg *et al.,* 1999a,b). In these lineages, perception of bacterial signaling molecules activates nuclear calcium spiking in the root epidermis, mediated by the CSSP (Oldroyd, 2013). In the case of rhizobial nodulation, these bacterial signaling molecules are lipo-chito-oligosaccharides (LCOs), which are similar to LCO factors produced by AM (glomeromycete) fungi (Maillet *et al.,* 2011). Perception of both AM

fungal and rhizobial LCO signals is mediated by LysM-motif receptor kinases (LysM-RKs) (De Mita *et al.,* 2014). *Frankia* signaling molecules remain uncharacterized, but are not digested by chitinase (Chabaud *et al.,* 2016). These signals appear not to be perceived by LysM-RKs, but also activate the CSSP in actinorhizal lineages (Markmann & Parniske, 2009; Svistoonoff *et al.,* 2014).

Plant perception of bacterial (or AM fungal) signaling pathways activates the CSSP through interaction with the leucine-rich repeat (LRR)-receptor kinase SYMRK, inducing nuclear calcium spiking. The nuclear pore complex formed by NUP85, NUP133, and NENA are required for nuclear calcium spiking, and may be involved in positioning several cation channels on the inner membrane of the nuclear envelope (Saito *et al.,* 2007; Kanamori *et al., 2006;* Groth *et al.,* 2010). These include MtMCA8, a calcium ATPase pump in the SERCA-type family (Capoen *et al.,* 2011) and potassium channels LjCASTOR (MtDMI1) (Ané *et al.,* 2004) and LjPOLLUX (Chen *et al.,* 2009), all of which are required for calcium spiking. This nuclear calcium spiking in turn induces a cascade of transcription factors leading to the reception of the symbiont by root hair curling and invagination, to form the transcellular IT (Madsen *et al.,* 2010; Oldroyd *et al.,* 2013; Singh *et al.,* 2014). Calcium spiking is perceived by CCAMK, which, in interaction with CYCLOPS, induces the expression of multiple genes which coordinate nodule organogenesis and the development of ITs, including the transcription factor genes *NIN, NSP1, NSP2* and the *NF-Y* complex (Soyano *et al.,* 2013; Soyano & Hayashi, 2014), and other genes such as *CERBERUS* and *VAPYRIN* (Madsen *et al.,* 2010; Murray *et al.,* 2011). Subtilases, a group of proteases involved in protein turnover in the

apoplastic space, are required for infection thread development in *C. glauca* and *L. japonicus* (Laplaze *et al.,* 2000; Takeda *et al.,* 2009).

The Elaeagnaceae represents an independent origin of actinorhizal nodulation from model nodulating lineages (Li *et al.,* 2015). The infection and development of nodules in this familydiffers substantially from these well-studied model lineages, based on examination of *Elaeagnus angustifolia* (Miller & Baker, 1985), and *Shepherdia argentea* (Racette & Torrey, 1989). In both species, *Frankia* infects the root intercellularly, through the middle lamella of epidermal cells, and moves through the apoplastic space to the nodule primordium without ITs (Miller & Baker, 1985; Racette & Torrey, 1989). This is typical of nodulation in the Rosales, in which the plasma membrane of each infected nodule cell invaginates to form new "vegetative hyphae" in each infected cell, though "invasive hyphae" ensheathed in transcellular ITs have been occasionally observed to cross from one infected cell to another, as in *Ceanothus* (Rhamnaceae) (Berry & Sunnell, 1990; Liu & Berry, 1991; Berg, 1999a,b). During intercellular infection, plant cell wall material (particularly pectic polysaccharides) is deposited around growing *Frankia* hyphae (Miller & Baker, 1985; Liu & Berry, 1991), and is partially dissolved and esterified as the hyphae grow through the intercellular spaces. It is unclear whether this digestion is a result of *Frankia* or plant enzymes, or both (Liu & Berry, 1991; Brewin, 2004). The Elaeagnaceae and Rhamnaceae are infected by EAN1pec, EUN1f and other strains from *Frankia* clade III, which infect five families within the orders Fagales and Rosales (Racette & Torrey, 1989; Navarro *et al.,* 1997; Clawson *et al.,* 1998; Benson et al., 2004).

While the CSSP signal transduction pathway was originally described in model

papilionoid legumes that are infected through root hair curling and the formation of transcellular ITs, there is some evidence that this pathway is also induced in lineages that are infected intercellularly (Imanishi *et al.,* 2011; Svistoonoff *et al.,* 2013; Svistoonoff *et al.,* 2014; Granqvist *et al.,* 2015). *D. glomerata* (Cucurbitales) is infected intercellularly, and shows upregulation of CSSP genes such as *CCAMK, CASTOR* and *CYCLOPS* in nodules (Demina *et al.,* 2013). In *Parasponia* (Cannabaceae, Rosales), the only non-legume to nodulate with rhizobia, bacteria enter intercellularly, through cracks in the epidermis subjacent to the formation of multicellular root hairs (Lancelle & Torrey, 1984; Lancelle & Torrey, 1985). More than 80 genes induced during nodulation in *Parasponia andersonii* are homologous with those induced in *M. truncatula,* including multiple CSSP genes (van Velzen *et al.,* 2017). *P. andersonii* also shows CSSP-mediated nuclear calcium spiking when exposed to bacterial nod signaling factors (Granqvist *et al.,* 2015). *Discaria trinveris,* an intercellularly-infected actinorhizal shrub in the Rhamnaceae (Valverde & Wall, 1999), also appears to employ the CSSP during nodulation, as evidenced by its spontaneous formation of nodules when transformed to express autoactive *ccamk* (Imanishi *et al.,* 2011). *CCAMK* is a central gene in the CSSP (Lévy *et al.,* 2004), and autoactive *ccamk* mutants form spontaneous nodules in nodulating lineages that use the CSSP (Madsen *et al.,* 2010; Svistoonoff *et al.,* 2013). *D. trinervis* is in the family Rhamnaceae, sister to Elaeagnaceae, though nodules in *D. trinervis* and *E. umbellata* are independently derived (Li *et al.,* 2015).

Until now, there has been little research into the genetic basis of nodulation in the Elaeagnaceae, though some nodule-specific genes, including polyubiquitins, chitinases, a chalcone isomerase and an auxin-repressed protein, have been isolated from mature *E.*

*umbellata* nodules (Kim *et al.,* 1999; Kim & An, 2002; An *et al.,* 2005; Kim *et al.,* 2007). It is not known whether the development of nodules in the Elaeagnaceae is mediated by genes homologous to those recruited in other nodulating lineages.

This study uses RNA-seq methods to examine genes induced during *Frankia* exposure in *Elaeagnus umbellata* (Elaeagnaceae), an actinorhizal shrub native to Eastern Asia that is invasive in Eastern North America (Czarapata, 2005). It also aims to analyze the phylogenetic relationship of *E. umbellata* genes to those involved in nodulation in other lineages. In so doing, this study examines whether the independent origin of nodulation in the Elaeagnaceae involved recruitment of genes homologous with those mediating nodulation in other lineages, and whether these homologs are orthologs or paralogs. Fundamentally, this study aims to answer the question of whether nodules in *E. umbellata* are deeply homologous with those in other nodulating lineages.

## 4.2 Methods

24 cuttings of an *E. umbellata* individual were collected on the grounds of Matthaei Botanical Gardens in Ann Arbor, Michigan. These cuttings were surface-sterilized in 0.525% sodium hypoclorite solution for 10 minutes, rinsed with dd$H_2O$ and rooted using IBA rooting hormone in soil that was autoclave-sterilized with two cycles of 30 minutes at 122 °C, and pots that were soaked in sodium hypoclorite for two hours. These cuttings were grown in a Thermo-Fisher 818 growth chamber at 25 °C with a 14 hr light/10 hr dark light cycle, and watered three times per week with ~200 ml of millipore deionized water. Test and control plants (12 each) were grown in identical conditions for 180 days, at which point plants in the test condition were inoculated with 5 ml of *Frankia*

strain EUN1f suspended in fructose-NH4CL media. At 4 time points following inoculation - 24 hours, 48 hours, 1 week and 3 weeks - lateral and woody roots were collected from three inoculated and three control cuttings, washed in ddH2O, and immediately frozen in liquid nitrogen.

Total RNA was extracted from these tissue collections following the protocol of Kalinowska *et al.* (2012), and assessed for RNA concentration and purity using a Nanodrop 2000**.** Four control samples and 12 test samples had adequate RNA concentration (>20 ng/ul) and purity for cDNA library preparation. Total RNA samples were transferred to the University of Michigan DNA Sequencing Core for cDNA library preparation and RNA-seq. Strand-specific cDNA libraries with polyA-selection were prepared, and pooled libraries were sequenced on an Illumina HiSeq platform, with 50 bp paired-end reads. De-novo transcriptome assembly was conducted using TRINITY v3.3 run on the Carbonate computer cluster at the Indiana University (Haas *et al.,* 2013). Differential gene expression was examined using the edgeR package; because of mismatched time points of samples for which adequate RNA was present, analyses were conducted using only the 48 hour and 3 week timepoint, which had two matched test and control samples each. These were analyzed using the exact test comparing inoculated (test) plants vs. uninoculated (control) plants (Robinson & Smyth, 2008).

For nodulation genes of interest, including the CSSP, representative sequences were downloaded from NCBI Genbank following a literature review. Primary transcript nucleotide sequences from 29 species (*Amborella trichocarpa, Aquilegia coerulea, Arabidopsis thaliana, Carica papaya, Citrus sinensis, Chenopodium quinoa, Chlamydomonas reinhardtii, Cucumis sativus, Daucus carota, Eucalyptus grandis,*

*Fragaria vesca, Glycine max, Gossypium raimondii, Malus domestica, Manihot esculenta, Marchantia polymorpha, Medicago truncatula, Mimulus guttatus, Musa acuminata, Oryza sativa, Phaseolus vulgaris, Physcomitrella patens, Populus trichocarpa, Prunus persica, Selaginella moellendorffii, Solanum lycopersicum, Trifolium pratense, Vitis vinifera,* and *Zea mays*) were downloaded from JGI Phytozome (http://www.phytozome.jgi.doe.gov). Additionally all *Lotus japonicus* cds sequences were downloaded from the Kazusa DNA Research Institute (http://www.kazusa.or.jp/lotus/). A local BLAST+ database of these nucleotide sequences was assembled in order to search for homologs of nodulation genes of interest from NCBI, using an e-value cutoff of 1e-10. The resulting sequence files were used to search the full Trinity assembly of *E. umbellata* transcripts (using an e-value cutoff of 1e-20). The retrieved sequences were aligned with characterized sequences from this study and the literature, manually curated to remove truncated or duplicated sequences, and assembled into a maximum likelihood gene phylogeny using RAxML with 1000 bootstrap replicates.

Additionally, in order to gain more confidence in our phylogenetic results by replication and to use more conserved amino acid sequences to search more distantly related plant clades, *E. umbellata* transcripts were translated to amino acid sequences using TRANSDECODER (Haas *et al.,* 2013). Amino acid sequences were downloaded from NCBI GenBank following a literature review, and were used to BLAST search a local BLAST+ database using blastp with an e-value cutoff of 1e-100. The resulting sequences were used to search the full Trinity translated assembly using blastp with an e-value cutoff of 1e-100. These sequences were used to search the 1KP database to gain

greater phylogenetic coverage. The resulting amino acid sequence files were aligned with MUSCLE (Edgar, 2004) and maximum likelihood phylogenetic analysis was run on RAxML (Stamatakis, 2006) with 100 bootstrap replicates. These gene phylogenies allow for interpretation of evolutionary history of genes recruited independently for nodulation in *E. umbellata.*

## 4.3 Results and Discussion

*Evolutionary History of Nodulation Genes*

This study represents the first transcriptomic examination of *Frankia* exposure in a species in the Elaeagnaceae. Thus, we analyzed the gene phylogenies of many assembled sequences to determine their relationship with genes involved in nodulation in other lineages, in order to assess the degree of deep homology between the Elaeagnaceae and other nodulating lineages. Because sampled tissue was from roots exposed to *Frankia,* we were able to assemble many *Elaeagnus* genes not captured in the 1KP dataset, which sampled young leaf tissue of *Elaeagnus pungens* (Fig. 4.2-4.4, Fig. 4.6-4.13). Our use of RNA-seq technology allowed for orders of magnitude more genetic sequence data over previous gene sequencing of *E. umbellata* nodules, which yielded less than a dozen genes (Kim *et al.,* 1999; Kim & An, 2002; An *et al.,* 2005; Kim *et al.,* 2007). Our transcriptome assembly recovered homologs of multiple genes mediating nodulation in other lineages, though few were significantly differentially expressed between inoculated and non-inoculated *E. umbellata* replicates in our dataset (Table 2).

Understanding the evolutionary history of genes involved in nodulation in different lineages is crucial to understanding the origins of nodulation (Taylor & Qiu,

2017; van Velzen *et al.,* 2018). Accordingly, we constructed gene phylogenies of 13 genes involved in nodulation in other lineages, using the phytozome and 1KP sequence databases to create wide taxonomic sampling to identify the evolutionary history of these genes across the land plant phylogeny, including the phylogenetic distribution of duplications and losses of these genes. By including gene sequences of *E. umbellata* homologs from our transcriptome assembly to these gene trees, we were able to provide valuable information from an independently nodulating lineage from which these gene sequences were not previously available. Gene phylogenies for most CSSP genes were largely congruent with the species phylogeny, supporting the idea that CSSP genes are generally vertically inherited as single-copy orthologs with relatively few duplications (Markmann *et al.,* 2008; Oldroyd, 2013). We did find an exception to this vertical inheritance: in the subtilase gene family, different paralogs were recruited for nodulation in different nodulating lineages (Fig. 4.1; Taylor & Qiu, 2017). Finally, our *E. umbellata NIN* and *RPG* sequences confirmed van Velzen's (2018) finding of widespread loss of these genes in non-nodulating lineages of the NFC by providing the only other sequence data from a nodulating lineage in the Rosales.

*E. umbellata Subtilases*

Subtilases are a large family of proteases that are involved in protein turnover in a variety of processes in land plants, including symbiotic interactions (Schaller *et al.,* 2012). This family expanded rapidly through several rounds of duplications early in the angiosperms, yielding multiple lineages in the *SBT1* subfamily that mediate a variety of symbiotic interactions (Taylor & Qiu, 2017), including pathogenesis, AM and nodulation

(Tornero *et al.,* 1996; Laplaze *et al.,* 2000; Takeda *et al.,* 2009). Subtilases have been shown to be involved in nodulation in both the rhizobial legume *Lotus japonicus* (Takeda *et al.,* 2009) and the actinorhizal *Alnus glutinosa, Allocasuarina verticillata,* and *Casuarina glauca* (Fagales) (Laplaze *et al.,* 2000).

We previously reported that the subtilases recruited for nodulation in the Fagales (Laplaze *et al.,* 2000) and the Fabales (Takeda *et al.,* 2009) are paralogous to one another, having diverged early in angiosperm evolution (Taylor & Qiu, 2017). Our subtilase phylogeny here replicates this finding with different taxon sampling and using nucleotide rather than amino acid sequence data (Fig. 4.1). Further, a subtilase transcript assembled from our *E. umbellata* transcriptome is orthologous with *LjSBTM4* (Fig. 4.1B; 87% BS), a subtilase required for proper nodulation in *Lotus japonicus* (Takeda *et al.,* 2009). We found no subtilase transcripts orthologous with *Alnus glutinosa AG12* nor *Casuarina glauca CG12* (Fig. 4.1C).

Subtilases are involved in transcellular IT development during nodulation in both actinorhizal Fagales and rhizobial legumes (Svistoonoff *et al.,* 2003; Takeda *et al.,* 2009). However, *E. umbellata* does not form transcellular ITs, instead being infected through the middle lamella of epidermal cells, with *Frankia* hyphae growing through the apoplastic space between cortical cells towards the nodule primordium (Miller & Baker, 1985; Racette & Torrey, 1989). Demina *et al.* (2013) found a similar pattern in examining the transcriptome of *D. glomerata* (Cucurbitales), which is also intercellularly infected but also showed upregulation of several genes associated with IT formation, including *VAPYRIN.* Several other IT related genes, such as lectin and expansin, were also differentially expressed following *Frankia* inoculation (Table 4.1). Intercellularly

infected plants still deposit plant cell wall material around the infecting nodule bacteria, and actinorhizal nodulators form IT-like sheaths around *Frankia* hyphae as they infect nodule cells (Brewin *et al.,* 2004; Pawlowski & Demchenko, 2012); perhaps this *E. umbellata SBTM4* is involved in this restructuring of plant cell wall components in the apoplastic space.

Furthermore, it is surprising that the *E. umbellata* subtilase would be orthologous with the legume *LjSBTM4* rather than *CG12*, the subtilase involved in IT development in actinorhizal *C. glauca*. IT formation in legumes and actinorhizal species is quite different: actinorhizal ITs (sometimess called "invasive hyphae") have no IT matrix, and the bacteria are instead in direct contact with the wall of the IT lumen (Pawlowski & Demchenko, 2012). The fact that the actinorhizal *E. umbellata* recruited a separate, distantly related paralog to the actinorhizal *C. glauca* copy constitutes further evidence that these symbioses evolved independently (Doyle, 1994). However, there is transcriptional conservation between *SBTM4* and *CG12;* transgenic *M. truncatula* expresses *CG12* in IT formation (Svistoonoff *et al.,* 2003; Svistoonoff *et al.,* 2004). Thus, the differential recruitment of different subtilase paralogs provides a potential counter-example to the "ortholog conjecture," the idea that orthologous genes are more likely to have similar functions than paralogous genes (Gabaldon & Koonin, 2013).

*The CSSP Genes*

We assembled *E. umbellata* orthologs of the CSSP nuclear pore components *NUP85, NUP133,* and *NENA,* and both nucleotide and amino acid-based phylogenies show strong congruence between the gene and species phylogeny, indicating vertical

inheritance of these single-copy orthologs. We recovered *NUP85* and *NUP133* sequences from across the land plants, constructing the most extensive phylogeny of these genes to date and replicating findings for other CSSP genes of vertical inheritance without retained duplications (Figs. 4.2, 4.3). The gene phylogeny of *NUP85* sequences is congruent with the species phylogeny; our assembled *E. umbellata* transcript is in a monophyletic clade with other sequences in the Rosales (94% BS, Fig. 4.2A), and shows the correct topology within that clade, with the *E. umbellata* sequence as sister to two sequences in *Rhamnus* species, the most closely related in the dataset (Fig. 4.2A). All *NUP85* sequences from the NFC also form a monophyletic group, though with low bootstrap support (19% BS). The gene phylogeny of *NUP133* is also largely congruent with the species phylogeny, and our assembled *E. umbellata* ortholog is monophyletic with all other Rosales homologs (95% BS, Fig. 4.3A). All NFC *NUP133* homologs are monophyletic (79% BS, Fig. 4.3A). Our *NENA* gene phylogeny also showed vertical inheritance with no conserved paralogs (Fig. 4.4). Our assembled *E. umbellata NENA* transcript is orthologous to those in the Rosaceae (Fig. 4.4; 38% BS).

Our gene phylogeny of LysM-RKs supports previous findings that *LjNFR1* is an ortholog of *AtCERK1* (Fig. 4.5A), and *Parasponia NFP* is orthologous to *NFR5* with 100% BS (Fig. 4.5B; Nakagawa *et al.,* 2011; Streng *et al.,* 2011; Op den Camp *et al.,* 2011). Our assembly uncovered several *E. umbellata* LysM-RKs, including one orthologous with *NFR1* with weak support (56% BS; Fig. 4.5A). If this weakly-supported relationship is real, it would be surprising, since *Frankia* does not appear to produce LCO signals; actinorhizal signal receptors likely belong to a different family (Markmann & Parniske, 2009; Chabaud *et al.,* 2016). Eurosid *SYMRK* sequences, which contain a third

LRR motif (Markmann *et al.,* 2008), formed a clade with weak support (36% BS), and within that clade we found weak bootstrap support for orthogroups of the NFC (25% BS). Within the NFC, there were monophyletic orthogroups in the Fabales (94% BS), Cucurbitales (99% BS), and Rosales (84%), including our assembled *E. umbellata SYMRK* transcript (Fig. 4.6).

Our phylogeny supports the antiquity of the paralogous ion channels *CASTOR* and *POLLUX,* which diverged before the divergence of angiosperms and gymnosperms (Fig. 4.7; Chen *et al.,* 2009; Wang *et al.*, 2010; Delaux *et al.,* 2013). Both orthologous gene lineages contain orthologs from the basal angiosperm *Amborella trichopoda* as well as several gymnosperms, rendering the *CASTOR* lineage older than the previously reported angiosperm origin (Delaux *et al.,* 2013). For the most part, the phylogeny of both paralogs is congruent with the species phylogeny, showing a history of vertical inheritance, though several major lineages are incongruent with the species phylogeny or show very low bootstrap support due to high levels of sequence conservation leading to a lack of ancestry informative sites (Fig. 4.7). Our assembled *E. umbellata POLLUX* was monophyletic with all other *POLLUX* orthologs in the Rosales (84% BS), though our gene phylogeny showed low support (9% BS) for an NFC *POLLUX* orthogroup (Fig. 7). Our MCA8 phylogeny showed a duplication event likely in the common ancestor of angiosperms, and another one within rosids (Fig. 4.8). Due to lack of gymnosperm and basal eudicot sequences, the exact timing of these duplication events is difficult to pinpoint at this stage, even though the recovered tree topology is well supported. Our assembled *E. umbellata* sequences fall into the three clades derived from these dupliction events, and one is orthologous to *MtMCA8,* a calcium ATPase pump required for calcium

spiking in *M. truncatula* (Fig. 4.8B; Capoen *et al.,* 2011). Interestingly, in the paralogous lineage that was derived from the duplication in early rosids, there were no sequences from the Fabales. Given the number of legume proteomes and transcriptomes searched in this analysis and the presence of homologs from the Rosales, Fagales and Cucurbitales, the absence of the sequence is most likely an indication of loss of the paralog in legumes or even Fabales, possibly indicating some legume-specific selection against retaining this gene.

The gene phylogenies of *CCAMK* and *CYCLOPS* were largely congruent with species phylogenies, suggesting vertical inheritance as a single-copy ortholog in both genes (Fig. 4.9, Fig. 4.13), as previously reported (Wang *et al.,* 2010). Our assembled *E. umbellata CYCLOPS* transcript was monophyletic with all *CYCLOPS* orthologs from the Rosales (97% BS), and the NFC *CYCLOPS* forms a weakly-supported clade (19% BS). Our *CCAMK* phylogeny largely followed the species phylogeny for major groups, for example supporting a monophyletic *CCAMK* orthogroup in the Fabaceae (95% BS, Fig 4.9), and replicated the findings of Wang *et al.* (2010) and Delaux *et al.* (2013) showing *CCAMK* as an ancient, vertically inherited gene dating back to charophyte algae. *CCAMK* in the Rosales was not resolved as a monophyletic group, however, and non-legume NFC came out as a poorly-supported grade (Fig. 4.9). Our assembled *E. umbellata CCAMK* was in a small, poorly supported clade with *Cucumis sativa* and *Carica papaya* (Brassicales) (12% BS).

*Genes Downstream of the CSSP*

The aquaporin nodulin-26 (*NIP2*) is expressed in *Glycine max* during nodulation and is a component of the symbiosome membrane (Rivers *et al.,* 1997). *NOD26/NIP2* is also implicated in AM symbiosis, and may be involved in ammonia uptake, based on yeast complementation data (Uehlein *et al.,* 2007). Our *NOD26/NIP2* phylogeny was generally congruent with the species phylogeny, but with weak bootstrap support throughout (Fig. 4.10), perhaps owing to the documented pattern of conservation of the NPA domain and extreme divergence of other regions in this gene family (Liu *et al.,* 2009). Our transcriptome assembly recovered an *E. umbellata NOD26* homolog which was monophyletic with several other Rosales *NOD26* homologs with weak support (15% BS). Despite incongruence between species and genes trees, our *E. umbellata NOD26* homolog tentatively appears to be orthologous to legume *NOD26,* as our BLAST search of several full genomes in the Rosaceae did not retrieve any more closely related Rosaceous aquaporin clades (Fig. 4.10).

Our *NLP/NIN* phylogeny largely agrees with the results of Soyano & Hayashi's (2014) neighbor-joining tree of amino acid sequences, which found a eudicot *NIN*/*NLP1* clade forming a monophyletic group with non-eudicot angiosperm *NLP1* sequences (Fig. 4.11). However, we found that eudicot *NIN/NLP1* actually consisted of two paralogous lineages that likely duplicated and diverged early in eudicot evolution, with homologs from the basal eudicots *Nelumbo sp.* (Proteales) and *Buxus sempervierns* (Buxales) as outgroup (Fig. 4.11). We found NFC *NIN* to be orthologous with *A. thaliana NLP1* (*AtNLP1*); this orthologous lineage is confined to eudicots, with a homolog from *Dillenia indica* in the basal position (Fig. 4.11D; 100% BS). Within this lineage, *NIN* is

orthologous with other *NLP1* homologs; only one NFC orthogroup is represented in the *NIN/NLP1* lineage, rather than the two that would be expected if *NIN* represented an NFC-specific paralog (Fig. 4.11D). We found that the genes previously identified as "*NLP1"* in *Lotus japonicus* and *Trema levigata* are in a separate rosid-specific lineage orthologous with *A. thaliana NLP4* and *NLP5* (100% BS), replicating a previous tentative finding of a parsimony phylogeny of 16 NIN-like proteins (Schauser *et al.,* 2005); we called this lineage NLP1.2. These two paralogous lineages (eudicot *NLP1* and rosid *AtNLP1.2*) were co-orthologous with *Oryza sativa NLP1,* and together this *NLP1* orthologous lineage (100% BS) arose during the origin of angiosperms, since it included an ortholog from the basalmost angiosperm *Amborella trichopoda* but no non-angiosperms (Fig. 4.11). Aside from *NIN* in the NFC, none of the functions of the genes in this clade are known. *A. thaliana NLP6* and *NLP7* are transcriptional regulators involved in nitrate response (Castaings *et al.,* 2009; Yonishi *et al.,* 2013), and these genes were orthologous to *Oryza sativa NLP3,* in a lineage including gymnosperm orthologs (85% BS, Fig. 4.11).

Our assembled *E. umbellata NIN* transcripts are monophyletic (100% BS) with all *NIN* sequences characterized as involved in nodulation: *Lotus japonicus NIN* and *Casuarina glauca NIN* (Fagales), which have been shown to be required for nodulation (Schauser *et al.,* 1999; Clavijo *et al.,* 2015), and *Parasponia andersonii NIN* and *Datisca glomerata NIN,* shown to be induced during nodulation (Demina *et al.,* 2013; van Velzen *et al.,* 2018). The *NIN* orthogroup has few orthologs from non-nodulating species (Fig. 4.11D), other than an intact *Ziziphus jujube* sequence, as well as *Trema spp.* and *Prunus persica* sequences that van Velzen *et al.* (2018) found to be pseudogenized. In this

respect, our phylogeny almost exactly replicated the findings of van Velzen *et al.,* (2018), suggesting loss of the *NIN* paralog in non-nodulating lineages. However, we did find three additional *NIN* sequences from non-nodulating lineages sequenced as part of the 1KP project: *Rhamnus caroliniana*, *Cannabis sativa,* and *Ficus religiosa,* sampled from young leaves, stem tissue, and leaf tissue, respectively – each of these sequences appears as truncated in our amino acid alignment, replicating the findings of van Velzen *et al.,* (2018). However, it should be noted the *NIN* itself is orthologous in evolutionary history, if not identical in function, to the *NLP1* genes found in other eudicots, and is not an NFC-specific paralog (Fig. 4.11D)

*RPG,* a gene involved in polar growth of ITs in *M. truncatula* (Arrighi *et al.,* 2008), has been shown to be repeatedly lost in non-nodulating lineages of the NFC (van Velzen *et al.,* 2018). Our *RPG* gene phylogeny replicates this finding (Fig. 4.12), with a monophyletic group of *RPG* orthologs from only nodulating lineages of the NFC (84% BS), with the exception of *Trema orientalis,* which has a pseudogenized copy of this gene (van Velzen *et al.,* 2018). Again, this NFC *RPG* lineage does not represent an NFC-specific paralog, and its position in the gene phylogeny is roughly congruent the position of the NFC in the species phylogeny, suggesting that *RPG* orthologs are retained in many non-nodulating lineages (Fig. 4.12).

*Differential Gene Expression*

No nodules or root swellings were observed on the roots of *Frankia*-inoculated *E. umbellata* cuttings; however, this study was designed to capture early signaling events in nodulation, and three weeks may not have been sufficient time for nodules to form (Wall

& Berry, 2007). Because of mismatched timepoints in samples with sufficient RNA for sequencing, only four control samples (2 at 48 hours and 2 at 3 weeks) were compared, using an exact test (Robinson & Smyth, 2008). Accordingly, differential gene expression should be considered with substantial caution. Several genes related to those involved in nodulation in other lineages were significantly differentially expressed (Table 4.1, Table 4.2). Several genes involved in nodule bacterial accommodation and formation of fixation threads were significantly differentially expressed, including polygalacturonase, lectin, peroxidase and expansin (Table 4.1). *E. umbellata* does not form transcellular ITs, but does form IT-like elements around infected cells. IT thread gene homologs were also found to be upregulated in *Datisca glomerata*, another intercellularly-infected actinorhizal species (Demina *et al.,* 2013). We found significant downregulation of two *E. umbellata* ABC transporter genes (Table 4.1). ABC transporters are downregulated during nodulation in *M. truncatula*, as part of a suppression of defense responses (Limpens *et al.,* 2013).

## 4.4 Conclusion

Despite not forming ITs during infection, our *E. umbellata* transcriptome showed significant differential expression of several genes implicated in IT growth during infection in other lineages, such as a lectin protein kinase and polygalacturonase (Table 1, Brewin, 2004). Further, we assembled several genes implicated in IT growth, such as *RPG* and subtilases (Arrighi *et al.,* 2008; Takeda *et al.,* 2009). Demina *et al.* (2013) found similar results in the nodulation transcriptome of *Datisca glomerata*, suggesting that the deposition of cell wall-like material in the apoplastic space during infection

might have a similar genetic basis to infection by means of transcellular infection threads. There have been many transitions between intracellular and intercellular infection mechanisms in the legumes (Sprent, 2001), and some nodulating species can switch infection modes based on stress (Goormachtig *et al.,* 2004). More research into the genetic basis of intercellularly-infected nodulating plant lineages could resolve the question of how much genetic similarity there is between these infection modes (Sinharoy *et al.,* 2009; Imanishi *et al.,* 2009).

Nodulation is a complex trait that evolved multiple times independently, by repeated recruitment of homologous genes to serve convergent functions (Doyle, 2011). However, studies of the genetic basis of nodulation have focused on just a few model lineages, and it remains unclear whether each instance of nodulation involved recruitment of the same homologous genes (Pawlowski *et al.*, 2011; Svistoonoff *et al.*, 2014). This study of the *E. umbellata* transcriptome assembled multiple homologs of many genes which have been shown to be involved in nodulation in other lineages. Phylogenetic analyses of these sequences provided clear evidence to support that in most genes, including *NUP85, NUP133, NENA, LysM-RK, SYMRK, CASTOR, POLLUX, MCA8, CCAMK, NOD26, NIN, RPG,* and *CYCLOPS,* orthologous sequences to the genes with characterized functions in legumes and non-legumes are present in *E. umbellata*.

In the case of subtilases, however, different paralogs have been recruited in different lineages (Taylor & Qiu, 2017). Surprisingly, our assembled *E. umbellata* subtilase sequence was orthologous to those in legumes with rhizobia symbionts, but paralogous to those in another actinorhizal lineage, Fagales, with *Frankia* symbionts (Fig. 1). This example shows that the homologous genes independently recruited for

nodulation in different lineages are not always orthologs. It provides both evidence of the non-homology of nodulation in different lineages (Doyle, 1994), and also points to the possible functional equivalence of different paralogous genes (Svistoonoff *et al.,* 2003; Svistoonoff *et al.,* 2004). Our phylogenetic analyses of these *E. umbellata* genes as well as a large number of other plant sequences also identified new gene clades and revealed previously unknown gene dupication events. How these dupicated genes are involveed in nodulation, especially in poorly studied actinorhizal lineages, requires further investigation in future studies. Overall, these gene histories add to our understanding of gene recruitment and deep homology in the evolutionary origin of nodulation.

**References:**


**An, C.S., Kim, H.B., Lee, S.H., Jang-Hyun, J., Oh, C.J. and Lee, H.,** 2005. Gene Expression in the Root Nodules of Elaeagnus umbellata. In *Biological Nitrogen Fixation, Sustainable Agriculture and the Environment* (pp. 207-208). Springer, Dordrecht.

**Ané, J.M., Kiss, G.B., Riely, B.K., Penmetsa, R.V., Oldroyd, G.E., Ayax, C., Lévy, J., Debellé, F., Baek, J.M., Kalo, P. and Rosenberg, C.,** 2004. Medicago truncatula DMI1 required for bacterial and fungal symbioses in legumes. *Science*, *303*(5662), pp.1364-1367.

**Arrighi, J.F., Godfroy, O., de Billy, F., Saurat, O., Jauneau, A. and Gough, C.,** 2008. The RPG gene of Medicago truncatula controls Rhizobium-directed polar growth during infection. *Proceedings of the National Academy of Sciences*, *105*(28), pp.9817-9822.

**Benson, D.R., Vanden Heuvel B.D., Potter D., Vanden Heuvel B.D., Potter D., Potter D.** 2004. Actinhorizal symbioses: diversity and biogeography. In: Gillings M., editor. Plant microbiology. BIOS Scientific Publishers Ltd.; Oxford.

**Berg, R. H.** 1999a. Frankia forms infection threads. *Can. J. Bot*., 77, 1327-1333.

**Berg, R.H.**, 1999b. Cytoplasmic bridge formation in the nodule apex of actinorhizal root nodules. *Canadian Journal of Botany*, *77*(9), pp.1351-1357.

**Berry, A. M., and Sunell, L. A.** 1990. The infection process and nodule development. Pages 61-81 in: The Biology of Frankia and Actinorhizal Plants. C. R. Schwintzer and J. D. Tjepkema, eds. Academic Press, New York.

**Brewin, N.J.,** 2004. Plant cell wall remodelling in the Rhizobium–legume symbiosis. *Critical Reviews in Plant Sciences*, *23*(4), pp.293-316.

**Capoen, W., Sun, J., Wysham, D., Otegui, M.S., Venkateshwaran, M., Hirsch, S., Miwa, H., Downie, J.A., Morris, R.J., Ané, J.M. and Oldroyd, G.E.,** 2011. Nuclear membranes control symbiotic calcium signaling of legumes. *Proceedings of the National Academy of Sciences*, *108*(34), pp.14348-14353.

**Castaings, L., Camargo, A., Pocholle, D., Gaudon, V., Texier, Y., Boutet‐Mercey, S., Taconnat, L., Renou, J.P., Daniel‐Vedele, F., Fernandez, E. and Meyer, C.,** 2009. The nodule inception-like protein 7 modulates nitrate sensing and metabolism in Arabidopsis. *The Plant Journal*, *57*(3), pp.426-435.

**Chabaud, M., Gherbi, H., Pirolles, E., Vaissayre, V., Fournier, J., Moukouanga, D., Franche, C., Bogusz, D., Tisa, L.S., Barker, D.G. and Svistoonoff, S.,** 2016. Chitinase‐resistant hydrophilic symbiotic factors secreted by Frankia activate both Ca2+ spiking and NIN gene expression in the actinorhizal plant Casuarina glauca. *New Phytologist*, *209*(1), pp.86-93.

**Chen, C., Fan, C., Gao, M. and Zhu, H.,** 2009. Antiquity and function of CASTOR and POLLUX, the twin ion channel-encoding genes key to the evolution of root symbioses in plants. *Plant Physiology*, *149*(1), pp.306-317.

**Clavijo, F., Diedhiou, I., Vaissayre, V., Brottier, L., Acolatse, J., Moukouanga, D., Crabos, A., Auguy, F., Franche, C., Gherbi, H. and Champion, A.,** 2015. The Casuarina NIN gene is transcriptionally activated throughout Frankia root infection as well as in response to bacterial diffusible signals. *New Phytologist*, *208*(3), pp.887-903.

**Clawson, M.L., Carú, M. and Benson, D.R.,** 1998. Diversity of Frankia strains in root nodules of plants from the families Elaeagnaceae and Rhamnaceae. *Applied and environmental microbiology*, *64*(9), pp.3539-3543.

**Czarapata, E.J.,** 2005. *Invasive plants of the upper Midwest: an illustrated guide to their identification and control*. Univ of Wisconsin Press.

**De Mita S, Streng A, Bisseling T, Geurts R. 2014.** Evolution of a symbiotic receptor through gene duplications in the legume–rhizobium mutualism. *New Phytologist* **201(3):** 961-972.

**Delaux, P.M., Séjalon-Delmas, N., Bécard, G. and Ané, J.M.,** 2013. Evolution of the plant–microbe symbiotic 'toolkit'. *Trends in plant science*, *18*(6), pp.298-304.

**Demina IV, Persson T, Santos P, Plaszczyca M, Pawlowski K. 2013.** Comparison of the nodule vs. root transcriptome of the actinorhizal plant *Datisca glomerata*: actinorhizal nodules contain a specific class of defensins. *PloS one* **8(8):** e72442.

**Doyle JJ. 1994.** Phylogeny of the legume family: an approach to understanding the origins of nodulation. *Annual Review of Ecology and Systematics*, 325-349.

**Doyle JJ. 2011.** Phylogenetic perspectives on the origins of nodulation. *Molecular Plant-Microbe Interactions* **24:** 1289–129

**Edgar, R.C., 2004.** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, *32*(5), pp.1792-1797.

**Fabre, S., Gully, D., Poitout, A., Patrel, D., Arrighi, J.F., Giraud, E., Czernic, P. and Cartieaux, F.,** 2015. Nod factor-independent nodulation in Aeschynomene evenia required the common plant-microbe symbiotic toolkit. *Plant physiology*, *169*(4), pp.2654-2664.

**Gabaldón, T. , and E. V. Koonin.** 2013 . Functional and evolutionary implications of gene orthology. *Nature Reviews. Genetics* 14 : 360 – 366.

**Galloway, J.N., Schlesinger, W.H., Levy, H., Michaels, A. and Schnoor, J.L.,** 1995. Nitrogen fixation: Anthropogenic enhancement‐environmental response. *Global biogeochemical cycles*, *9*(2), pp.235-252.

**Granqvist, E., Sun, J., Op den Camp, R., Pujic, P., Hill, L., Normand, P., Morris, R.J., Downie, J.A., Geurts, R. and Oldroyd, G.E.,** 2015. Bacterial‐induced calcium oscillations are common to nitrogen‐fixing associations of nodulating legumes and non‐legumes. *New Phytologist*, 207(3), pp.551-558.

**Goormachtig, S., Capoen, W., James, E.K. and Holsters, M.,** 2004. Switch from intracellular to intercellular invasion during water stress-tolerant legume nodulation. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(16), pp.6303-6308.

**Gopalasubramaniam, 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: Structure, Function, and Bioinformatics*, *72*(1), pp.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. *The Plant Cell*, *22*(7), pp.2509-2526.

**Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M. and MacManes, M.D.,** 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, *8*(8), p.1494.

**Hocher V, Alloisio N, Auguy F, Founier P, Doumas P, Pujic P, Gherbi H.** 2011. Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant Physiology* Online publication.

**Imanishi, L., Vayssières, A., Franche, C., Bogusz, D., Wall, L. and Svistoonoff, S.,** 2011. Transformed hairy roots of Discaria trinervis: a valuable tool for studying actinorhizal symbiosis in the context of intercellular infection. *Molecular plant-microbe interactions*, *24*(11), pp.1317-1324.

**Kalinowska, E., Chodorska, M., Paduch-Cichal, E. and Mroczkowska, K.,** 2012. An improved method for RNA isolation from plants using commercial extraction kits. *Acta Biochimica Polonica*, *59*(3).

**Kanamori, N., Madsen, L.H., Radutoiu, S., Frantescu, M., Quistgaard, E.M., Miwa, H., Downie, J.A., James, E.K., Felle, H.H., Haaning, L.L. and Jensen, T.H.,** 2006. A nucleoporin is required for induction of Ca2+ spiking in legume nodule development and essential for rhizobial and fungal symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), pp.359-364.

**Kim, H.B., Lee, S.H. and An, C.S.,** 1999. Isolation and characterization of a cDNA clone encoding asparagine synthetase from root nodules of Elaeagnus umbellata. *Plant science*, *149*(2), pp.85-94.

**Kim, H.B. and An, C.S.,** 2002. Differential expression patterns of an acidic chitinase and a basic chitinase in the root nodule of Elaeagnus umbellata. *Molecular plant-microbe interactions*, *15*(3), pp.209-215.

**Kim, H.B., Lee, H., Oh, C.J., Lee, N.H. and An, C.S.,** 2007. Expression of EuNOD-ARP1 encoding auxin-repressed protein homolog is upregulated by auxin and localized to the fixation zone in root nodules of Elaeagnus umbellata. *Molecules & Cells (Springer Science & Business Media BV)*, *23*(1).

**Konishi, M. and Yanagisawa, S.,** 2013. Arabidopsis NIN-like transcription factors have a central role in nitrate signalling. *Nature Communications*, *4*, p.1617.

**Lancelle, S.A. and Torrey, J.G.,** 1984. Early development of Rhizobium-induced root nodules of *Parasponia rigida*. I. Infection and early nodule initiation. *Protoplasma*, *123*(1), pp.26-37.

**Lancelle, S.A. and Torrey, J.G.,** 1985. Early development of Rhizobium-induced root nodules of Parasponia rigida. II. Nodule morphogenesis and symbiotic development. *Canadian journal of botany*, *63*(1), pp.25-35.

**Laplaze, L., Duhoux, E., Franche, C., Frutz, T., Svistoonoff, S., Bisseling, T., Bogusz, D. and Pawlowski, K.,** 2000. Casuarina glauca prenodule cells display the same differentiation as the corresponding nodule cells. *Molecular plant-microbe interactions*, *13*(1), pp.107-112.

**Lee, H., Hur, C.G., Oh, C.J., Kim, H.B., Park, S.Y. and An, C.S.,** 2004. Analysis of the root nodule-enhanced transcriptome in soybean. *Molecules & Cells (Springer Science & Business Media BV)*, *18*(1).

**Lévy, J., Bres, C., Geurts, R., Chalhoub, B., Kulikova, O., Duc, G., Journet, E.P., Ané, J.M., Lauber, E., Bisseling, T. and Dénarié, J.,** 2004. A putative Ca2+ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science*, *303*(5662), pp.1361-1364.

**Li, H. L., W. Wang, P.E. Mortimer, R. Q. Li, D. Z. Li, K.D. Hyde, J.C. Xu, et al.** 2015. Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. Scientific Reports 5 : 14023.

**Limpens, E., Moling, S., Hooiveld, G., Pereira, P.A., Bisseling, T., Becker, J.D. and Küster, H.,** 2013. Cell-and tissue-specific transcriptome analyses of Medicago truncatula root nodules. *PloS one*, *8*(5), p.e64377.

**Liu, Q. and Berry, A.M.,** 1991. The infection process and nodule initiation in the Frankia-Ceanothus root nodule symbiosis. *Protoplasma*, *163*(2), pp.82-92.

**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*(1), p.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. *Nature communications*, *1*, p.10.

**Maillet F, Poinsot V, André O, Puech-Pagès V, Haouy A, Gueunier M, Cromer L *et al*. 2011.** Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature*, **469(7328):** 58-63.

**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 biology*, *6*(3), p.e68.

**Markmann, K. and Parniske, M.,** 2009. Evolution of root endosymbiosis with bacteria: How novel are nodules?. *Trends in plant science*, *14*(2), pp.77-86.

**Mbengue, M., Camut, S., de Carvalho-Niebel, F., Deslandes, L., Froidure, S., Klaus-Heisen, D., Moreau, S., Rivas, S., Timmers, T., Hervé, C. and Cullimore, J.,** 2010. The Medicago truncatula E3 ubiquitin ligase PUB1 interacts with the LYK3 symbiotic receptor and negatively regulates infection and nodulation. *The Plant Cell*, *22*(10), pp.3474-3488.

**Miller, I.M. and Baker, D.D.,** 1985. The initiation, development and structure of root nodules in *Elaeagnus angustifolia* L.(Elaeagnaceae). *Protoplasma*, *128*(2-3), pp.107-119.

**Murray, J.D., Muni, R.R.D., Torres‐Jerez, I., Tang, Y., Allen, S., Andriankaja, M., Li, G., Laxmi, A., Cheng, X., Wen, J. and Vaughan, D.,** 2011. Vapyrin, a gene essential for intracellular progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of Medicago truncatula. *The Plant Journal*, *65*(2), pp.244-252.

**Nakagawa, T., Kaku, H., Shimoda, Y., Sugiyama, A., Shimamura, M., Takanashi, K., Yazaki, K., Aoki, T., Shibuya, N. and Kouchi, H.,** 2011. From defense to symbiosis: limited alterations in the kinase domain of LysM receptor‐like kinases are crucial for evolution of legume–Rhizobium symbiosis. *The Plant Journal*, *65*(2), pp.169-180.

**Navarro, E., Nalin, R., Gauthier, D. and Normand, P.,** 1997. The nodular microsymbionts of Gymnostoma spp. are Elaeagnus-infective Frankia strains. *Applied and environmental microbiology*, *63*(4), pp.1610-1616.

**Oldroyd GE. 2013.** Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews Microbiology* **11(4):** 252-263.

**Op den Camp R, Streng A, De Mita S, Cao Q, Polone E, Lie W, Ammiraju JSS *et al.* 2011.** LysM-type mycorrhizal receptor recruited for Rhizobium symbiosis in nonlegume Parasponia. *Science* **331:** 909–912.

**Pawlowski K, Sprent JI. 2008.** Comparison between actinorhizal and legume symbiosis. In *Nitrogen-fixing actinorhizal symbioses* (pp. 261-288). Springer Netherlands.

**Pawlowski K, Bogusz D, Ribeiro A, Berry AM.** 2011. Progress on research on actinorhizal plants. *Functional Plant Biology* **38(9):** 633-638.

**Pawlowski, K. and Demchenko, K.N.**, 2012. The diversity of actinorhizal symbiosis. *Protoplasma*, *249*(4), pp.967-979.

**Racette, S. and Torrey, J.G.,** 1989. Root nodule initiation in Gymnostoma (Casuarinaceae) and Shepherdia (Elaeagnaceae) induced by Frankia strain HFPGpI1. *Canadian Journal of Botany*, *67*(10), pp.2873-2879.

**Rivers, R.L., Dean, R.M., Chandy, G., Hall, J.E., Roberts, D.M. and Zeidel, M.L.,** 1997. Functional analysis of nodulin 26, an aquaporin in soybean root nodule symbiosomes. *Journal of Biological Chemistry*, *272*(26), pp.16256-16261.

**Robinson, MD, and Smyth, GK** (2008). Small sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* 9, 321–332.

**Saito, K., Yoshikawa, M., Yano, K., Miwa, H., Uchida, H., Asamizu, E., Sato, S., Tabata, S., Imaizumi-Anraku, H., Umehara, Y. and Kouchi, H.,** 2007. NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in Lotus japonicus. *The Plant Cell*, *19*(2), pp.610-624.

**Schaller, A., Stintzi, A. and Graff, L.,** 2012. Subtilases–versatile tools for protein turnover, plant development, and interactions with the environment. *Physiologia Plantarum*, *145*(1), pp.52-66.

**Schauser, L., Roussis, A., Stiller, J. and Stougaard, J.,** 1999. A plant regulator controlling development of symbiotic root nodules. Nature, 402(6758), p.191.

**Schauser, L., Wieloch, W. and Stougaard, J.,** 2005. Evolution of NIN-like proteins in Arabidopsis, rice, and Lotus japonicus. *Journal of molecular evolution*, *60*(2), pp.229-237.

**Singh, S., Katzer, K., Lambert, J., Cerri, M. and Parniske, M., 2014.** CYCLOPS, a DNA-binding transcriptional activator, orchestrates symbiotic root nodule development. *Cell Host & Microbe*, *15*(2), pp.139-152.

**Sinharoy, S., Saha, S., Chaudhury, S.R. and DasGupta, M., 2009**. Transformed hairy roots of Arachis hypogea: A tool for studying root nodule symbiosis in a non–infection thread legume of the Aeschynomeneae tribe. *Molecular plant-microbe interactions*, *22*(2), pp.132-142.

**Smil, V.,** 1999. Nitrogen in crop production: An account of global flows. *Global biogeochemical cycles*, *13*(2), pp.647-662.

**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. *Proceedings of the National Academy of Sciences*, *92*(7), pp.2647-2651.

**Soyano, T., Kouchi, H., Hirota, A. and Hayashi, M.,** 2013. Nodule inception directly targets NF-Y subunit genes to regulate essential processes of root nodule development in Lotus japonicus. *PLoS Genetics*, *9*(3), p.e1003352.

**Soyano, T., and M. Hayashi.** 2014 . Transcriptional networks leading to symbiotic nodule organogenesis. *Current Opinion in Plant Biology* 20: 146 – 154.

**Sprent JI.** 2001. Nodulation in legumes. London: Royal Botanic Gardens Kew.

**Stamatakis, A.,** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*(21), pp.2688-2690.

**Streng, A., op den Camp, R., Bisseling, T. and Geurts, R.,** 2011. Evolutionary origin of rhizobium Nod factor signaling. *Plant signaling & behavior,* 6(10), pp.1510-1514.

**Svistoonoff, S., Laplaze, L., Auguy, F., Runions, J., Duponnois, R., Haseloff, J., Franche, C. and Bogusz, D.,** 2003. cg12 expression is specifically linked to infection of root hairs and cortical cells during Casuarina glauca and Allocasuarina verticillata actinorhizal nodule development. *Molecular plant-microbe interactions*, *16*(7), pp.600-607.

**Svistoonoff, S., Laplaze, L., Liang, J., Ribeiro, A., Gouveia, M.C., Auguy, F., Fevereiro, P., Franche, C. and Bogusz, D.,** 2004. Infection-related activation of the cg12 promoter is conserved between actinorhizal and legume-rhizobia root nodule symbiosis. *Plant physiology*, *136*(2), pp.3191-3197.

**Svistoonoff, S., Benabdoun, F.M., Nambiar-Veetil, M., Imanishi, L., Vaissayre, V., Cesari, S., Diagne, N., Hocher, V., De Billy, F., Bonneau, J. and Wall, L.,** 2013. The independent acquisition of plant root nitrogen-fixing symbiosis in Fabids recruited the same genetic pathway for nodule organogenesis. *PLoS One*, *8*(5), p.e64515.

**Svistoonoff, S., Hocher, V. and Gherbi, H.,** 2014. Actinorhizal root nodule symbioses: what is signalling telling on the origins of nodulation?. *Current opinion in plant biology*, *20*, pp.11-18.

**Svistoonoff, S., Tromas, A., Diagne, N., Alloisio, N., Laplaze, L., Champion, A., Bonneau, J., Bogusz, D. and Hocher, V.,** 2015. How Transcriptomics Revealed New Information on Actinorhizal Symbioses Establishment and Evolution. *Biological Nitrogen Fixation*, pp.425-432.

**Swensen, S.M.,** 1996. The evolution of actinorhizal symbioses: evidence for multiple origins of the symbiotic association. *American Journal of Botany*, pp.1503-1512.

**Swensen, S.M. and Benson, D.R.,** 2007. Evolution of actinorhizal host plants and Frankia endosymbionts. In *Nitrogen-fixing actinorhizal symbioses* (pp. 73-104). Springer, Dordrecht.

**Takeda, N., Sato, S., Asamizu, E., Tabata, S. and Parniske, M.,** 2009. Apoplastic plant subtilases support arbuscular mycorrhiza development in Lotus japonicus. *The Plant Journal*, *58*(5), pp.766-777.

**Taylor, A. and Qiu, Y.L.,** 2017. Evolutionary history of subtilases in land plants and their involvement in symbiotic interactions. *Molecular Plant-Microbe Interactions*, *30*(6), pp.489-501.

**Timmers, A.C., Auriac, M.C. and Truchet, G.,** 1999. Refined analysis of early symbiotic steps of the Rhizobium-Medicago interaction in relationship with microtubular cytoskeleton rearrangements. *Development*, *126*(16), pp.3617-3628.

**Tornero, P., Conejero, V. and Vera, P.,** 1996. Primary structure and expression of a pathogen-induced protease (PR-P69) in tomato plants: similarity of functional domains to subtilisin-like endoproteases. *Proceedings of the National Academy of Sciences*, *93*(13), pp.6332-6337.

**Valverde, C. and Wall, L.G.,** 1999. Time course of nodule development in the *Discaria trinervis* (Rhamnaceae)–Frankia symbiosis. *The New Phytologist*, *141*(2), pp.345-354.

**van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., Santuari, L., Cao, Q., Sharma, T., Shen, D. and Roswanjaya, Y.,** 2018. Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proceedings of the National Academy of Sciences*, p.201721395.

**Vázquez-Limón, C., Hoogewijs, D., Vinogradov, S.N. and Arredondo-Peter, R.,** 2012. The evolution of land plant hemoglobins. *Plant science*, *191*, pp.71-81.

**Wall, L.G.,** 2000. The actinorhizal symbiosis. *Journal of plant growth regulation*, *19*(2), pp.167-182.

**Wall, L.G. and Berry, A.M.,** 2007. Early interactions, infection and nodulation in actinorhizal symbiosis. In *Nitrogen-fixing actinorhizal symbioses* (pp. 147-166). Springer, Dordrecht.

**Wang B, Yeun LH, Xue JY, Yang L, Ane JM, Qiu YL. 2010.** Presence of three mycorrhizal genes in the common ancestor of land plants suggests a key role of mycorrhizas in the colonization of land by plants. *New Phytologist* **186:** 514–525.

**Zhang, Q., Blaylock, L.A. and Harrison, M.J.,** 2010. Two Medicago truncatula half-ABC transporters are essential for arbuscule development in arbuscular mycorrhizal symbiosis. *The Plant Cell*, *22*(5), pp.1483-1497.

Table 4.1. Expression profiles of *E. umbellata* transcripts that are significantly differentially expressed in our pooled timepoints exact test and their closest homolog in *Arabidopsis thaliana.*

| Transcript ID | Fold Change | P-Value | False Discovery Rate | Arabidopsis best hit | Protein Name |
|---|---|---|---|---|---|
| DN188540_c2_g1 | -14.55 | 3.55E-20 | 2.32E-17 | AT4G08150.1 | Homeobox domain containing protein |
| DN192889_c1_g2 | -11.77 | 4.57E-20 | 2.88E-17 | AT1G61800.1 | phosphate/phosphate translocator |
| DN195723_c0_g2 | -13.97 | 5.67E-20 | 3.38E-17 | AT4G38400.1 | expansin |
| DN191329_c1_g4 | 10.59 | 1.56E-18 | 8.71E-16 | AT4G32300.1 | lectin protein kinase family protein |
| DN196288_c1_g1 | -12.44 | 2.84E-18 | 1.49E-15 | AT2G37130.1 | peroxidase |
| DN192452_c0_g1 | 11.06 | 2.78E-17 | 1.32E-14 | AT3G28860.1 | multidrug resistance protein |
| DN193539_c1_g2 | -9.29 | 1.36E-13 | 4.22E-11 | AT3G51895.1 | sulfate transporter |
| DN192314_c4_g2 | -6.58 | 2.91E-12 | 7.65E-10 | AT5G05340.1 | peroxidase precursor |
| DN191992_c1_g1 | -6.22 | 1.75E-11 | 4.28E-09 | AT1G48100.1 | polygalacturonase |
| DN185850_c1_g3 | -5.43 | 9.27E-11 | 2.10E-08 | AT3G53960.1 | peptide transporter PTR2 |
| DN194615_c6_g1 | -8.50 | 1.50E-10 | 3.30E-08 | AT4G21760.1 | monolignol beta-glucoside |
| DN192205_c0_g2 | -5.22 | 1.58E-10 | 3.40E-08 | AT4G18910.1 | aquaporin |
| DN195696_c1_g2 | -5.58 | 2.60E-10 | 5.35E-08 | AT2G19070.1 | transferase family protein |
| DN189437_c2_g4 | 5.75 | 1.97E-09 | 3.62E-07 | AT3G18180.1 | glycosyltransferase |
| DN195410_c2_g4 | -3.91 | 1.94E-08 | 2.80E-06 | AT2G01770.1 | integral membrane protein |
| DN187519_c4_g6 | -5.95 | 2.30E-08 | 3.28E-06 | AT4G27440.1 | oxidoreductase, short chain dehydrogenase/reductase |
| DN192580_c1_g1 | -3.85 | 2.77E-08 | 3.88E-06 | AT3G61490.1 | polygalacturonase |
| DN189573_c6_g3 | -4.42 | 1.14E-07 | 1.41E-05 | AT1G08290.1 | C2H2 zinc finger protein |
| DN184020_c0_g1 | -6.16 | 1.35E-07 | 1.65E-05 | AT2G29380.1 | protein phosphatase 2C |
| DN193565_c3_g7 | -3.55 | 1.38E-07 | 1.67E-05 | AT5G52860.1 | ABC-2 type transporter |
| DN196328_c2_g1 | -3.41 | 1.42E-07 | 1.71E-05 | AT1G22710.1 | sucrose transporter |
| DN191442_c4_g1 | -3.34 | 1.56E-07 | 1.82E-05 | AT5G24930.1 | CCT/B-box zinc finger protein |
| DN193716_c1_g1 | -3.31 | 1.63E-07 | 1.91E-05 | AT1G69780.1 | homeobox associated leucine zipper |
| DN187322_c3_g1 | -6.28 | 1.92E-07 | 2.20E-05 | AT1G25530.1 | amino acid transporter |
| DN191444_c7_g2 | -4.04 | 2.68E-07 | 2.93E-05 | AT2G13610.1 | ABC-2 type transporter |
| DN188965_c0_g1 | -3.85 | 4.79E-07 | 5.01E-05 | AT1G23380.2 | Homeobox domain containing protein |

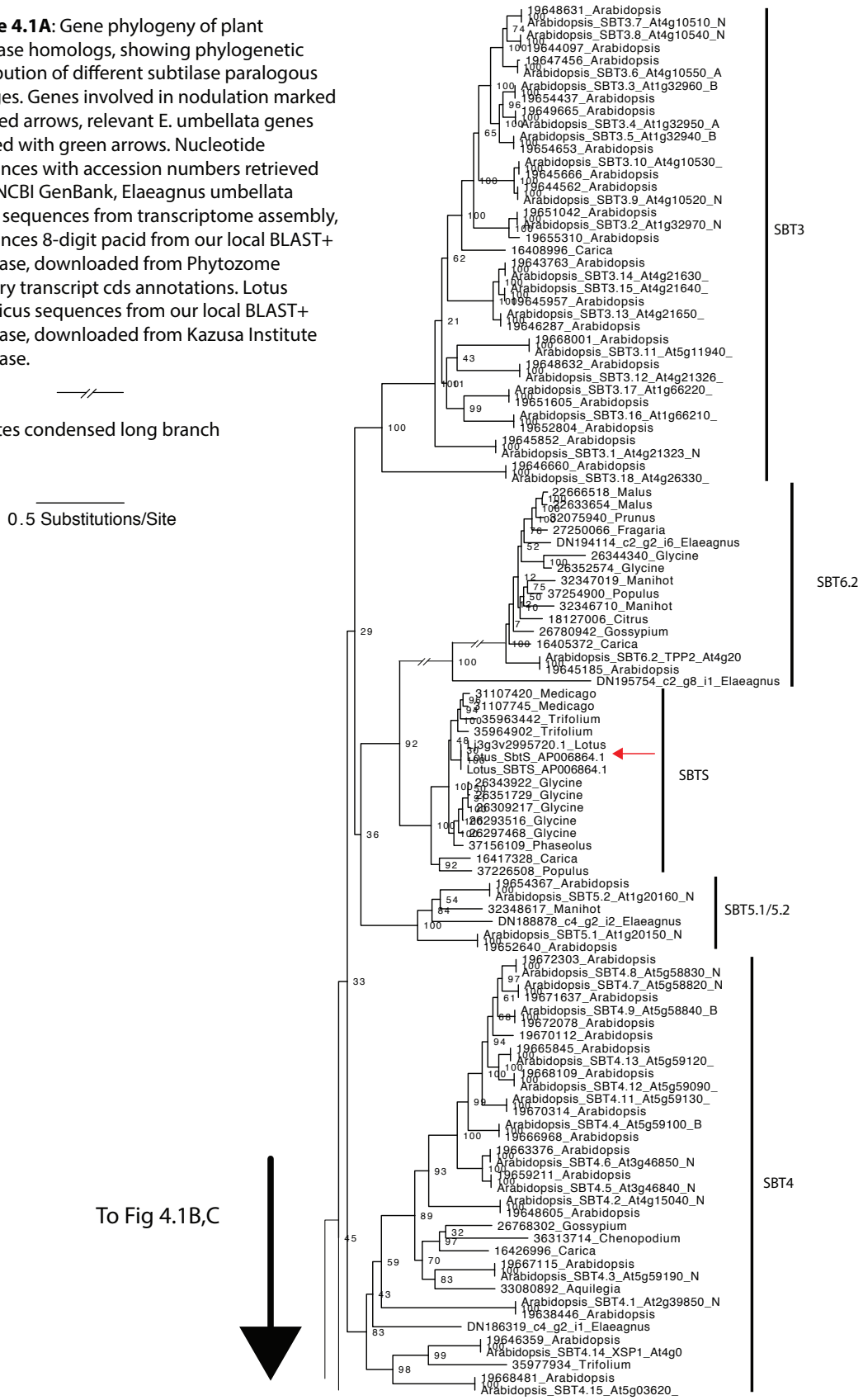
Table 4.2. Expression profiles on *E. umbellata* transcripts homologous to genes involved in nodulation in other lineages in our pooled timepoints exact test

| Homolog | Transcript ID | Fold Change | P-Value | False Discovery Rate |
|---|---|---|---|---|
| NOD26 | DN192205_c0_g2 | -5.22 | 1.58E-10 | 3.40E-08 |
| NSHB1 | DN193242_c0_g2 | -3.02 | 1.96E-06 | 0.0002 |
| CASTOR | DN194038_c0_g1 | 1.01 | 0.0020 | 0.0449 |
| SBTS | DN194728_c0_g2 | -1.02 | 0.0055 | 0.0938 |
| vapyrin | DN190823_c1_g1 | 0.70 | 0.0094 | 0.1349 |
| CYCLOPS | DN197191_c1_g1 | 0.73 | 0.0128 | 0.1663 |
| NFR1 | DN194536_c1_g3 | 0.34 | 0.0683 | 0.4508 |
| NIN | DN195843_c2_g2 | 0.32 | 0.0919 | 0.5157 |
| SYMRK | DN195580_c2_g3 | 0.25 | 0.1302 | 0.5997 |
| CCAMK | DN193510_c1_g1 | 0.19 | 0.1929 | 0.6992 |
| NSP2 | DN189172_c1_g4 | 0.09 | 0.3487 | 0.8398 |
| RPG | DN191061_c0_g1 | 0.19 | 0.3496 | 0.8398 |
| POLLUX | DN189268_c0_g2 | 0.11 | 0.3657 | 0.8486 |
| SBTM1 | DN196250_c1_g1 | -0.03 | 0.6021 | 0.9327 |
| NUP133 | DN192647_c1_g2 | 0.02 | 0.6602 | 0.9448 |
| NUP85 | DN191961_c1_g1 | -0.01 | 0.8534 | 0.9797 |
| NENA | DN185194_c5_g1 | 0.00 | 0.9331 | 0.9935 |

**Figure 4.1A**: Gene phylogeny of plant subtilase homologs, showing phylogenetic distribution of different subtilase paralogous lineages. Genes involved in nodulation marked with red arrows, relevant E. umbellata genes marked with green arrows. Nucleotide sequences with accession numbers retrieved from NCBI GenBank, Elaeagnus umbellata cDNA sequences from transcriptome assembly, sequences 8-digit pacid from our local BLAST+ database, downloaded from Phytozome primary transcript cds annotations. Lotus japonicus sequences from our local BLAST+ database, downloaded from Kazusa Institute database.

Denotes condensed long branch

0.5 Substitutions/Site




To Fig 4.1B,C

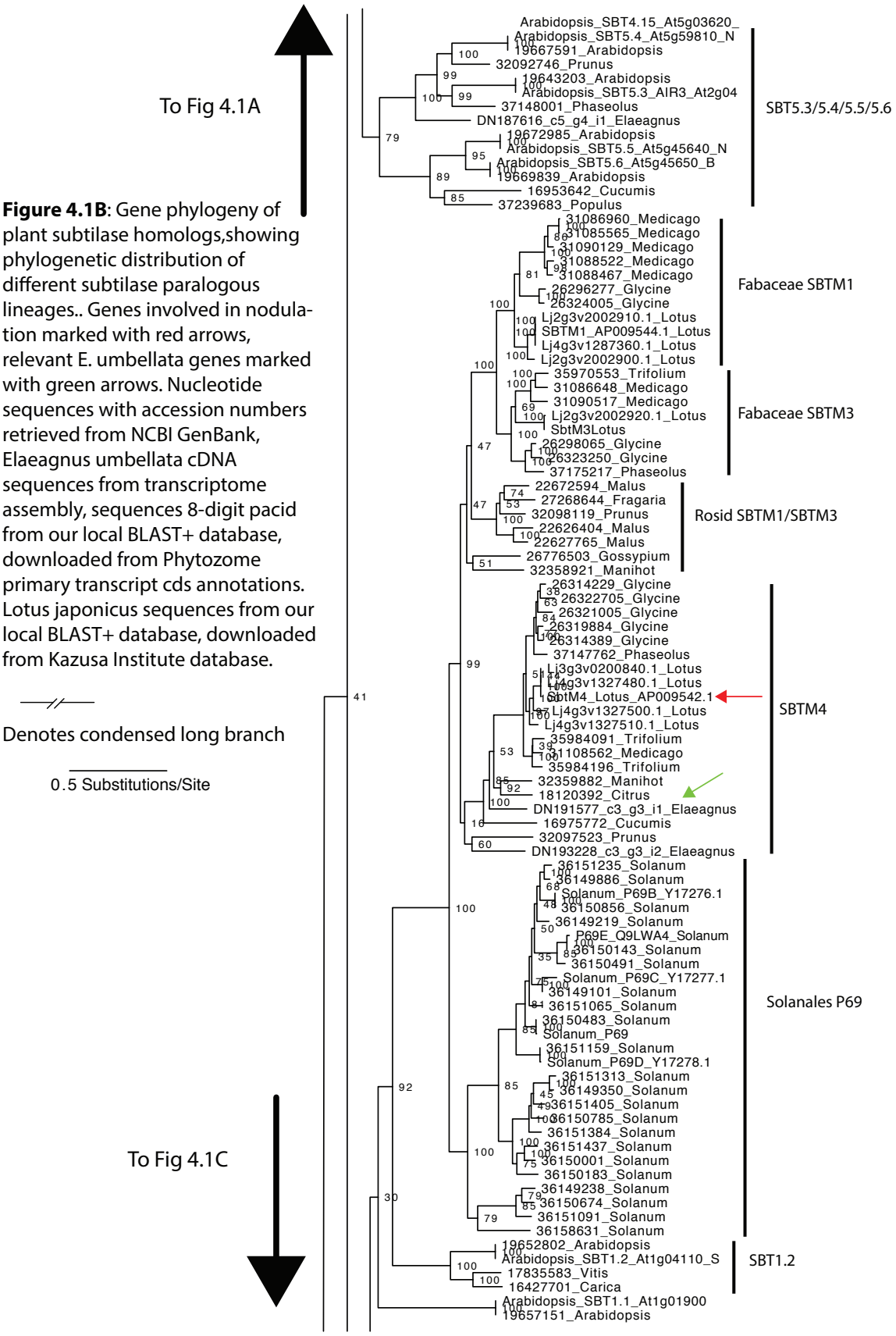



**Figure 4.1B**: Gene phylogeny of plant subtilase homologs,showing phylogenetic distribution of different subtilase paralogous lineages.. Genes involved in nodulation marked with red arrows, relevant E. umbellata genes marked with green arrows. Nucleotide sequences with accession numbers retrieved from NCBI GenBank, Elaeagnus umbellata cDNA sequences from transcriptome assembly, sequences 8-digit pacid from our local BLAST+ database, downloaded from Phytozome primary transcript cds annotations. Lotus japonicus sequences from our local BLAST+ database, downloaded from Kazusa Institute database.

**Figure 4.1C**: Gene phylogeny of plant subtilase homologs, showing phylogenetic distribution of different subtilase paralogous lineages. Genes involved in nodulation marked with red arrows, relevant E. umbellata genes marked with green arrows. Nucleotide sequences with accession numbers retrieved from NCBI GenBank, Elaeagnus umbellata cDNA sequences from transcriptome assembly, sequences 8-digit pacid from our local BLAST+ database, downloaded from Phytozome primary transcript cds annotations. Lotus japonicus sequences from our local BLAST+ database, downloaded from Kazusa Institute database.

Denotes condensed long branch

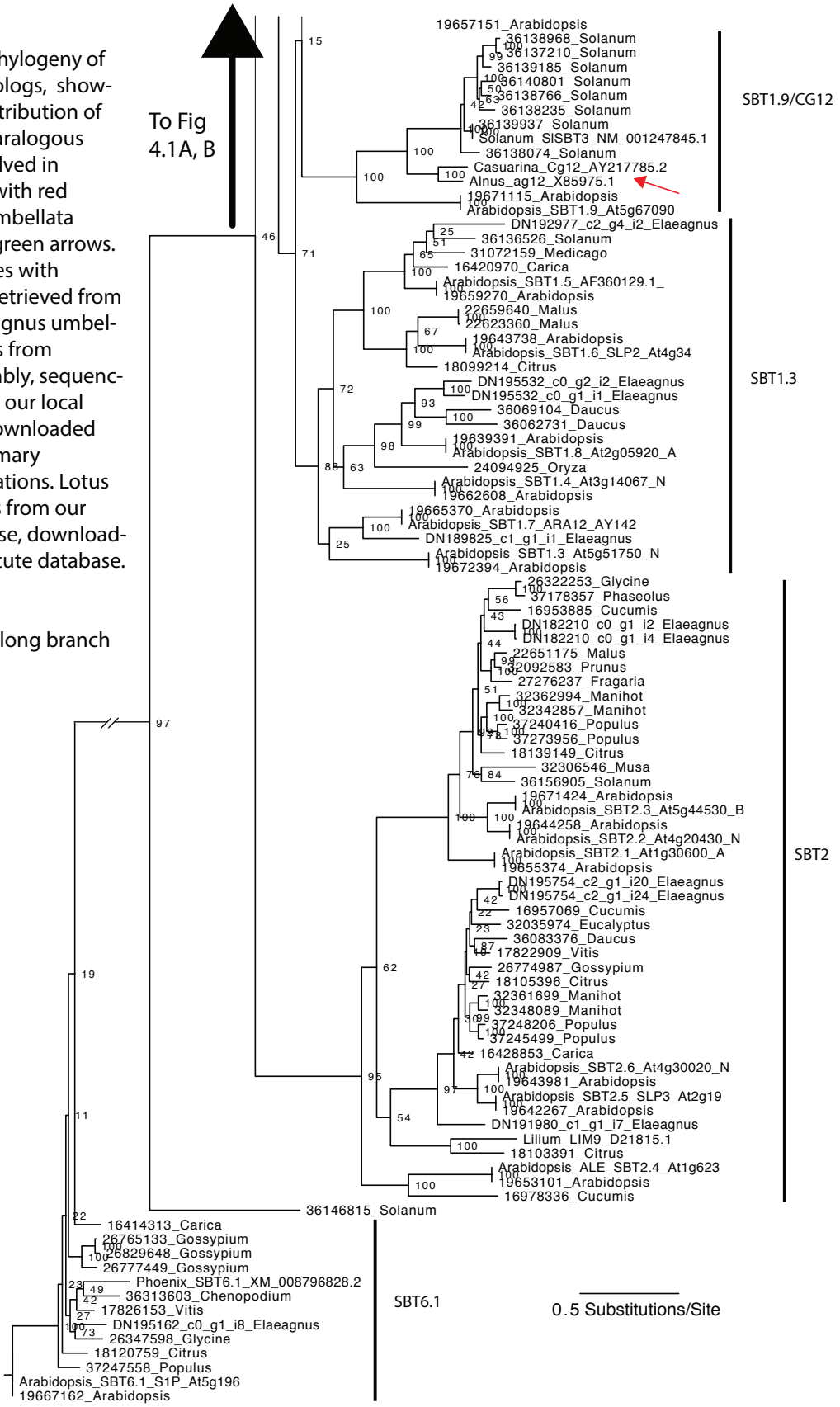
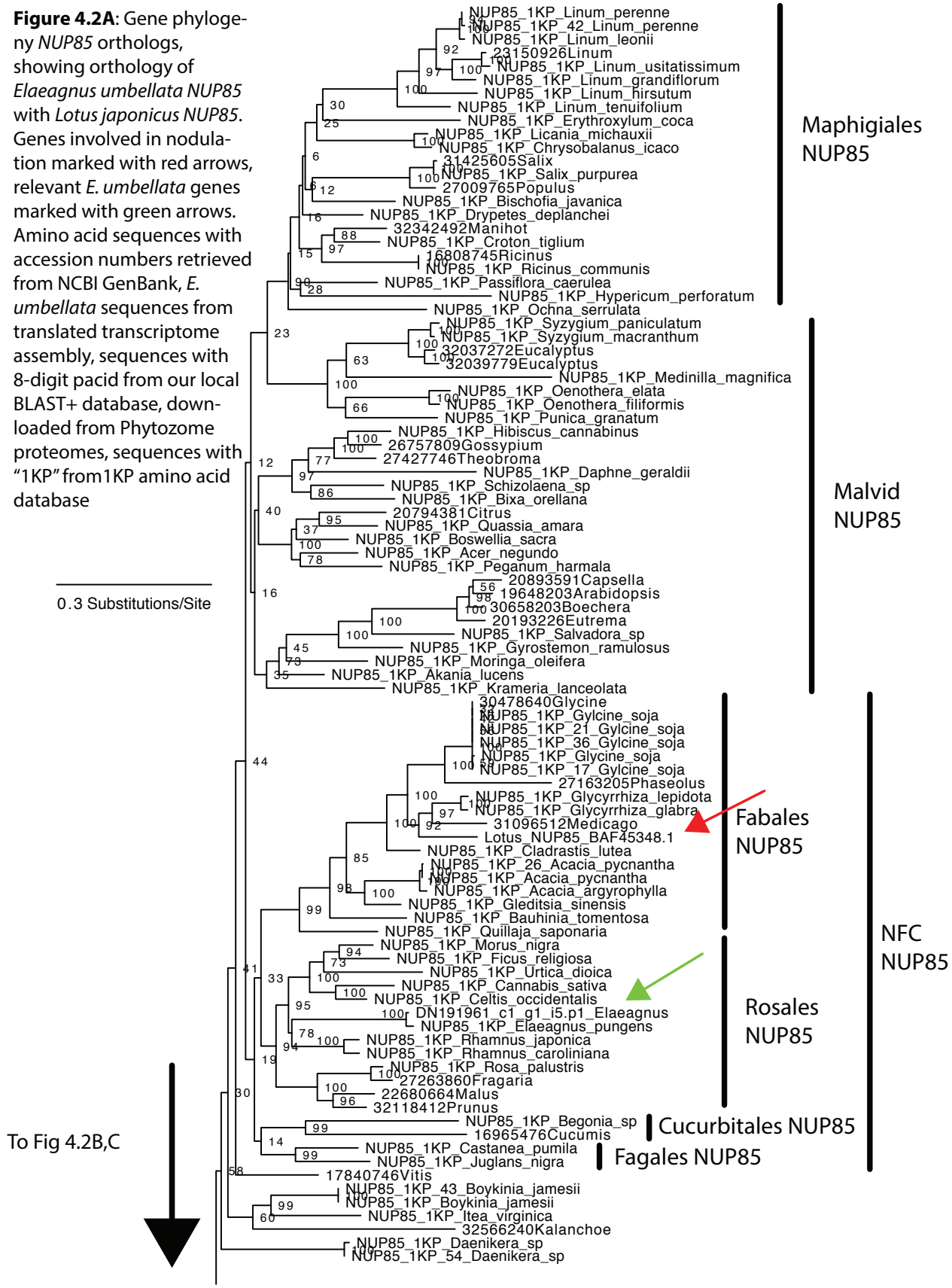


**Figure 4.2A**: Gene phylogeny *NUP85* orthologs, showing orthology of *Elaeagnus umbellata NUP85* with *Lotus japonicus NUP85*. Genes involved in nodulation marked with red arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *E. umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database



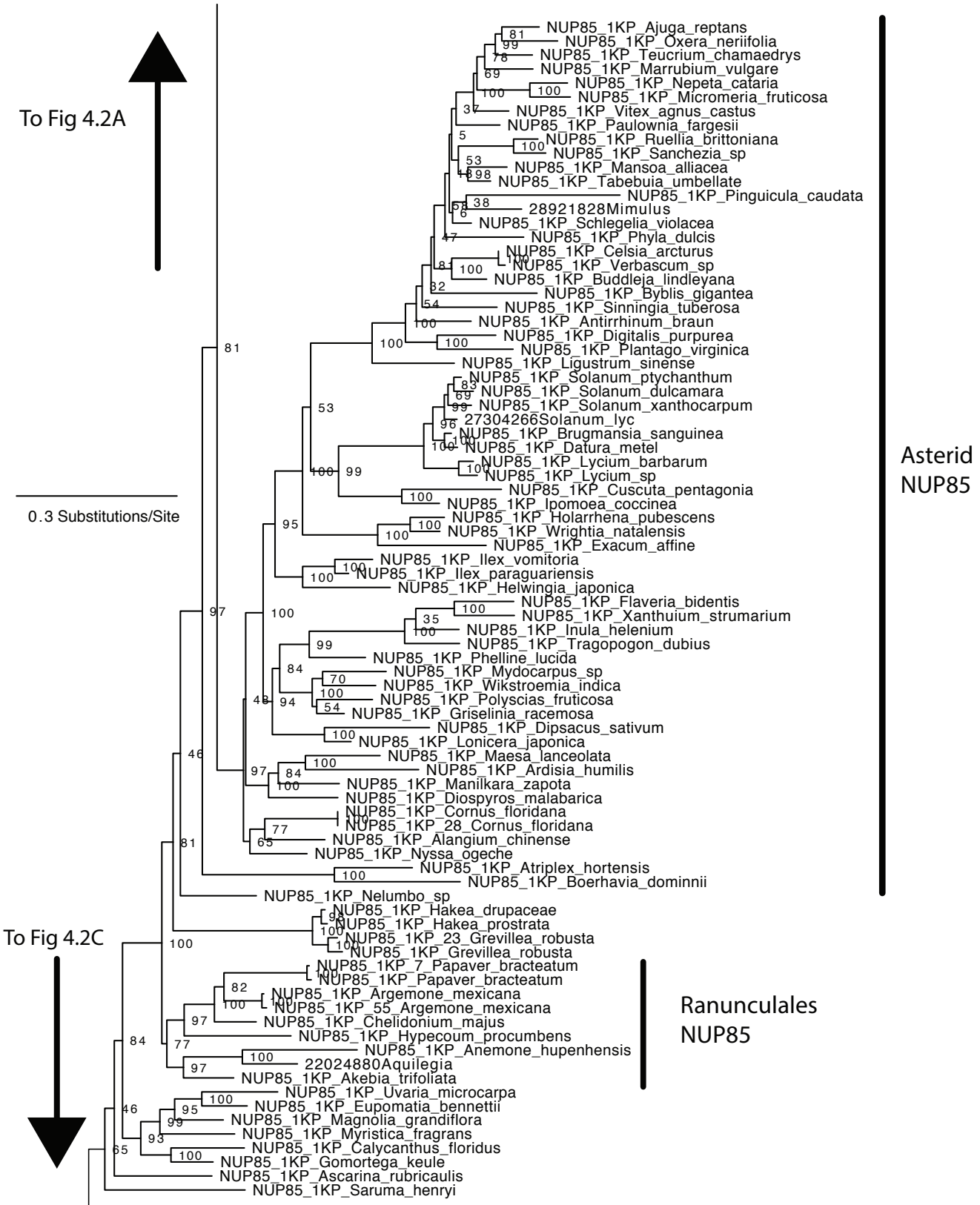



**Figure 4.2B**: Gene phylogeny of land plant *NUP85* orthologs, showing congruence of gene phylogeny with species phylogeny and vertical inheritance. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

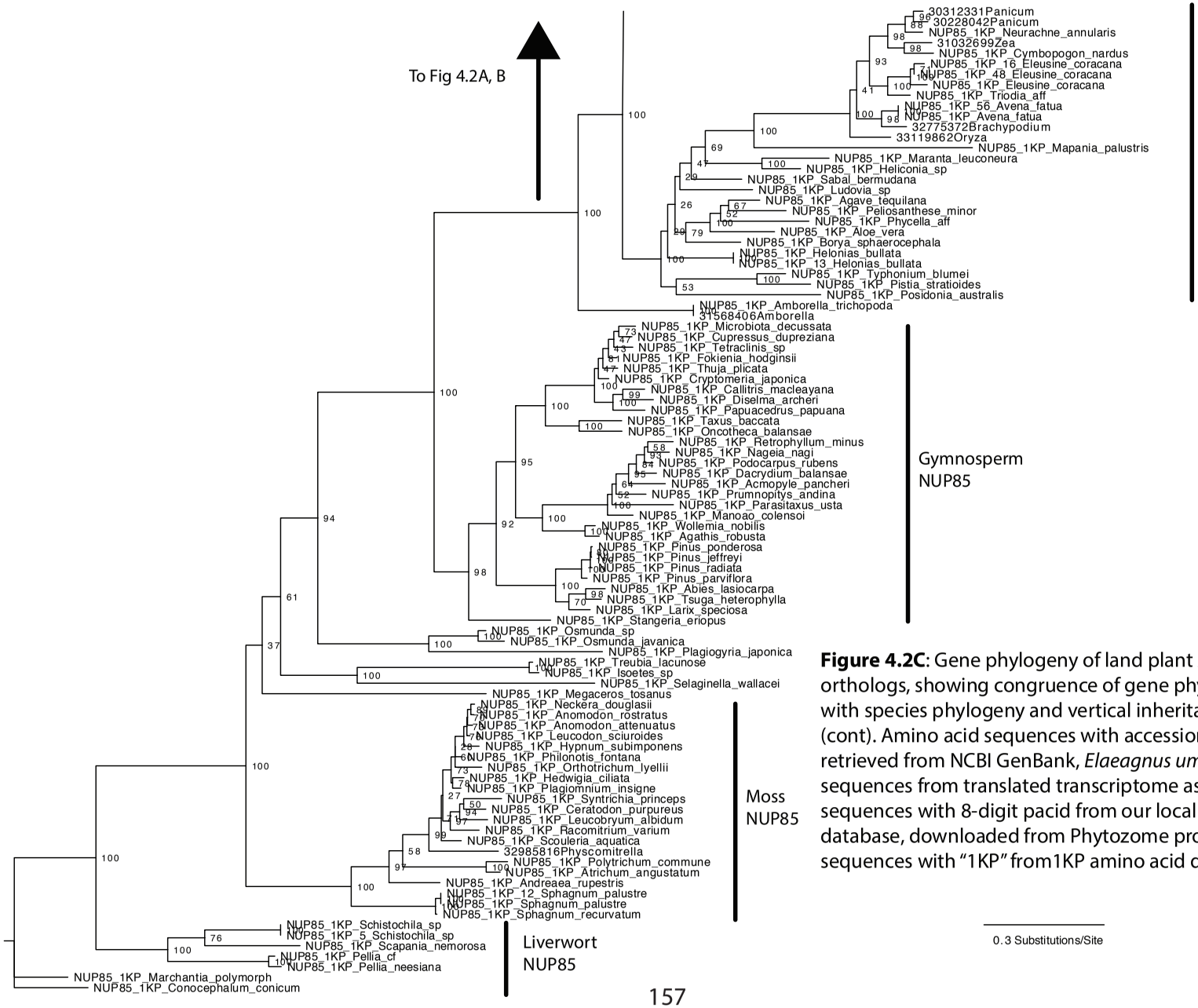
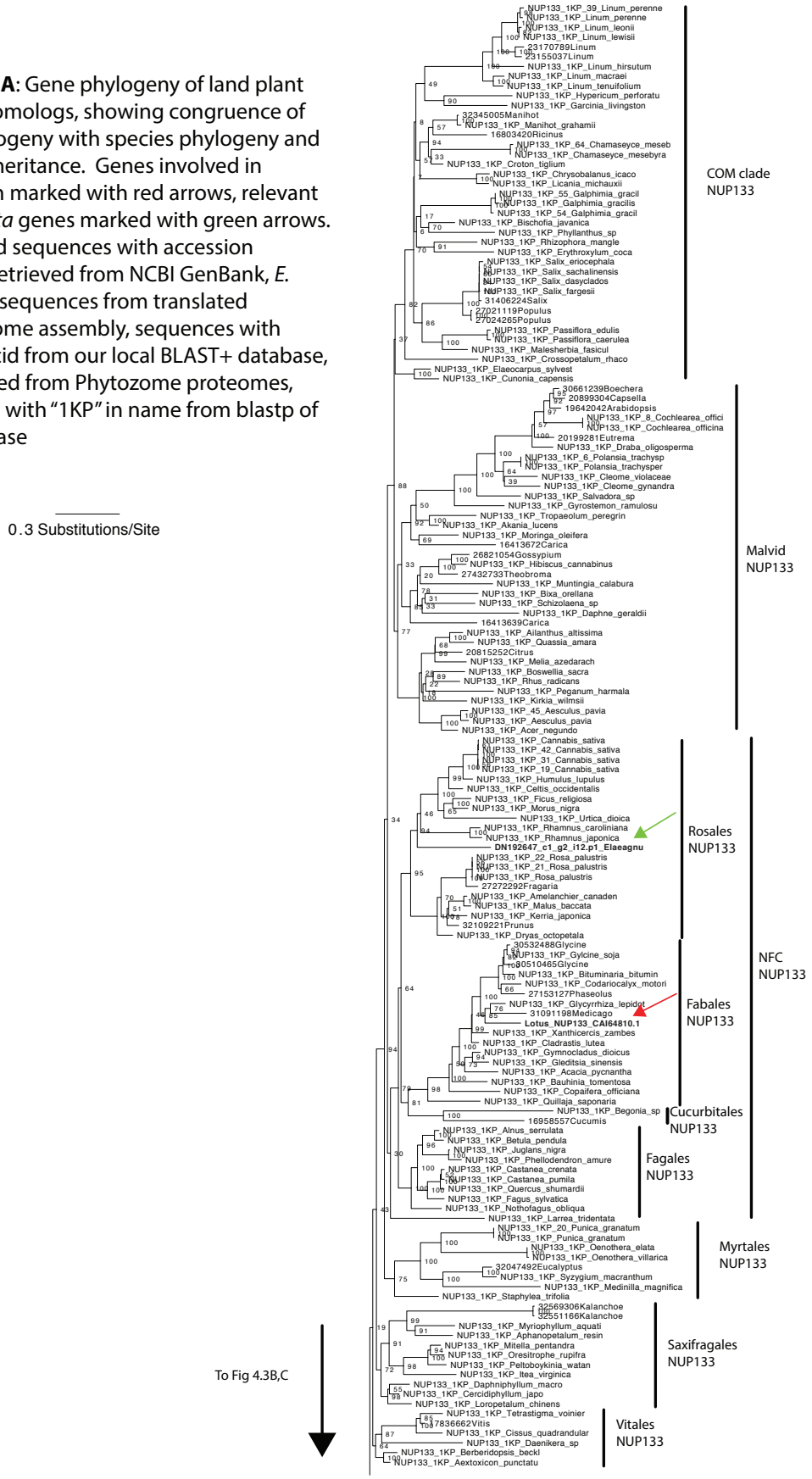



**Figure 4.2C**: Gene phylogeny of land plant *NUP85* orthologs, showing congruence of gene phylogeny with species phylogeny and vertical inheritance (cont). Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

**Figure 4.3A**: Gene phylogeny of land plant *NUP133* homologs, showing congruence of gene phylogeny with species phylogeny and vertical inheritance. Genes involved in nodulation marked with red arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *E. umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" in name from blastp of 1KP database



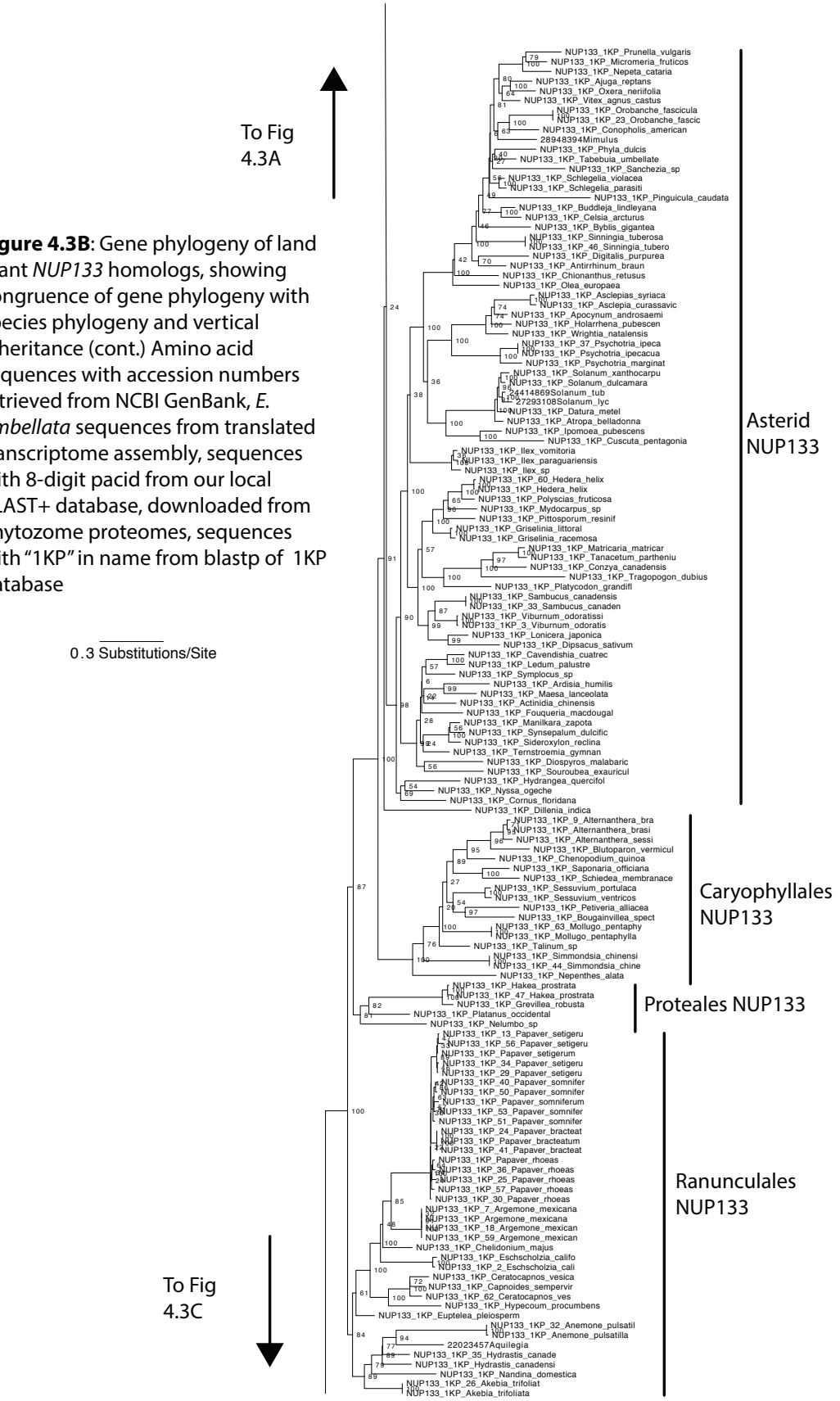



**Figure 4.3B**: Gene phylogeny of land plant *NUP133* homologs, showing congruence of gene phylogeny with species phylogeny and vertical inheritance (cont.) Amino acid sequences with accession numbers retrieved from NCBI GenBank, *E. umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" in name from blastp of 1KP database

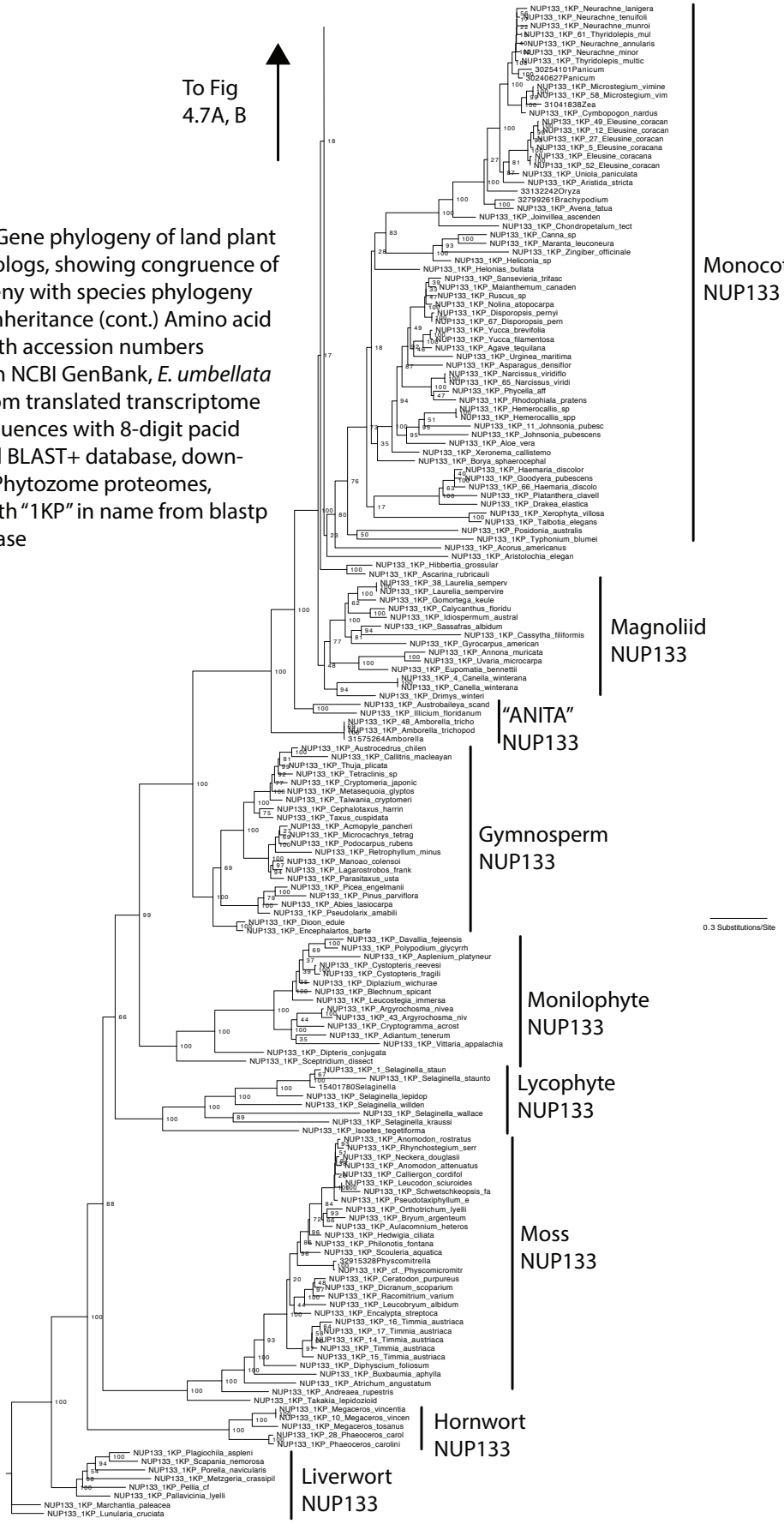
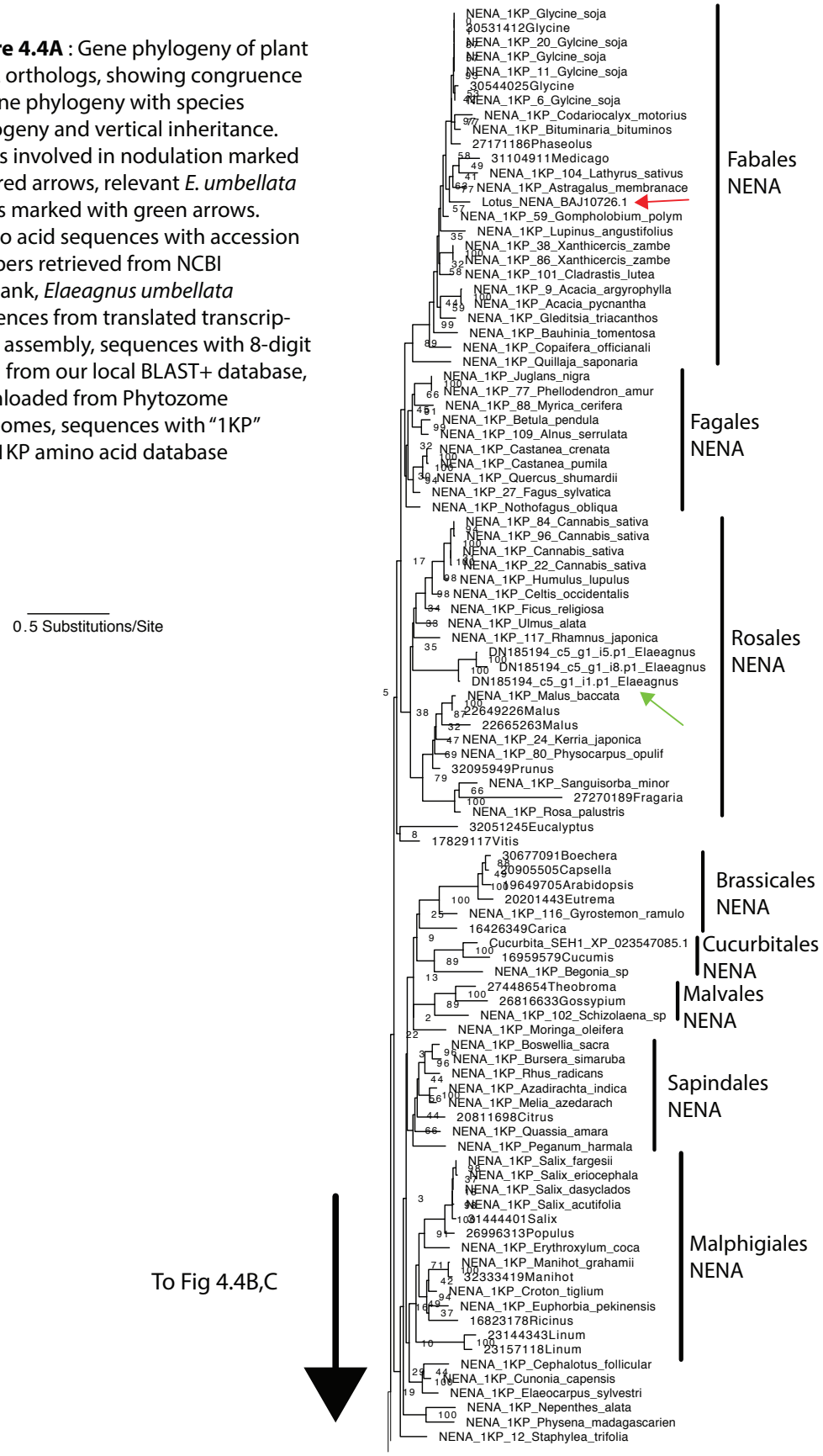



**Figure 4.3C**: Gene phylogeny of land plant *NUP133* homologs, showing congruence of gene phylogeny with species phylogeny and vertical inheritance (cont.) Amino acid sequences with accession numbers retrieved from NCBI GenBank, *E. umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" in name from blastp of 1KP database

**Figure 4.4A** : Gene phylogeny of plant *NENA* orthologs, showing congruence of gene phylogeny with species phylogeny and vertical inheritance. Genes involved in nodulation marked with red arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database



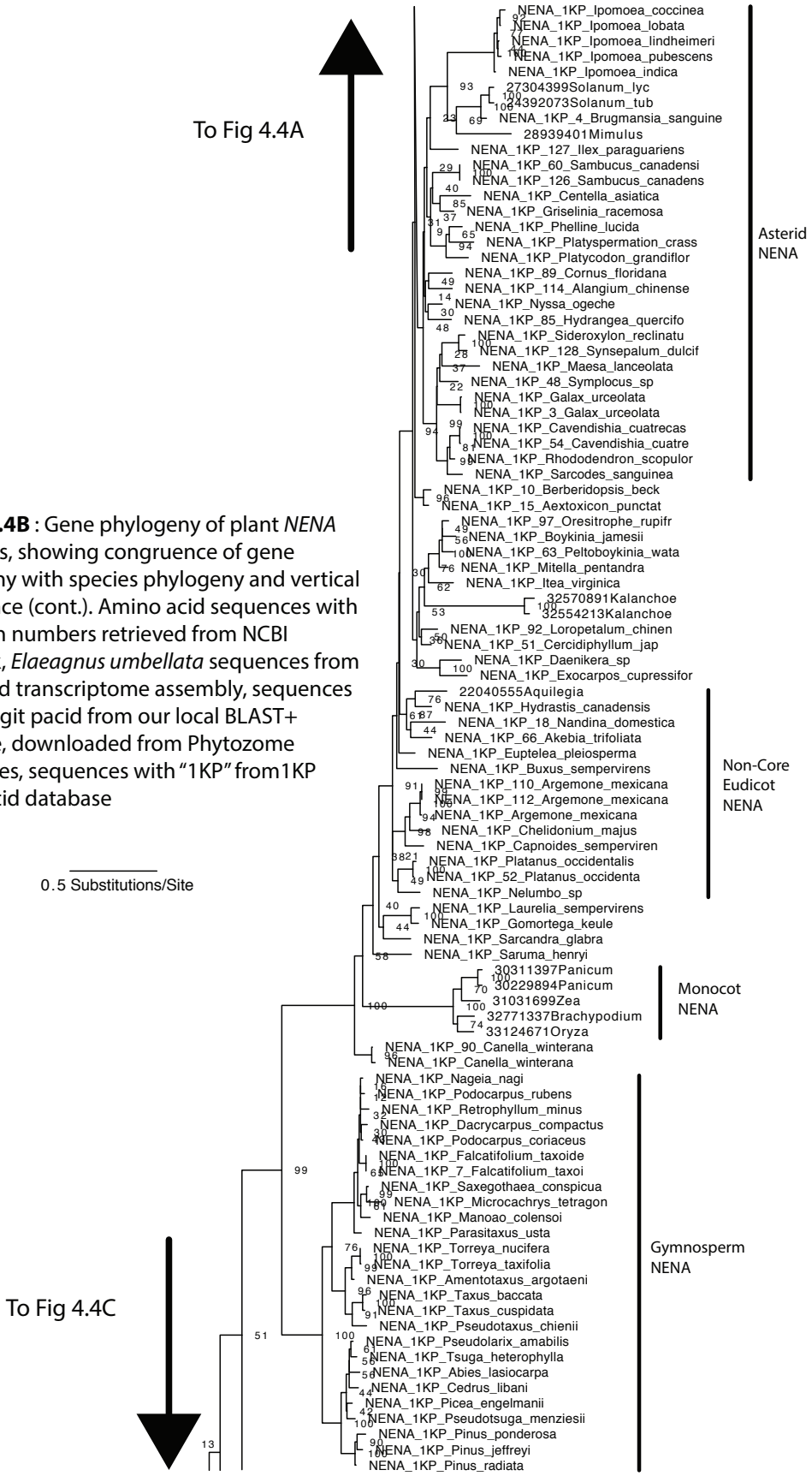



**Figure 4.4B** : Gene phylogeny of plant *NENA* orthologs, showing congruence of gene phylogeny with species phylogeny and vertical inheritance (cont.). Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

**Figure 4.4C** : Gene phylogeny of plant *NENA* orthologs, showing congruence of gene phylogeny with species phylogeny and vertical inheritance (cont.). Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

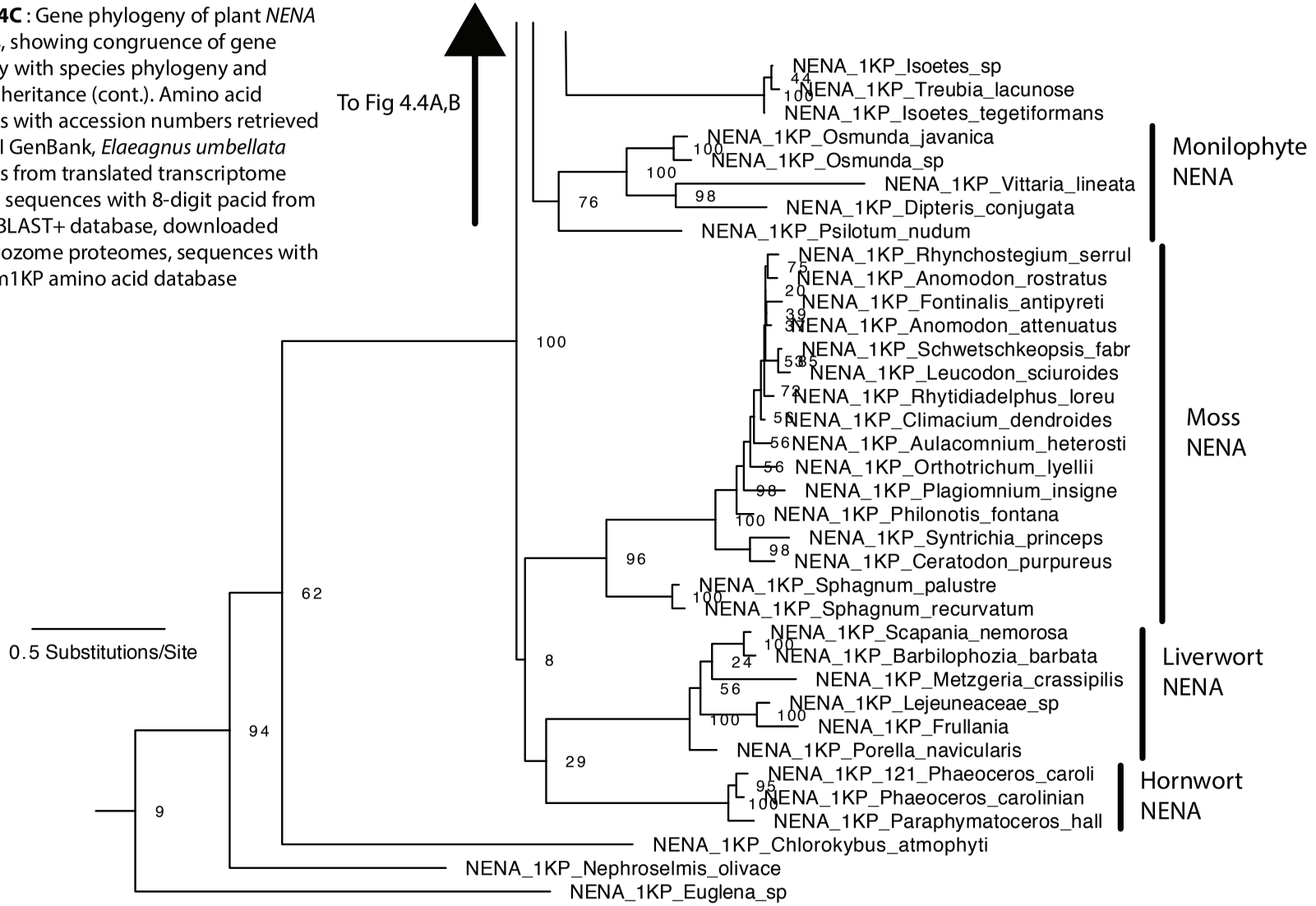


**Figure 4.5A**: Gene phylogeny of plant LysM-RK homologs, showing phylogenetic distribution of different LysM-RK paralogous lineages. Genes involved in nodulation marked with red arrows, genes defining clades marked with blue arrows, relevant *E. umbellata* genes marked with green arrows. Nucleotide sequences with accession numbers retrieved from NCBI GenBank, Elaeagnus umbellata cDNA sequences from transcriptome assembly, sequences 8-digit pacid from our local BLAST+ database, downloaded from Phytozome primary transcript cds annotations. Lotus japonicus sequences from our local BLAST+ database, downloaded from Kazusa Institute database.

Denotes condensed long branch

0.4 Substitutions/Site

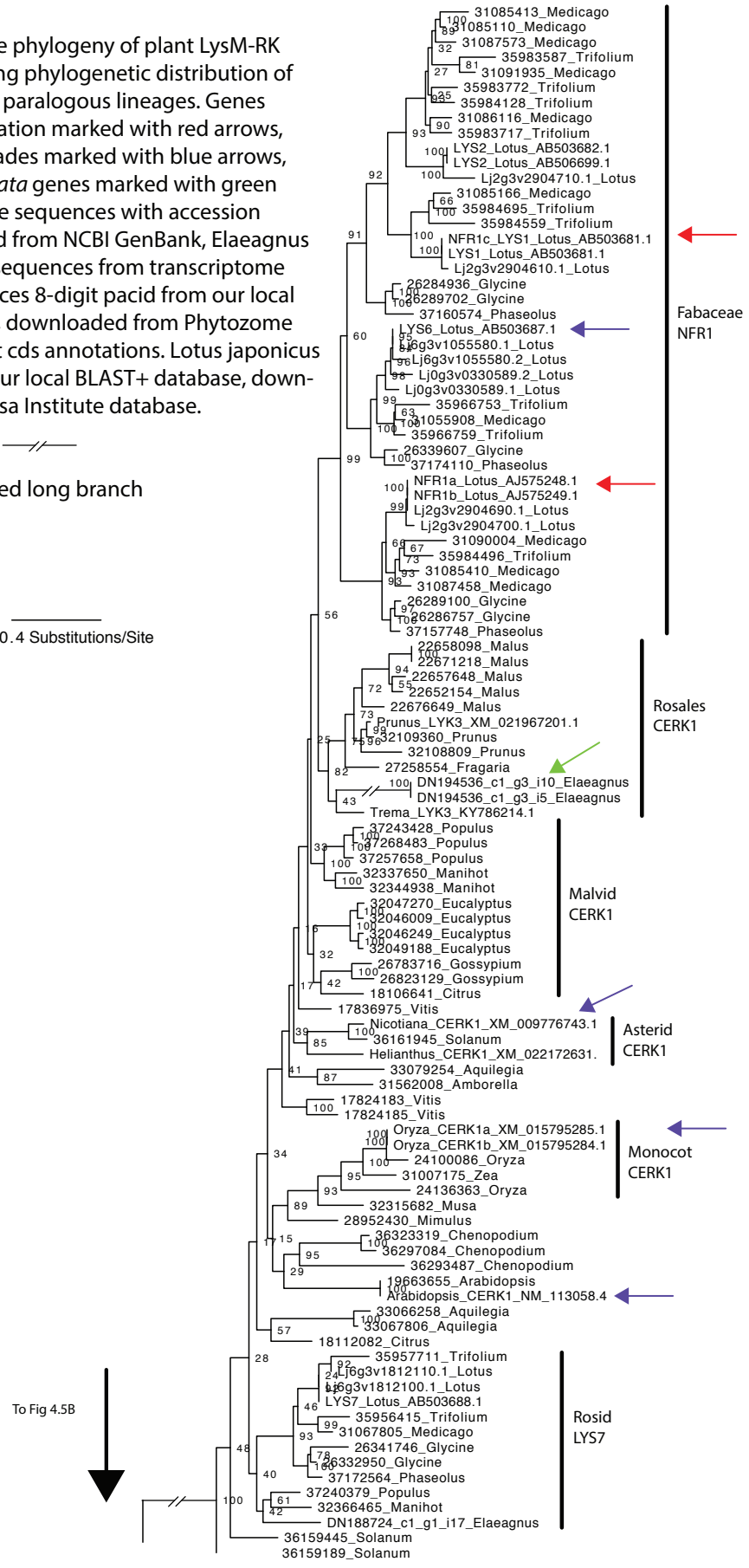


**Figure 4.5B:** Gene phylogeny of plant LysM-RK homologs, showing phylogenetic distribution of different LysM-RK paralogous lineages. Genes involved in nodulation marked with red arrows, genes defining clades marked with blue arrows, relevant *E. umbellata* genes marked with green arrows. Nucleotide sequences with accession numbers retrieved from NCBI GenBank, Elaeagnus umbellata cDNA sequences from transcriptome assembly, sequences 8-digit pacid from our local BLAST+ database, downloaded from Phytozome primary transcript cds annotations. Lotus japonicus sequences from our local BLAST+ database, downloaded from Kazusa Institute database.

Denotes condensed long branch

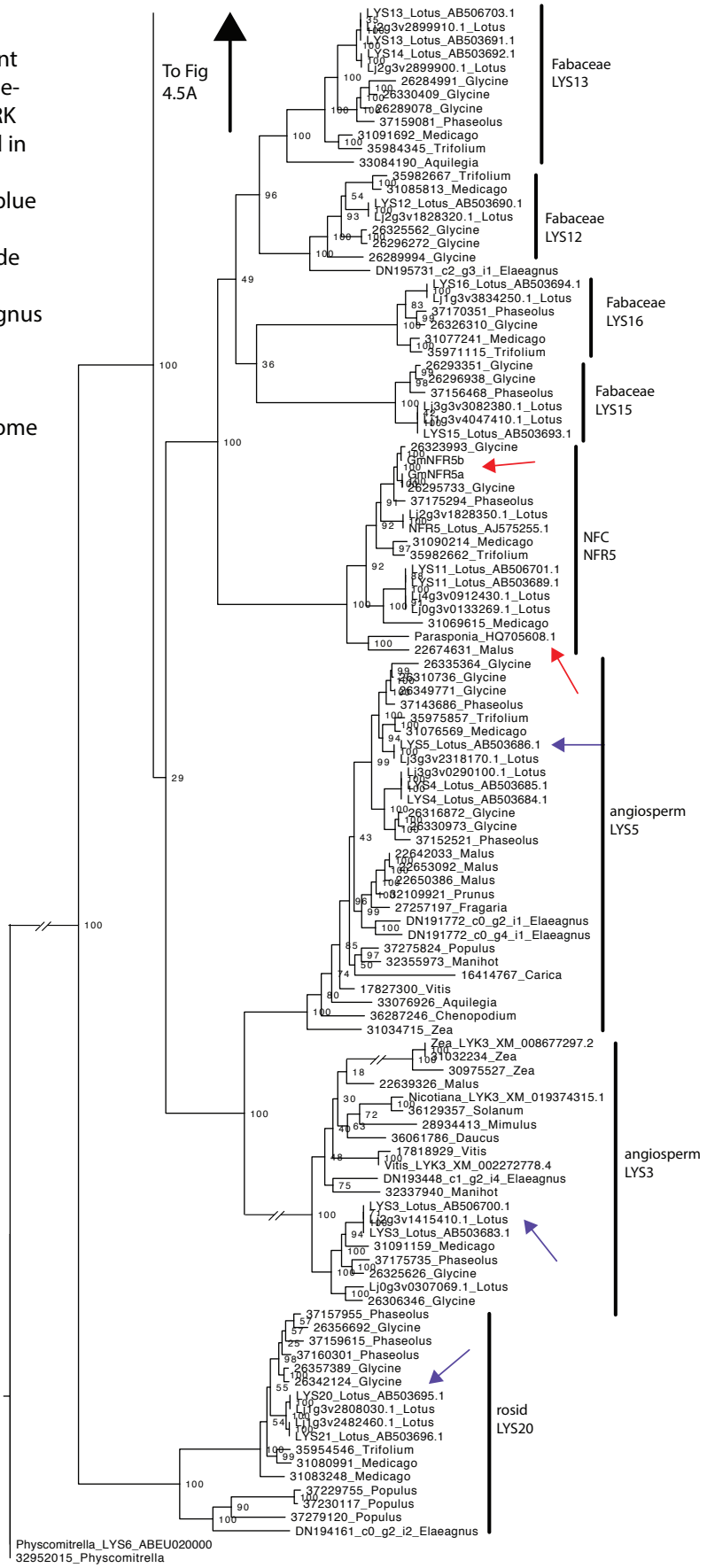


**Figure 4.6A** : Gene phylogeny of plant *SYMRK* orthologs, showing congruence of gene phylogeny with species phylogeny and vertical inheritance. Genes involved in nodulation marked with red arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

Denotes condensed long branch

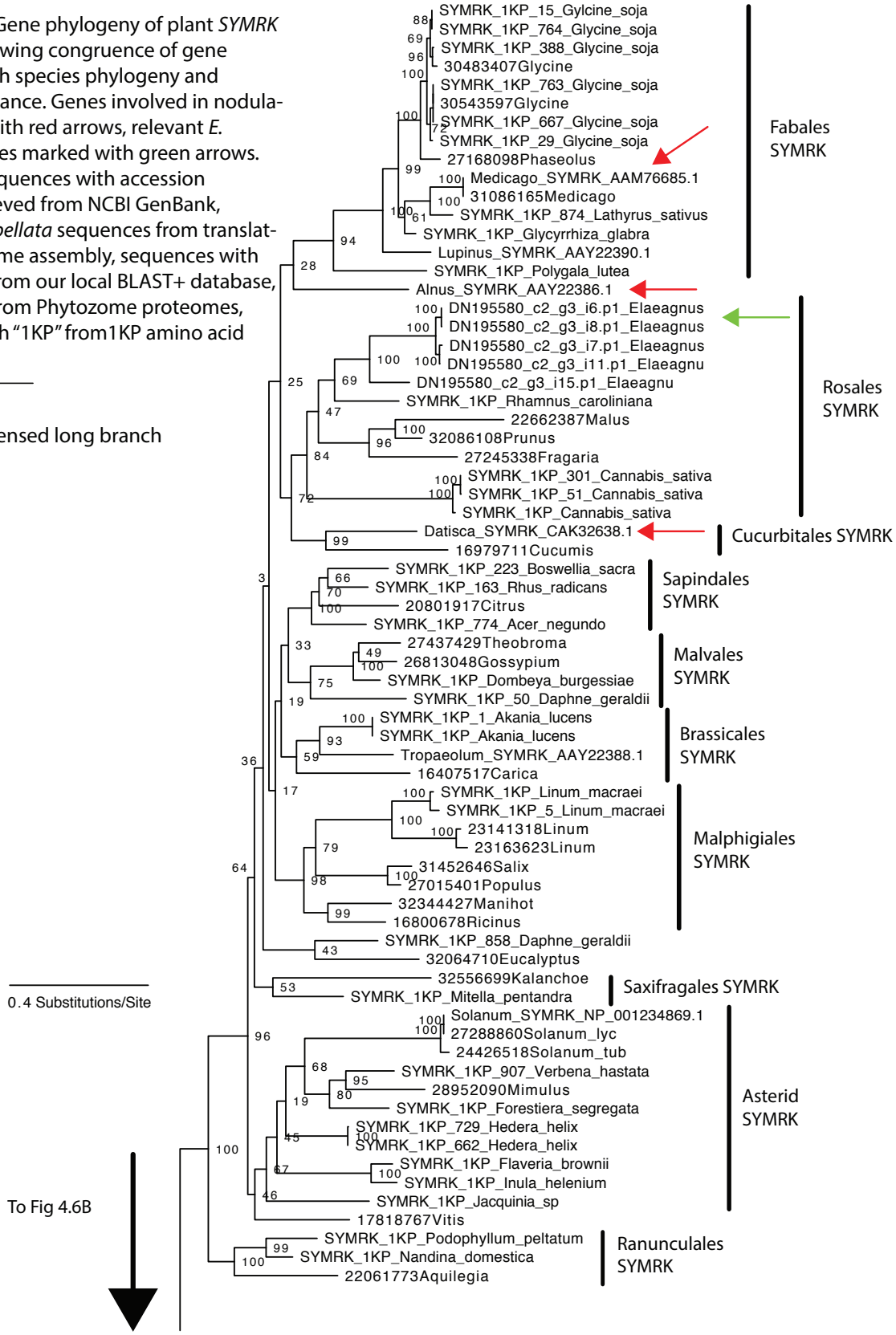


**Figure 4.6B**: Gene phylogeny of plant *SYMRK* orthologs, showing congruence of gene phylogeny with species phylogeny and vertical inheritance. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

Denotes condensed long branch

0.4 Substitutions/Site

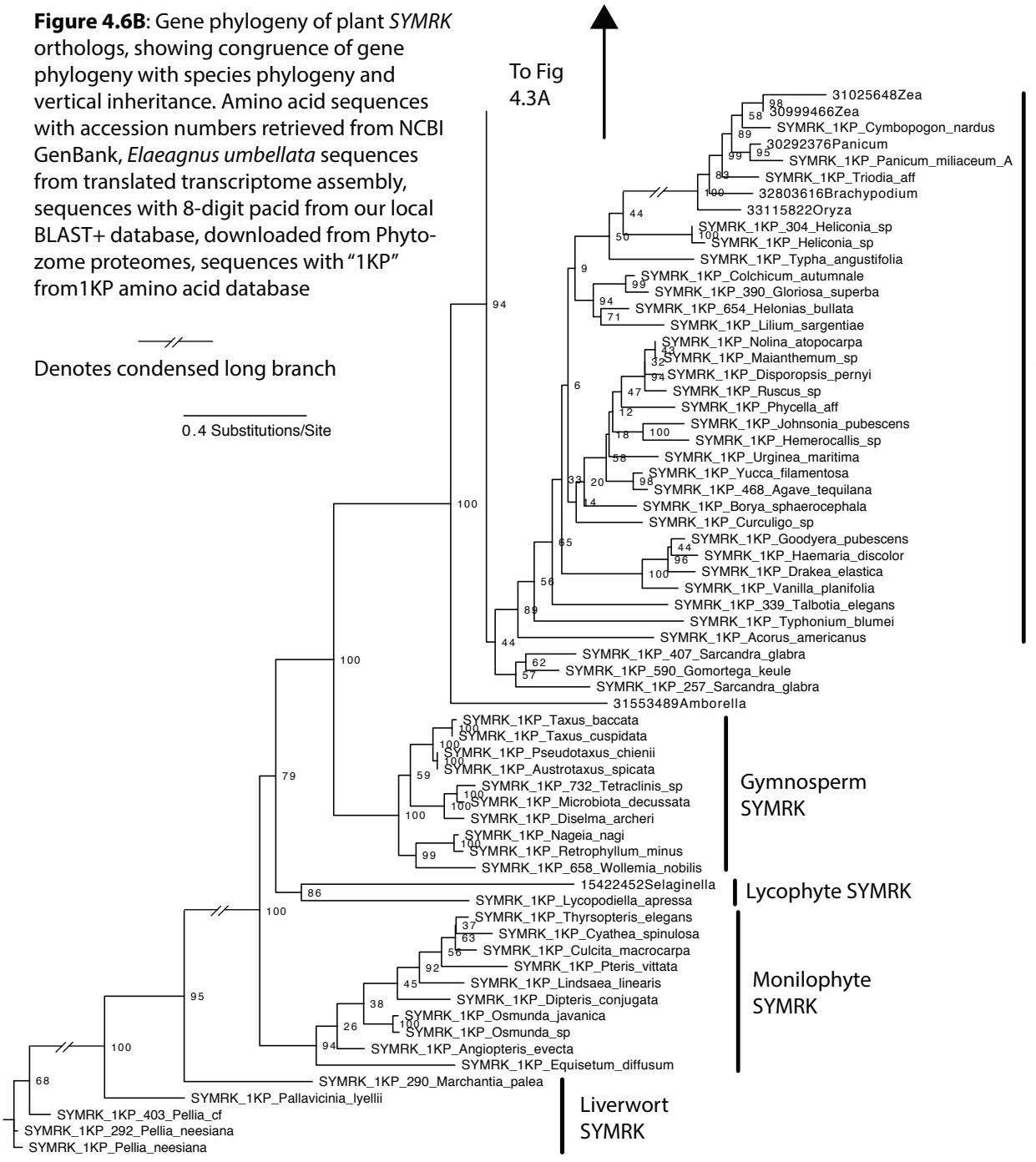
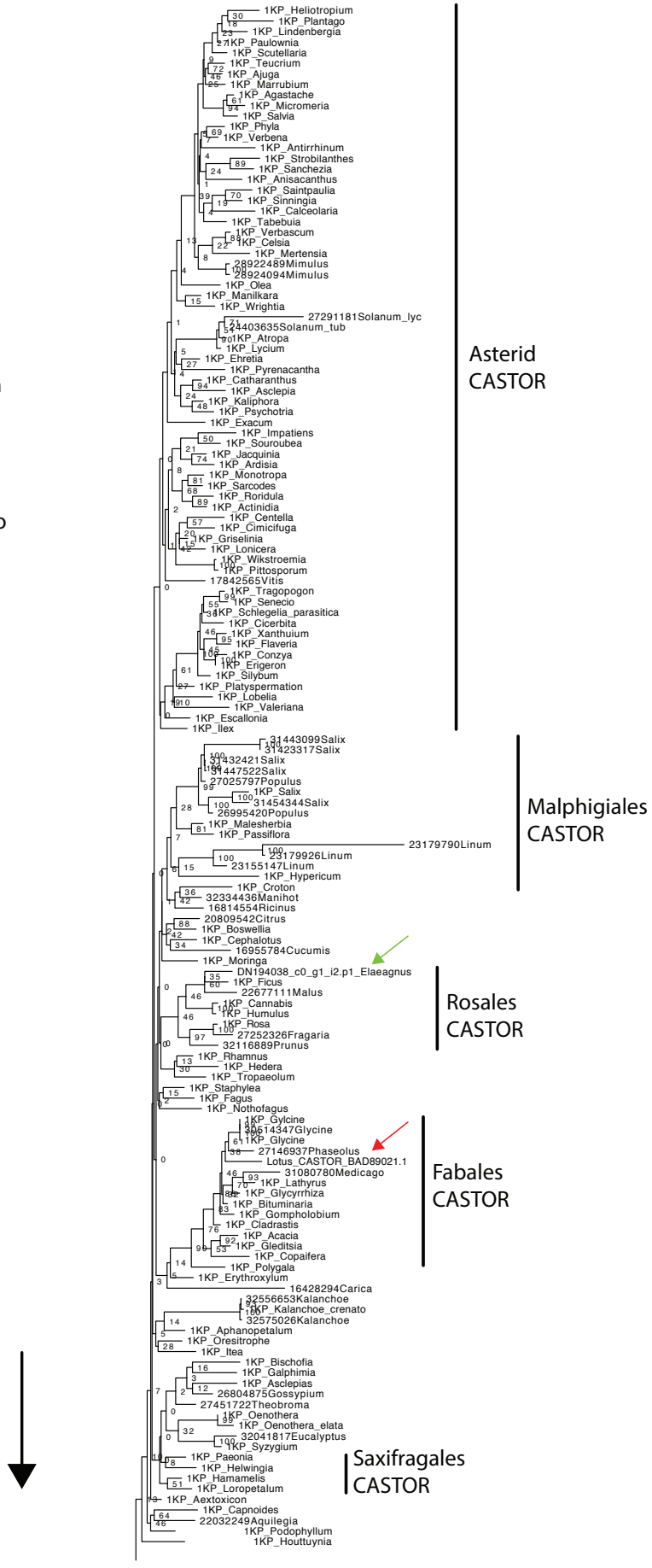


**Figure 4.7A** : Gene phylogeny of plant *CASTOR* orthologs, showing congruence of gene phylogeny with species phylogeny and vertical inheritance. Genes involved in nodulation marked with red arrows, relevant E. umbellata genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, down-loaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

—//—

Denotes condensed long branch

1.0 Substitutions/Site




To Fig 4.7B

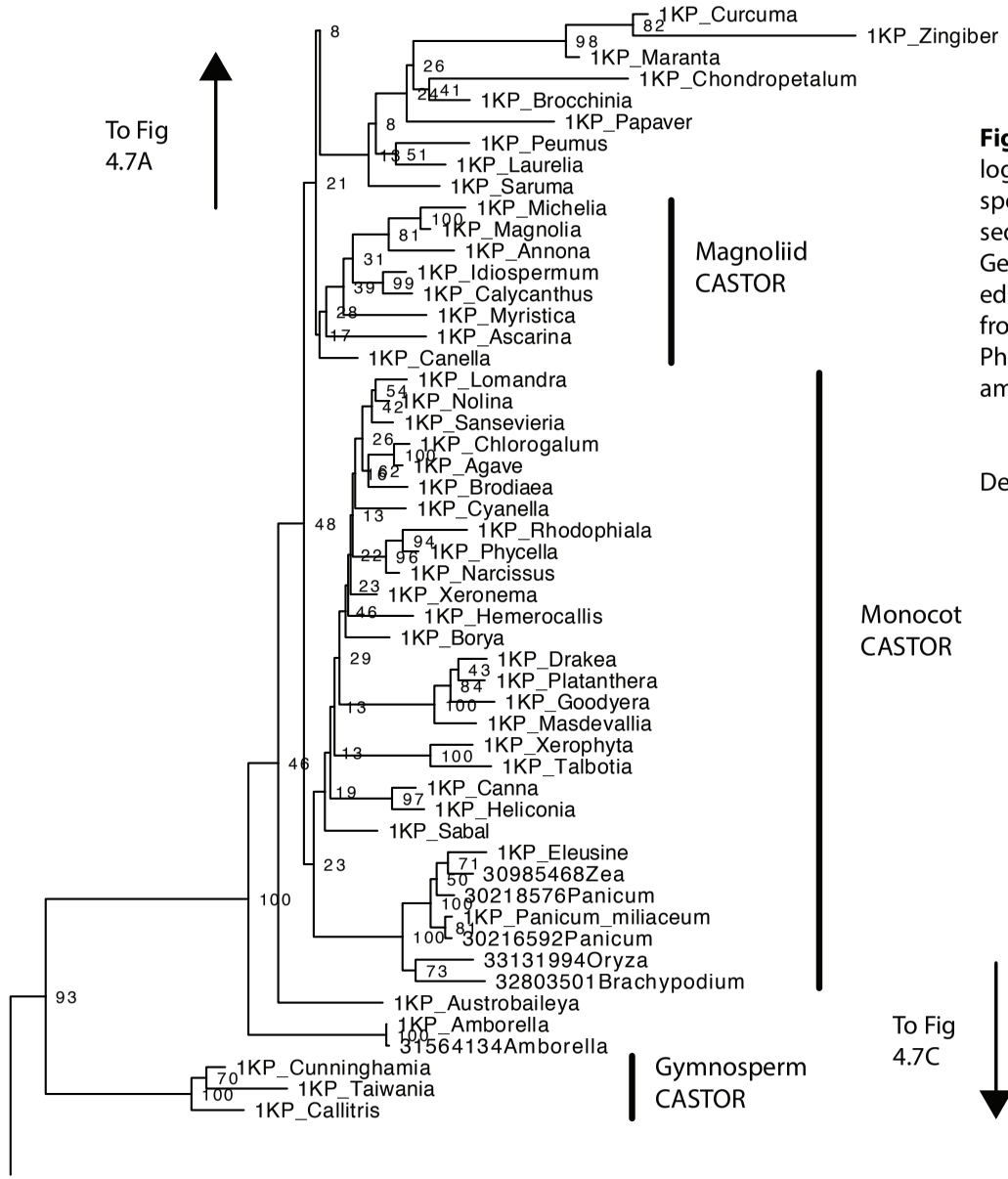



**Figure 4.7B** : Gene phylogeny of plant *CASTOR* orthologs, showing congruence of gene phylogeny with species phylogeny and vertical inheritance. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

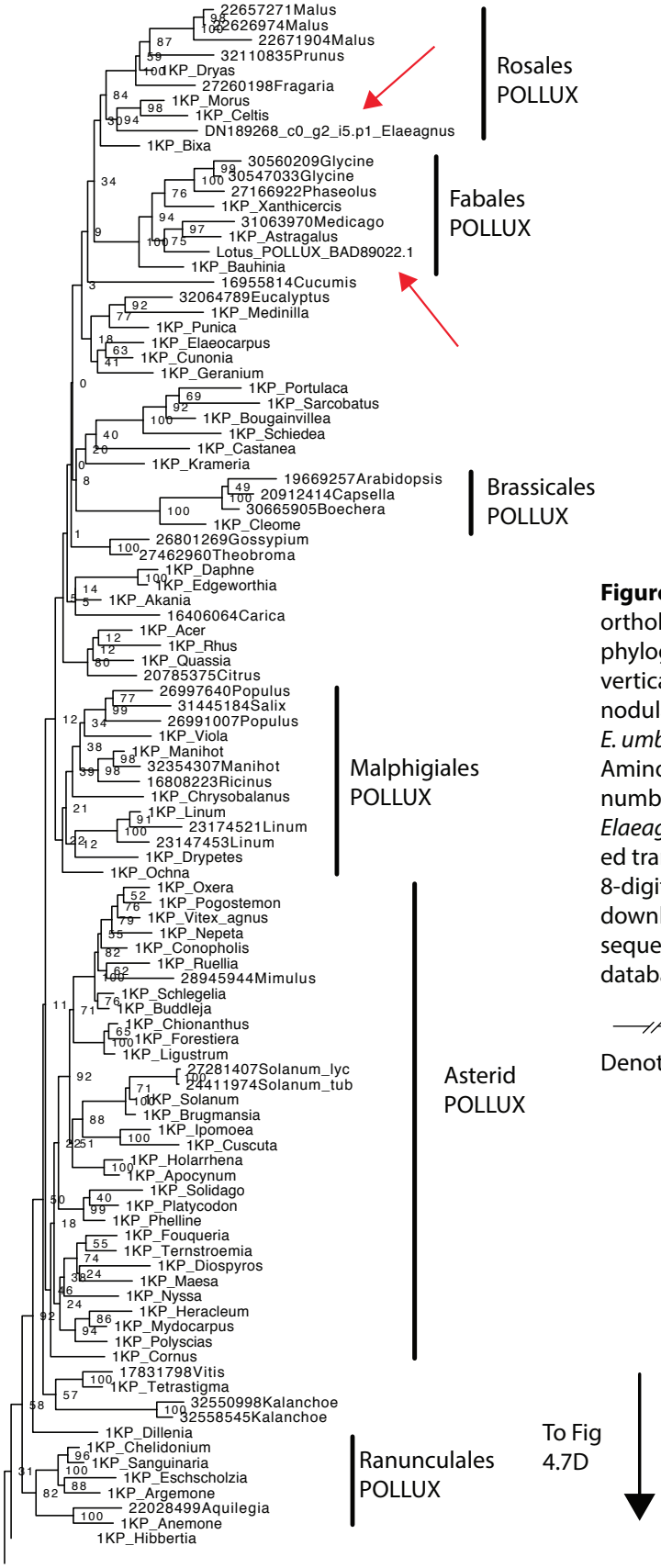



**Figure 4.7C**: Gene phylogeny of plant *POLLUX* orthologs, showing congruence of gene phylogeny with species phylogeny and vertical inheritance. Genes involved in nodulation marked with red arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

Denotes condensed long branch

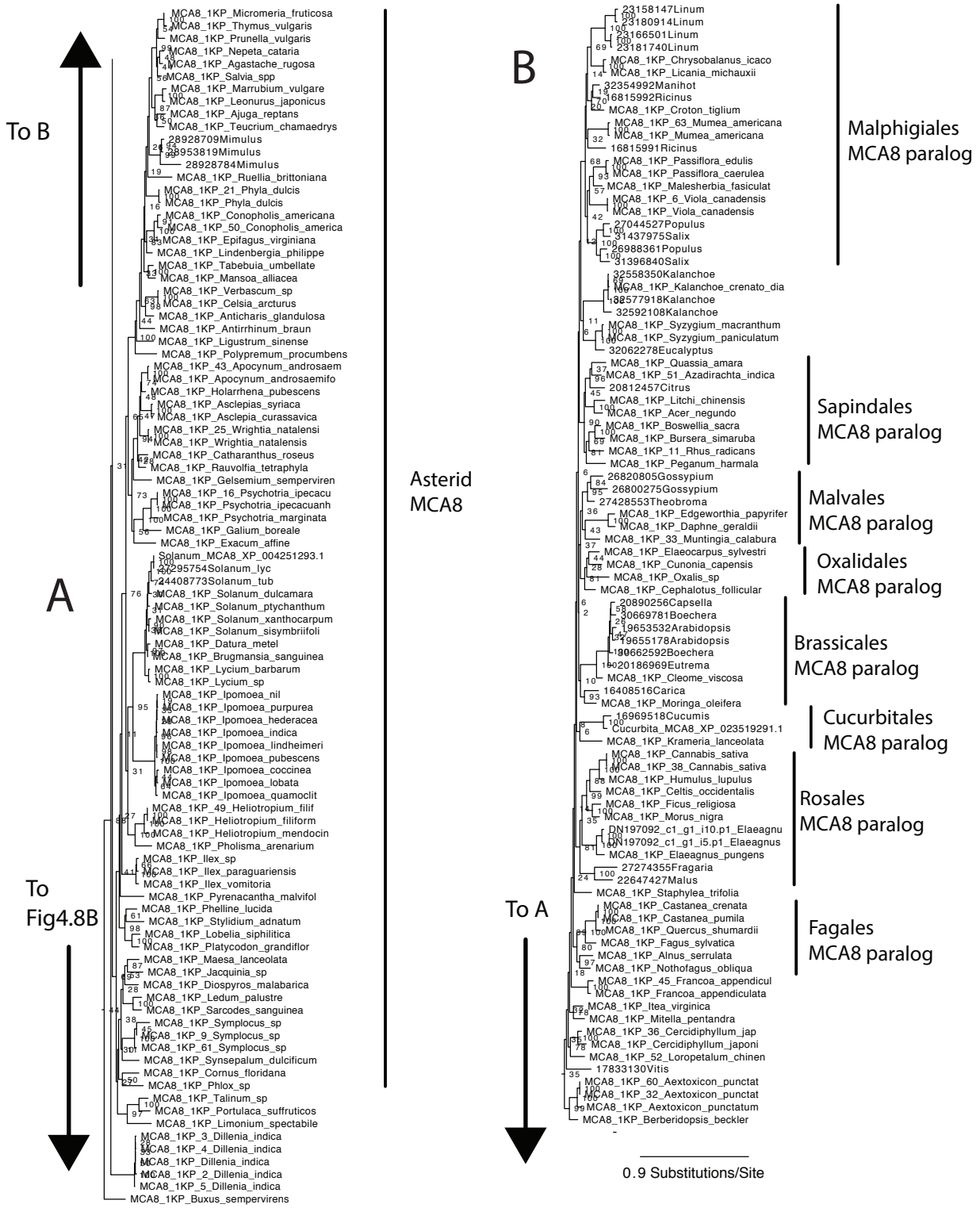



**Figure 4.8A** : Gene phylogeny of plant *MCA8* homologs, showing phylogenetic distribution of paralogous gene lineages. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

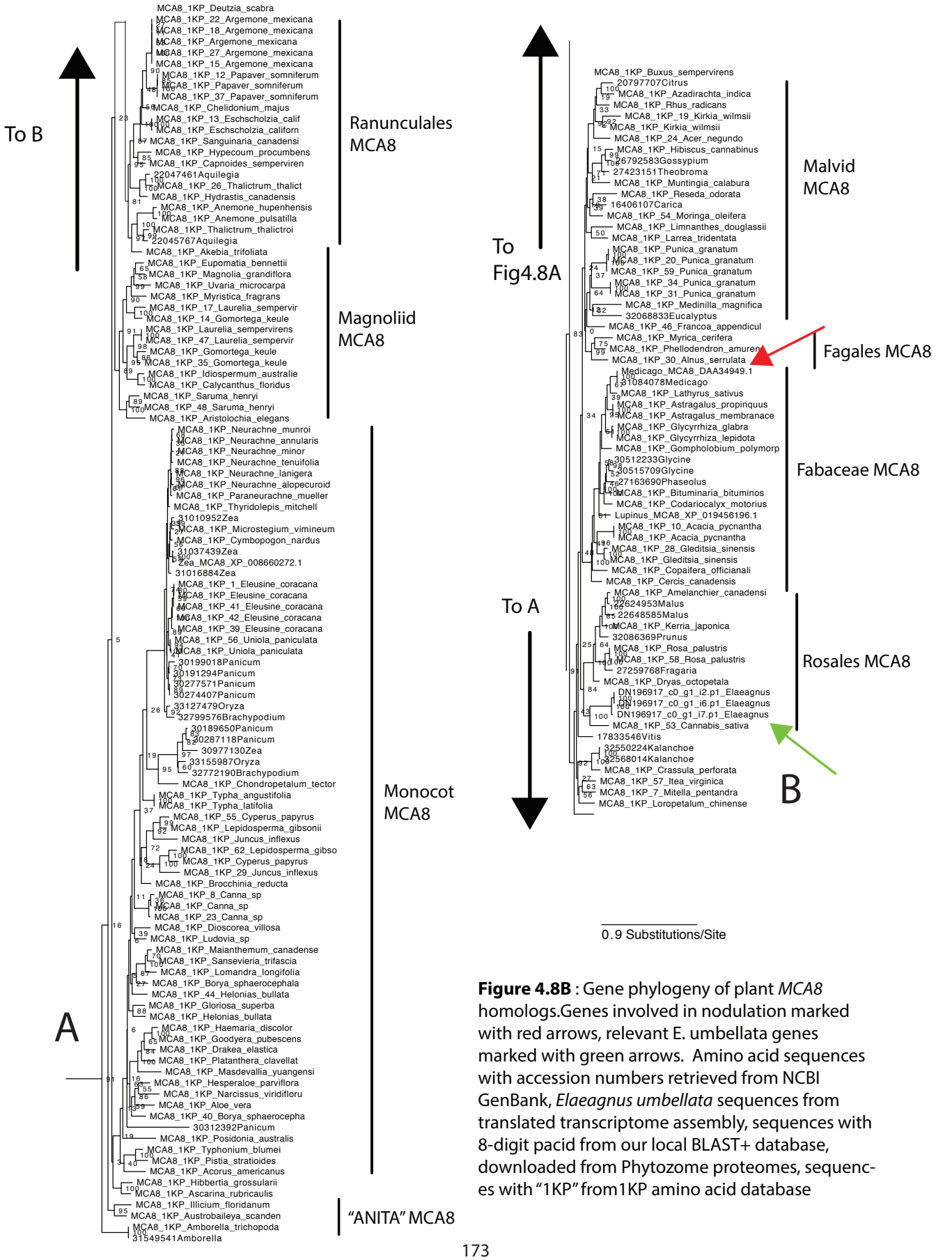



**Figure 4.8B** : Gene phylogeny of plant *MCA8* homologs.Genes involved in nodulation marked with red arrows, relevant E. umbellata genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

**Figure 4.9**: Gene phylogeny of plant *CCAMK* homologs, showing antiquity and vertical inheritance of *CCAMK*. Genes involved in nodulation marked with red arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

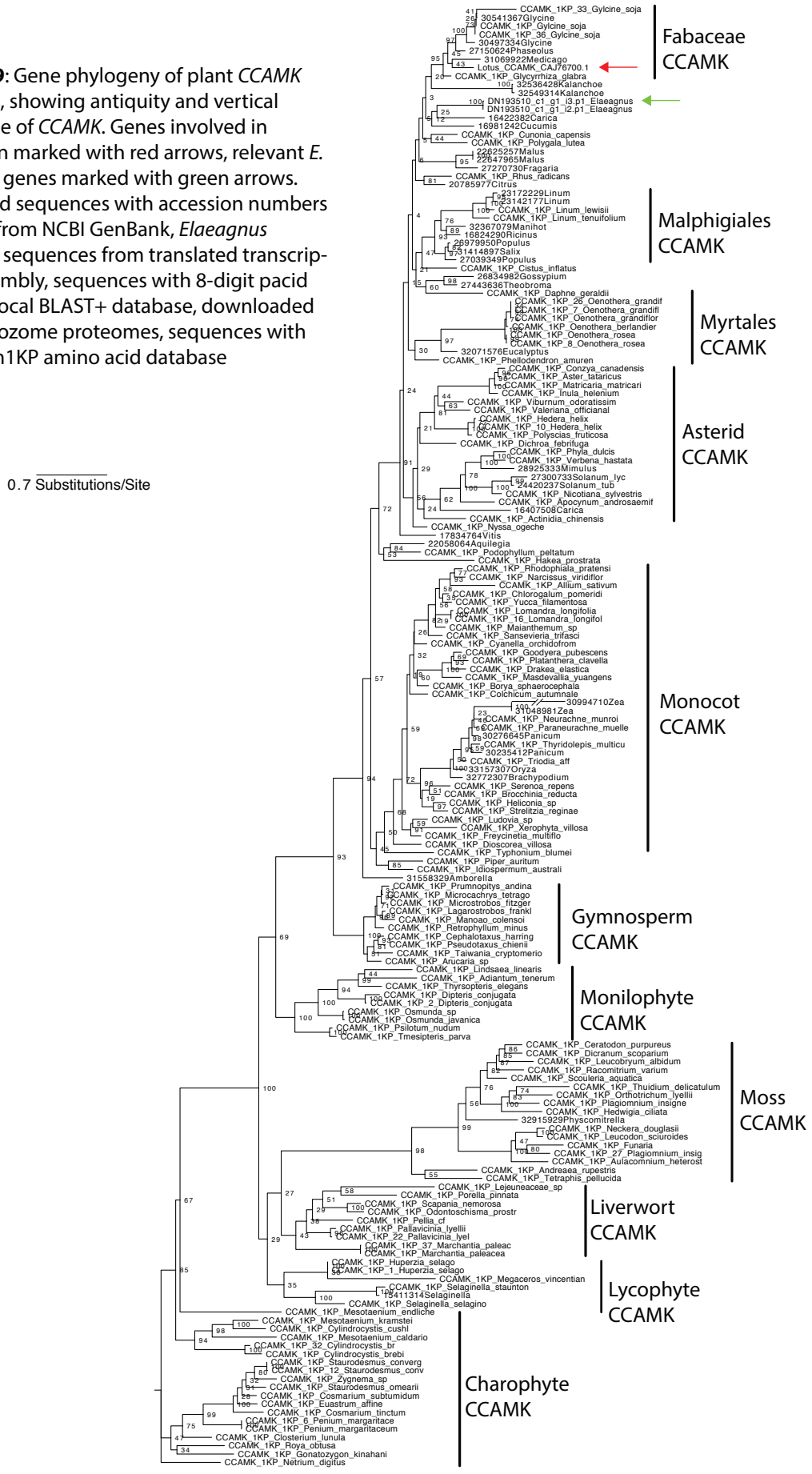
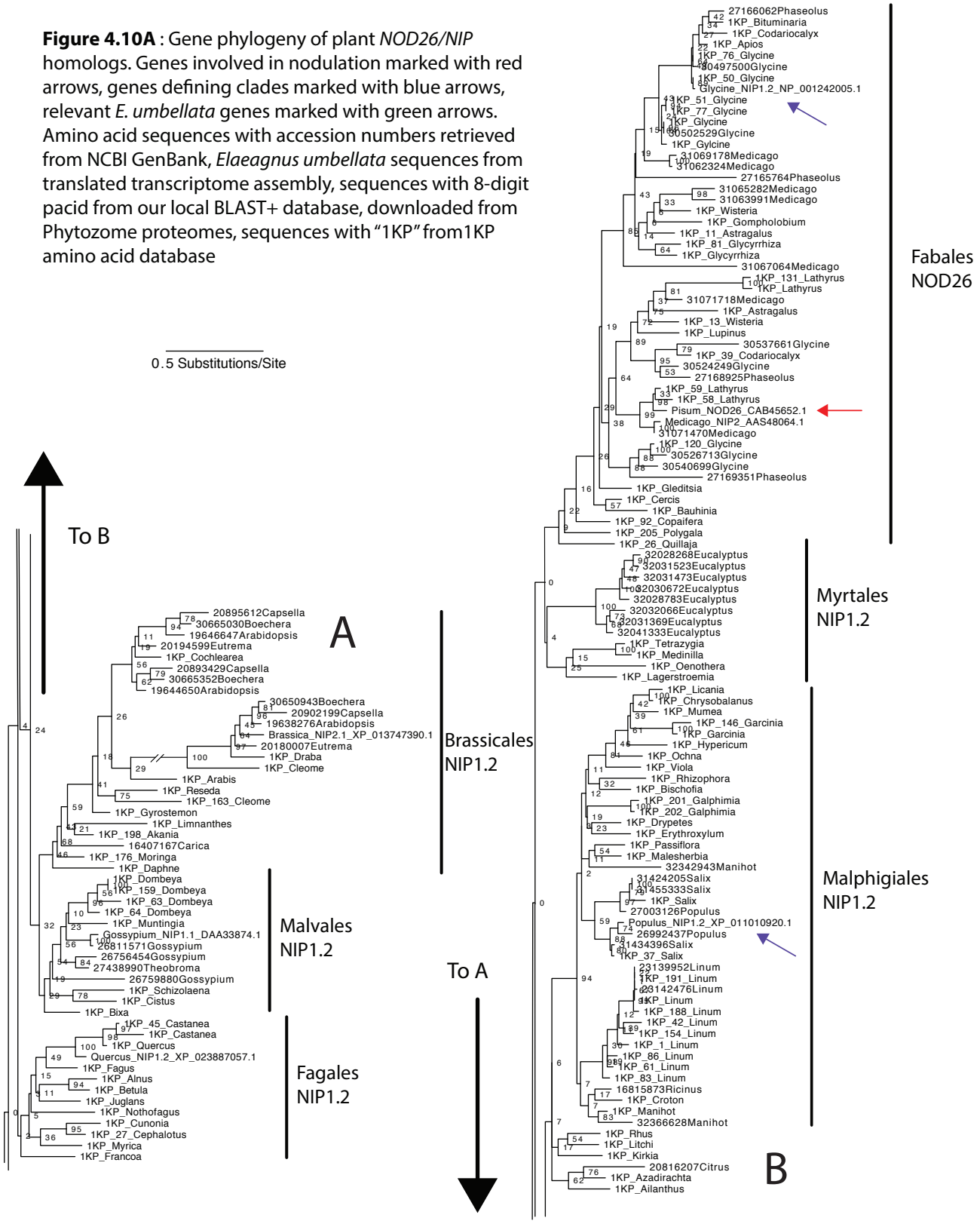


**Figure 4.10A** : Gene phylogeny of plant *NOD26/NIP* homologs. Genes involved in nodulation marked with red arrows, genes defining clades marked with blue arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database



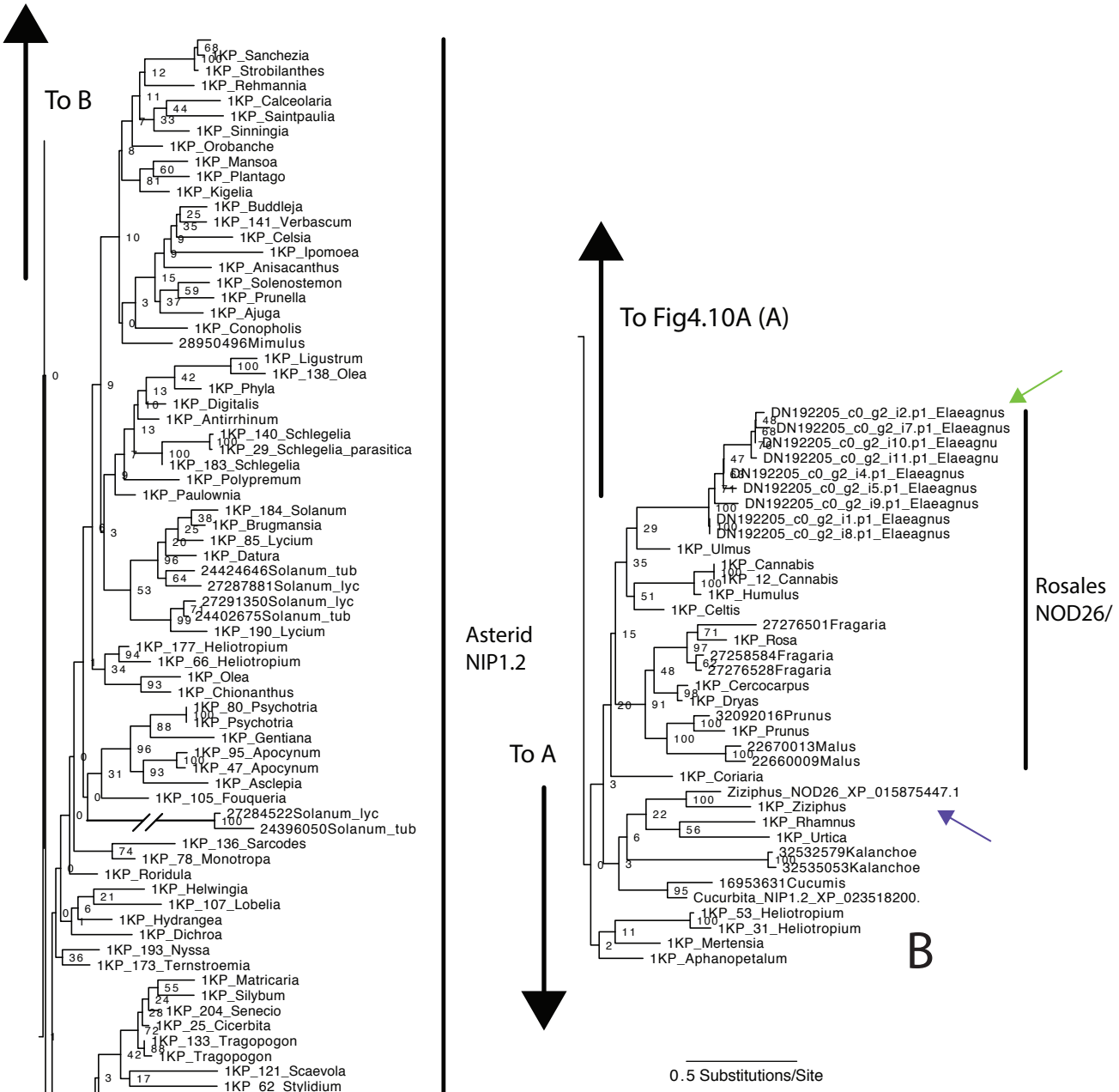



**Figure 4.10B** : Gene phylogeny of plant *NOD26/NIP* homologs. Genes involved in nodulation markedwith red arrows, genes defining clades markedwith blue arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

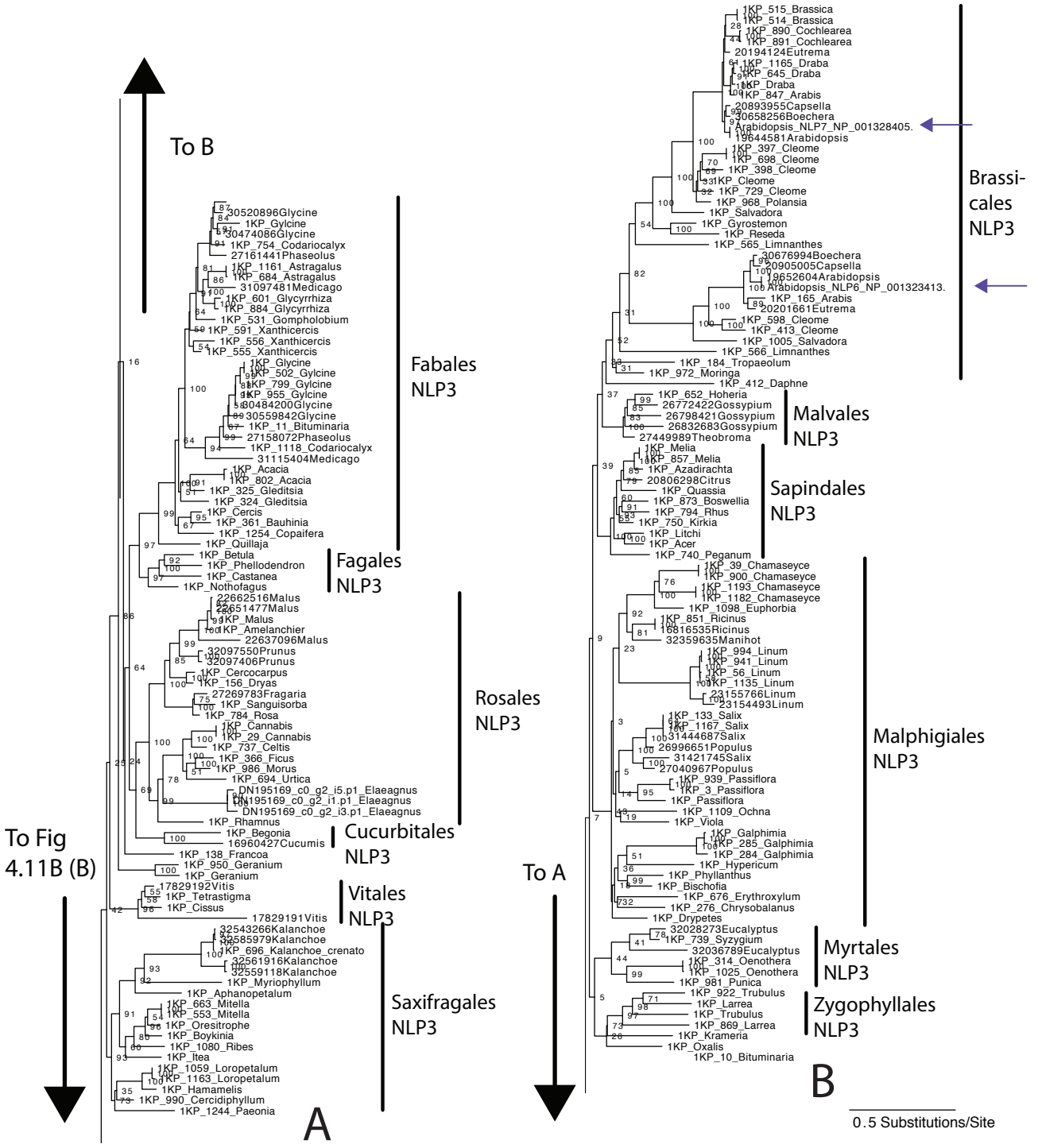



**Figure 4.11A** : Gene phylogeny of plant *NIN/NLP* homologs, showing orthology of *Elaeagnus umbellata NIN* with *NIN* in other nodulating clades. Genes involved in nodulation markedwith red arrows, genes defining clades markedwith blue arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

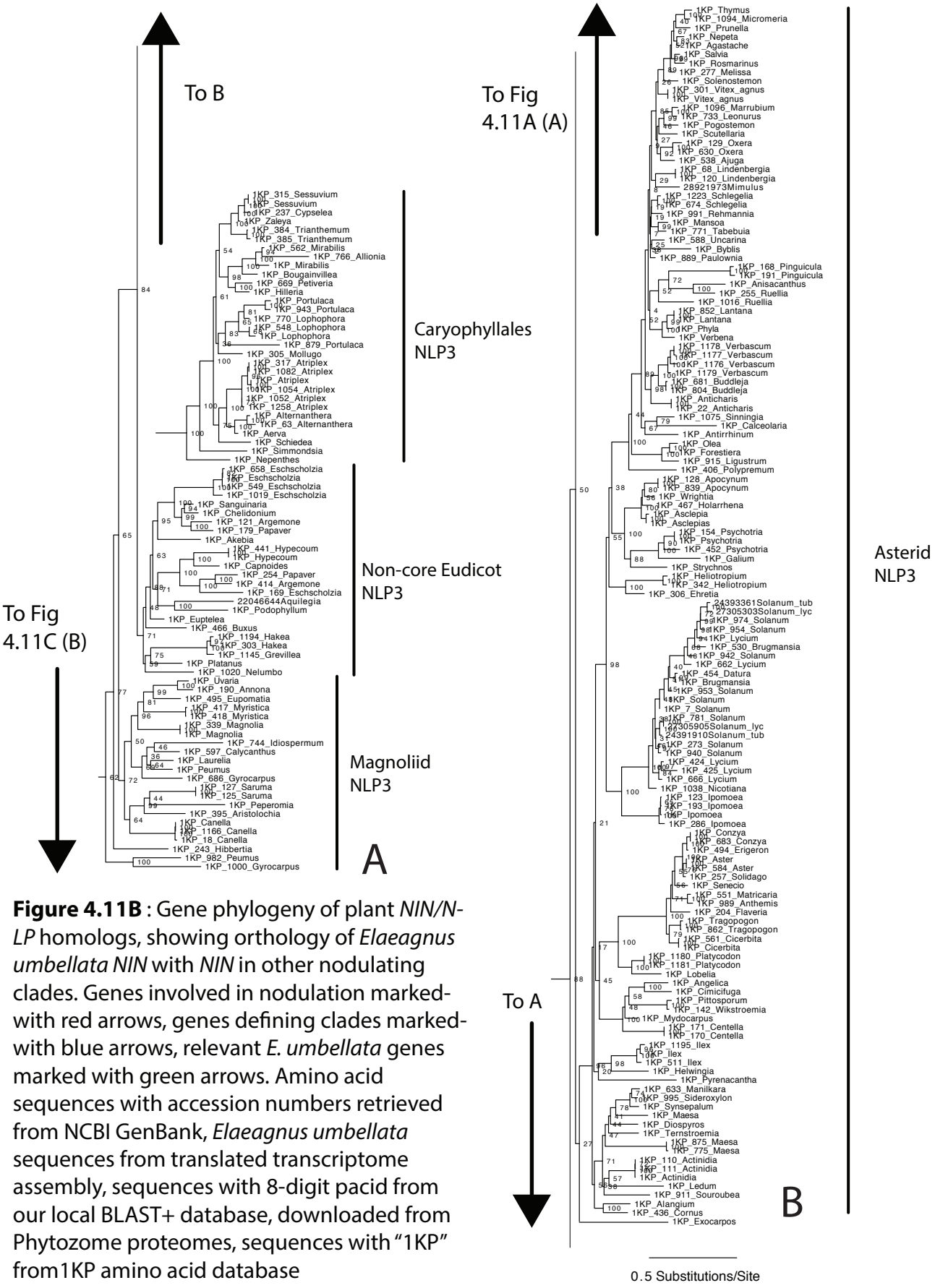



**Figure 4.11B** : Gene phylogeny of plant *NIN/N-LP* homologs, showing orthology of *Elaeagnus umbellata NIN* with *NIN* in other nodulating clades. Genes involved in nodulation marked-with red arrows, genes defining clades marked-with blue arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

**Figure 4.11C** : Gene phylogeny of plant NIN/NLP homologs, showing phylogenetic distribution of paralogous gene lineages. Genes involved in nodulation markedwith red arrows, genes defining clades markedwith blue arrows, relevant E. umbellata genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, Elaeagnus umbellata sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

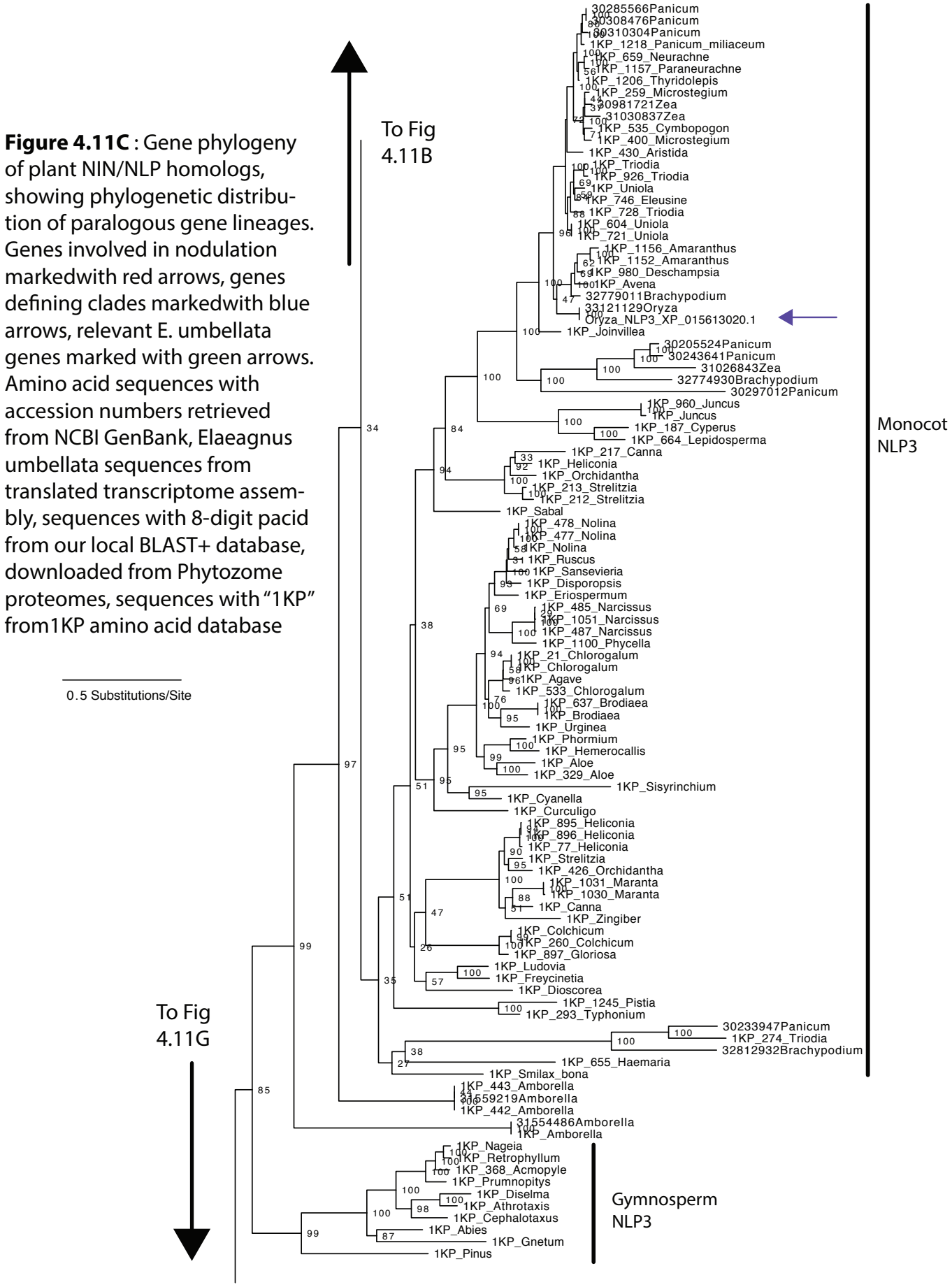
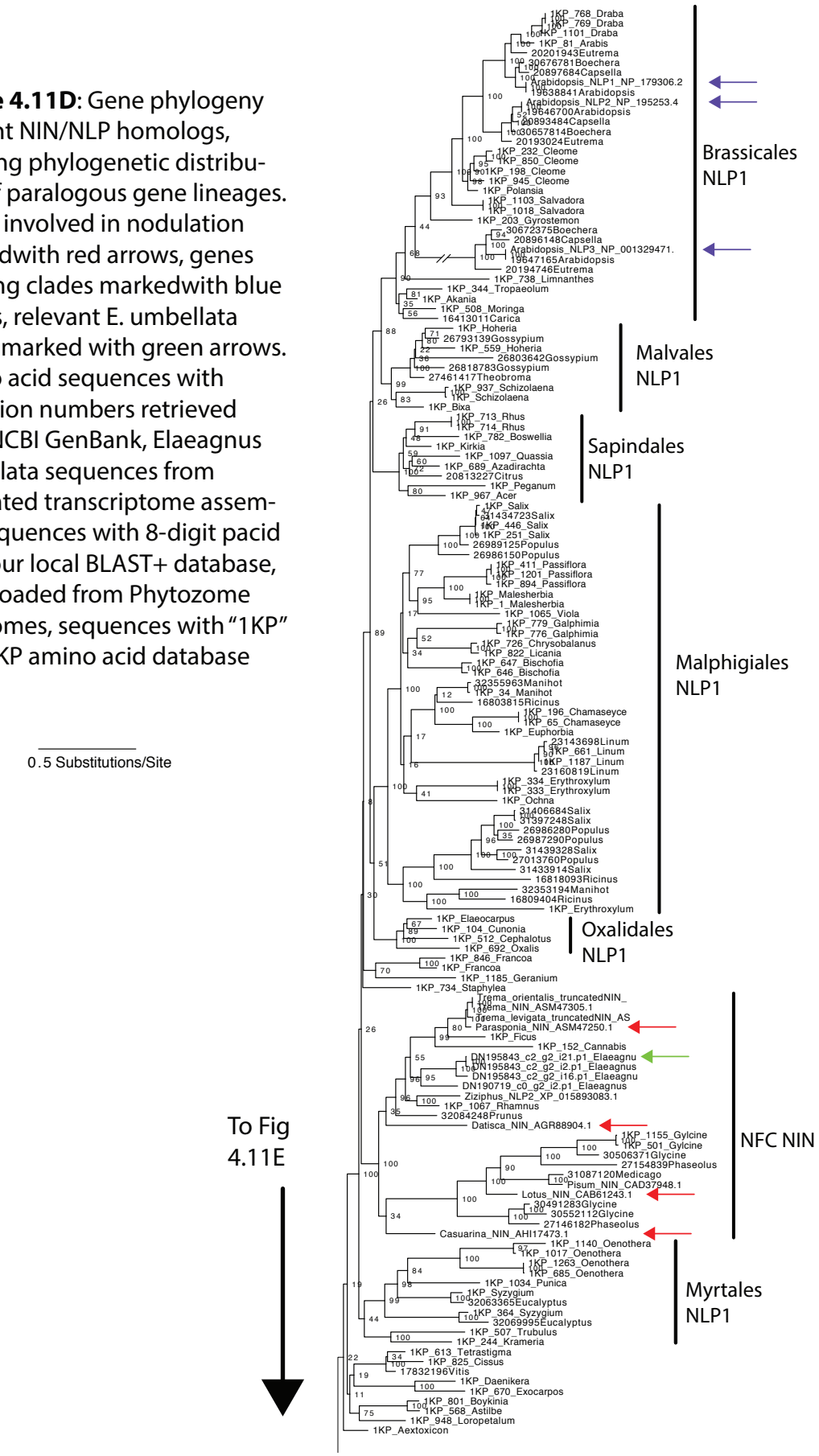


**Figure 4.11D**: Gene phylogeny of plant NIN/NLP homologs, showing phylogenetic distribution of paralogous gene lineages. Genes involved in nodulation markedwith red arrows, genes defining clades markedwith blue arrows, relevant E. umbellata genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, Elaeagnus umbellata sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database



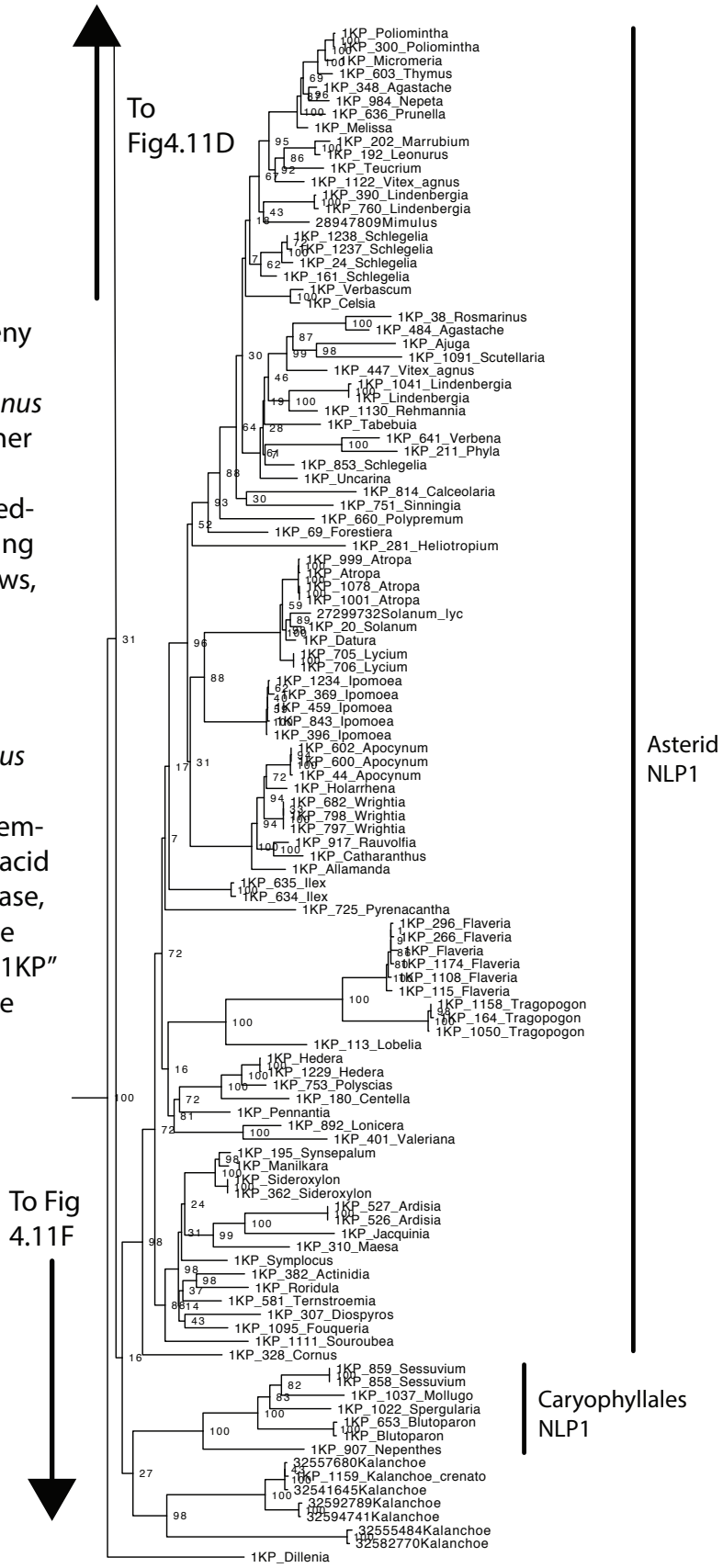



**Figure 4.11E** : Gene phylogeny of plant *NIN/NLP* homologs, showing orthology of *Elaeagnus umbellata NIN* with *NIN* in other nodulating clades. Genes involved in nodulation marked-with red arrows, genes defining clades markedwith blue arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

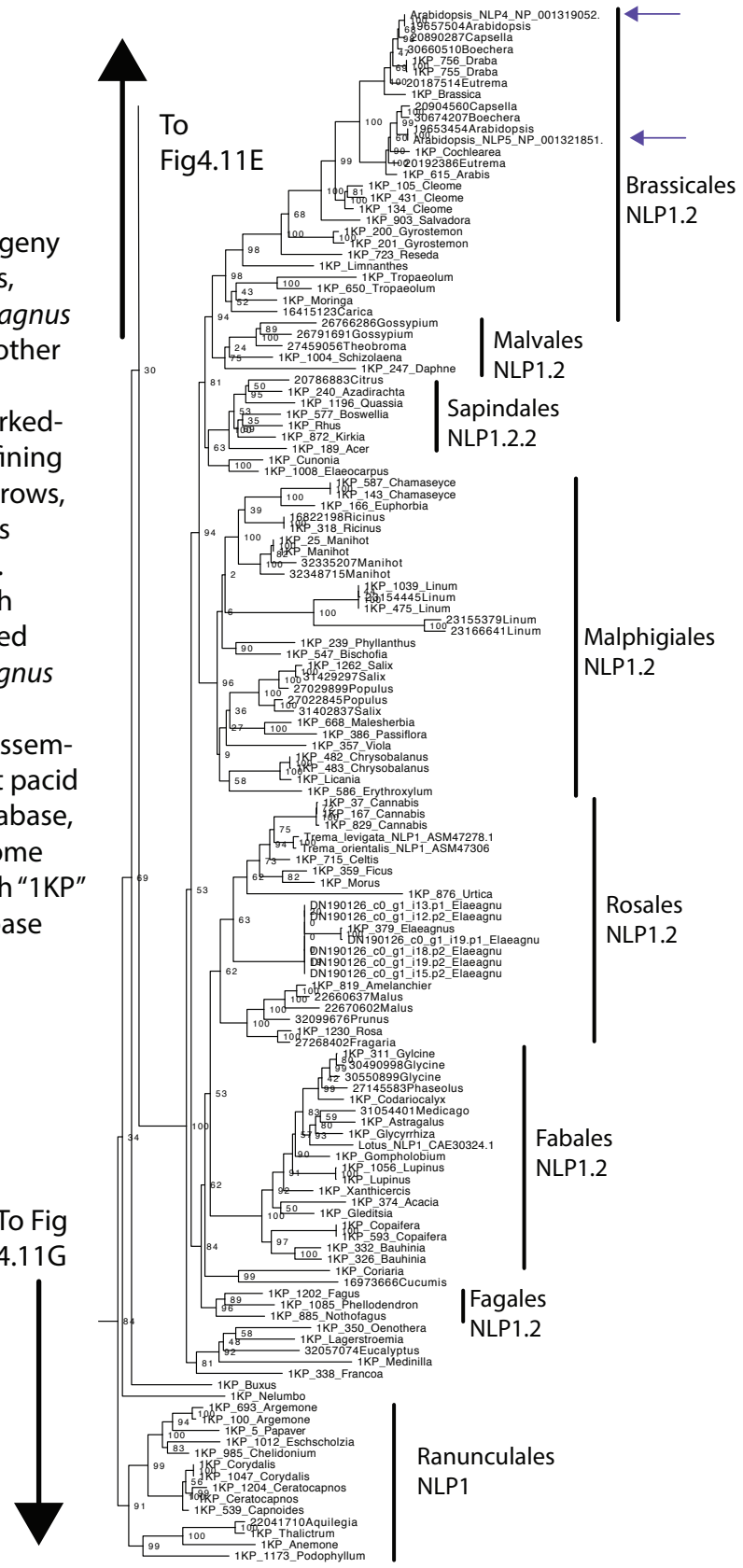



**Figure 4.11F** : Gene phylogeny of plant *NIN/NLP* homologs, showing orthology of *Elaeagnus umbellata NIN* with *NIN* in other nodulating clades. Genes involved in nodulation marked-with red arrows, genes defining clades markedwith blue arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

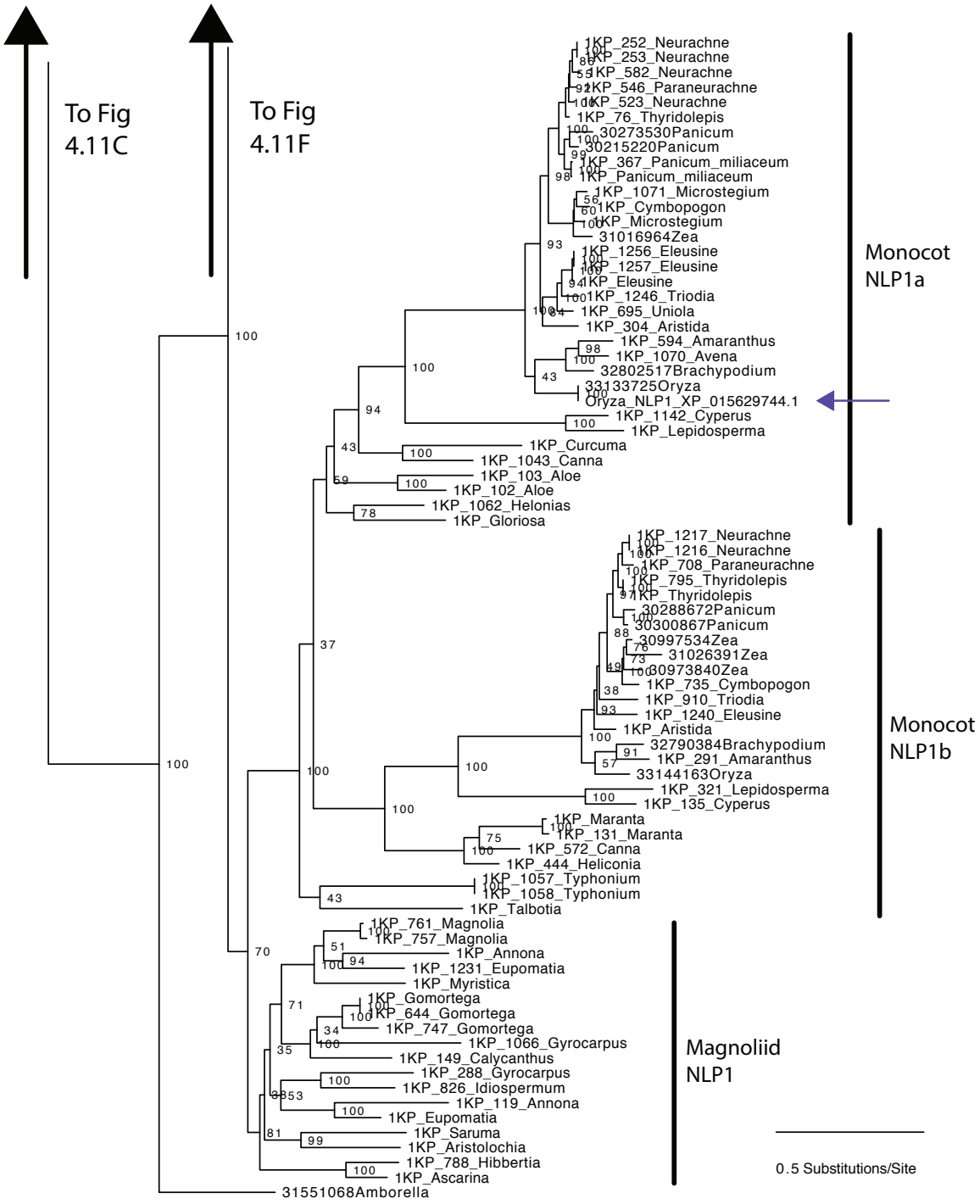



**Figure 4.11G** : Gene phylogeny of plant NIN/NLP homologs, showing phylogenetic distribution of paralogous gene lineages. Genes involved in nodulation markedwith red arrows, genes defining clades markedwith blue arrows, relevant E. umbellata genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, Elaeagnus umbellata sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

**Figure 4.11H** : Gene phylogeny of plant *NIN/NLP* homologs, showing phylogenetic distribution of paralogous gene lineages. Genes involved in nodulation markedwith red arrows, genes defining clades markedwith blue arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

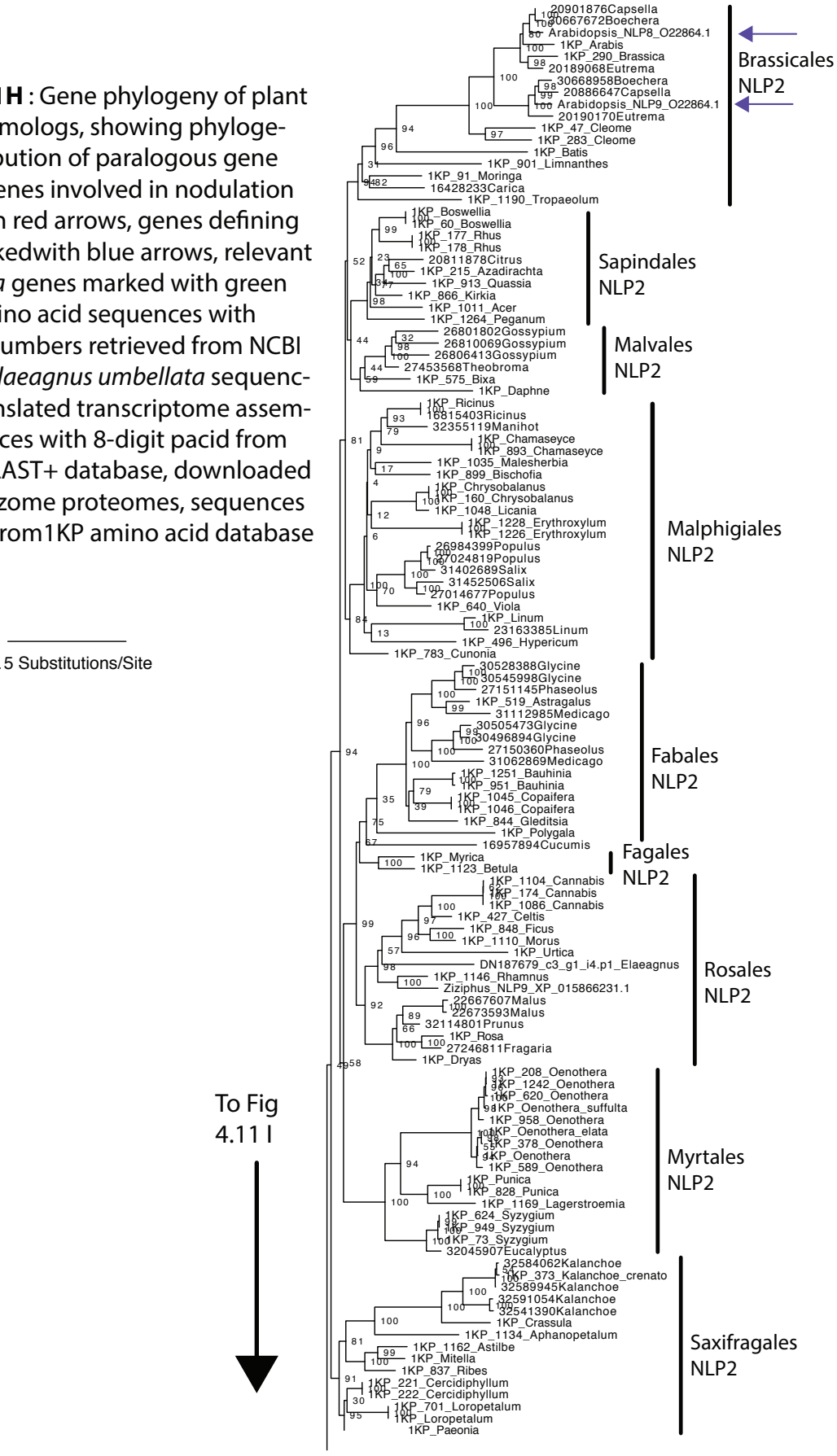
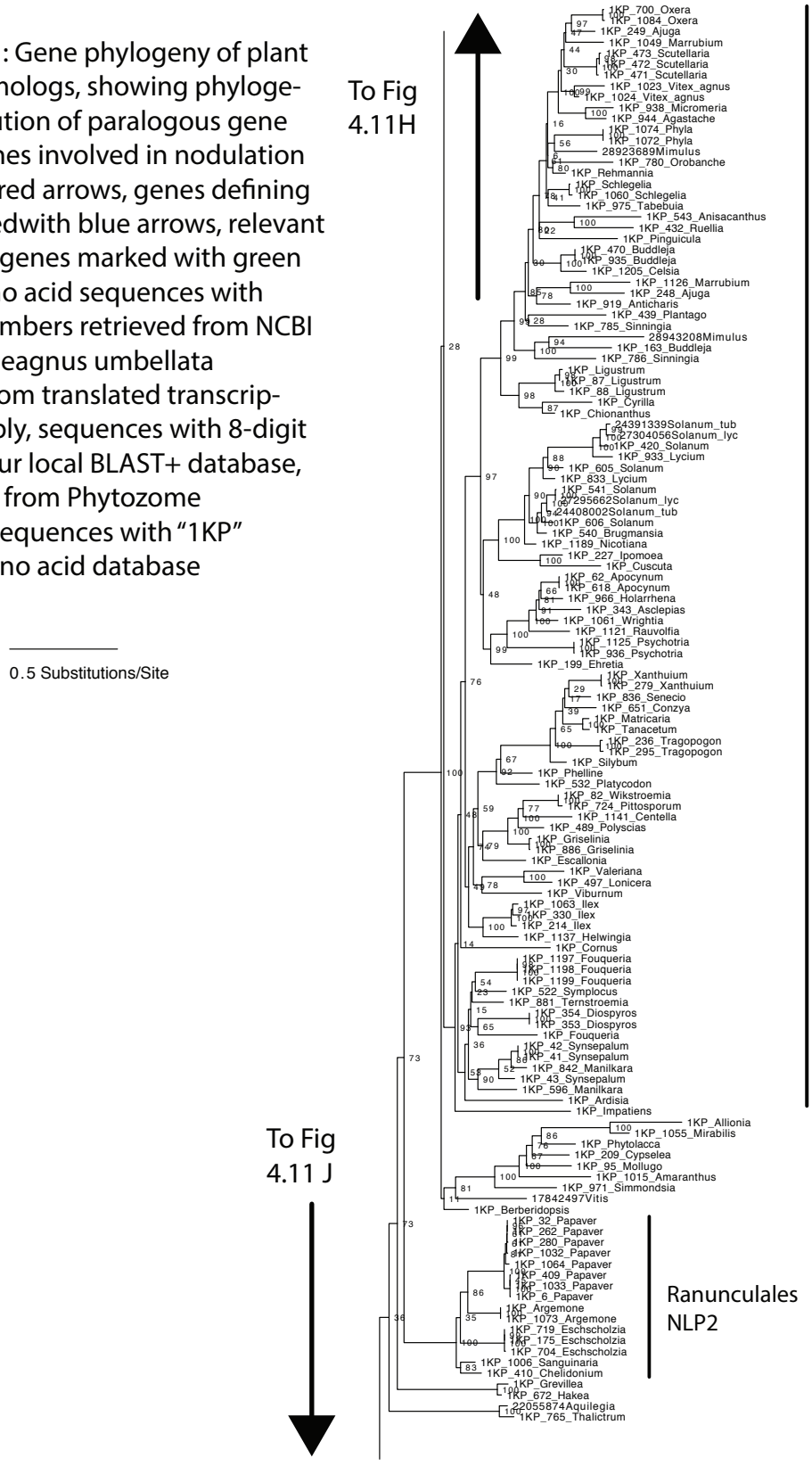


**Figure 4.11I** : Gene phylogeny of plant NIN/NLP homologs, showing phylogenetic distribution of paralogous gene lineages. Genes involved in nodulation markedwith red arrows, genes defining clades markedwith blue arrows, relevant E. umbellata genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, Elaeagnus umbellata sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database



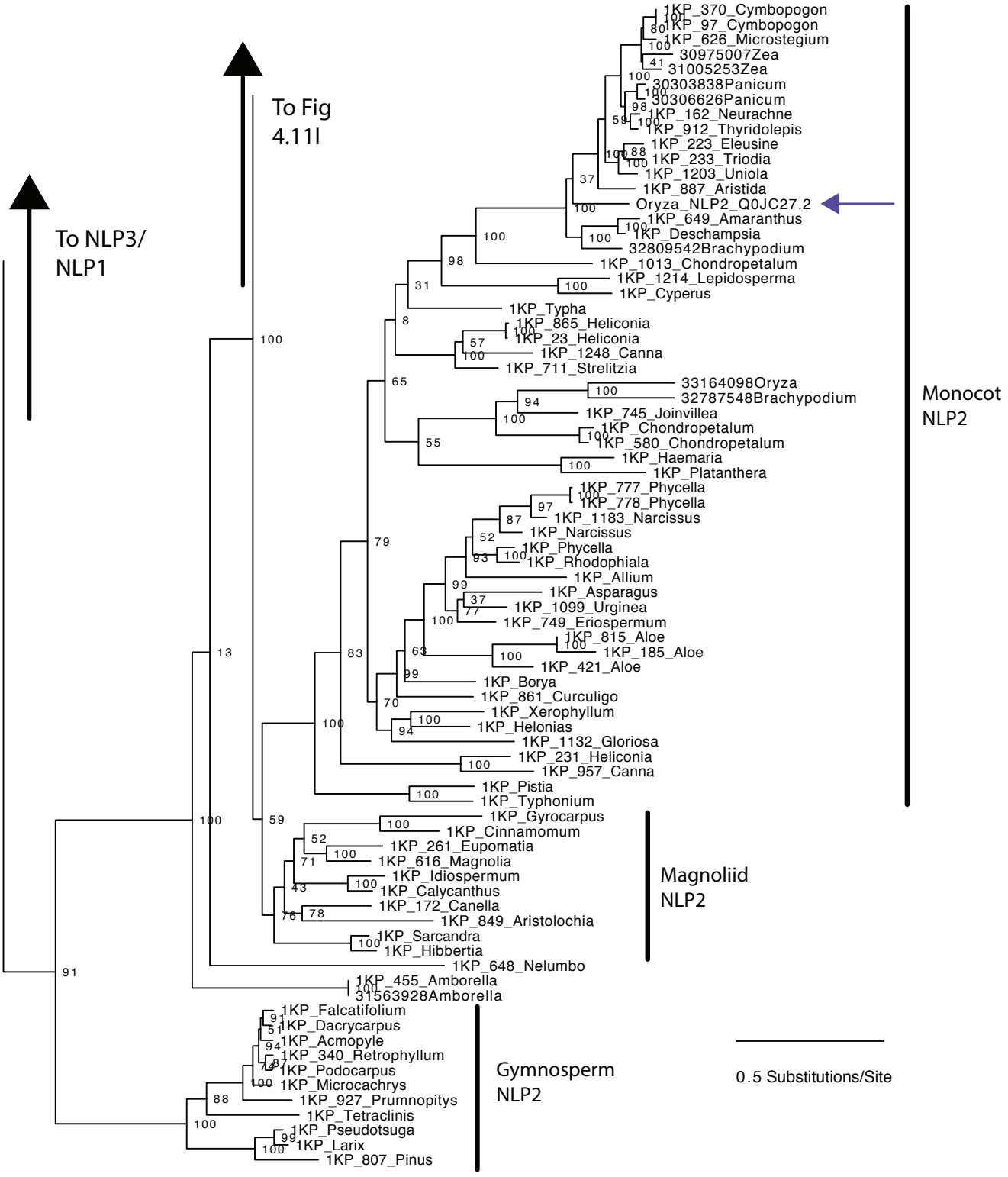



**Figure 4.11J** : Gene phylogeny of plant *NIN/NLP* homologs, showing orthology of *Elaeagnus umbellata NIN* with *NIN* in other nodulating clades. Genes involved in nodulation markedwith red arrows, genes defining clades markedwith blue arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

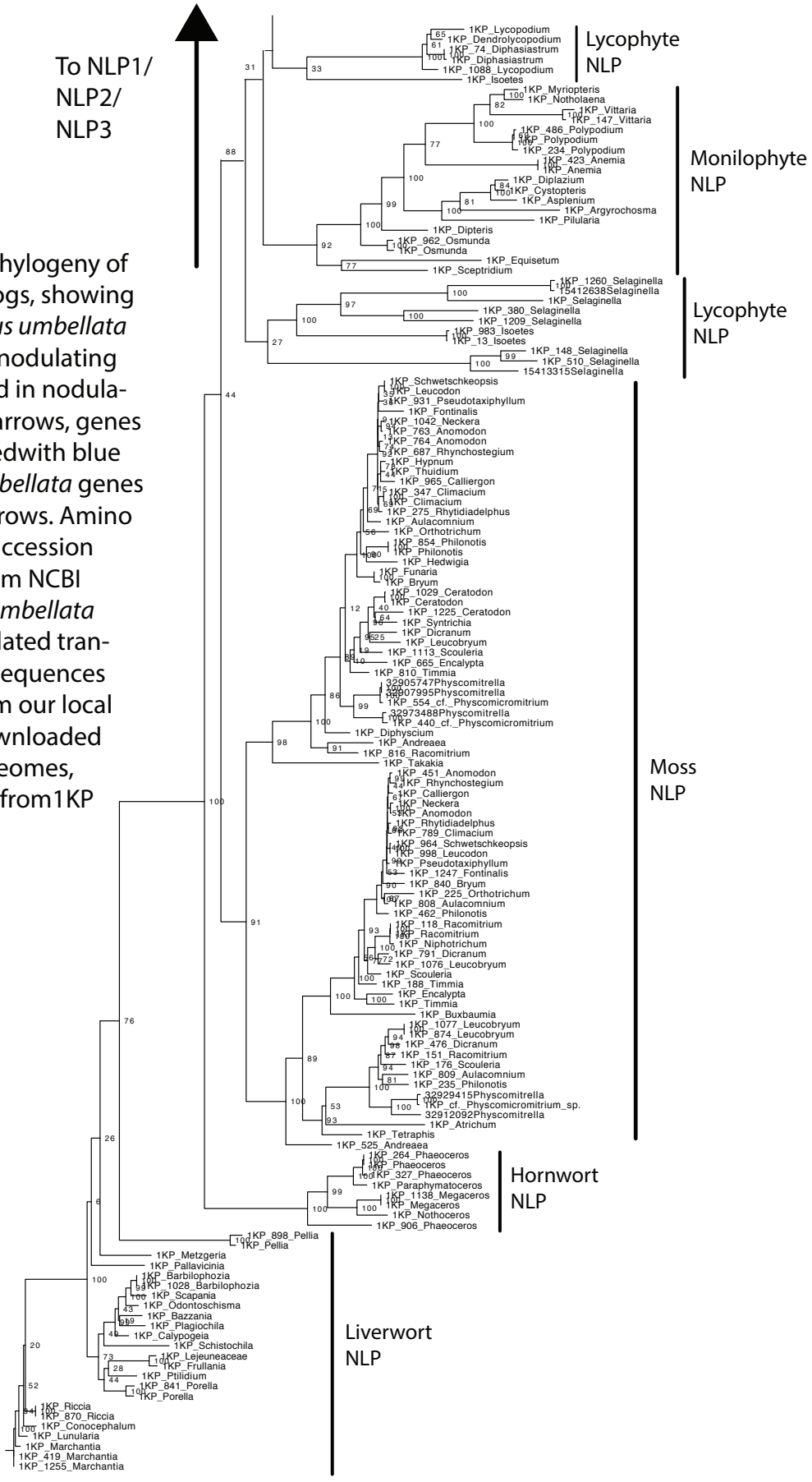



**Figure 4.11J** : Gene phylogeny of plant *NIN/NLP* homologs, showing orthology of *Elaeagnus umbellata NIN* with *NIN* in other nodulating clades. Genes involved in nodulation markedwith red arrows, genes defining clades markedwith blue arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

**Figure 4.12**: Gene phylogeny of plant *RPG* homologs. Genes involved in nodulation marked with red arrows, genes defining clades marked with blue arrows, relevant *E. umbellata* genes marked with green arrows. Amino acid sequences with accession numbers retrieved from NCBI GenBank, Elaeagnus umbellata sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, downloaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

Denotes condensed long branch

0.6 Substitutions/Site

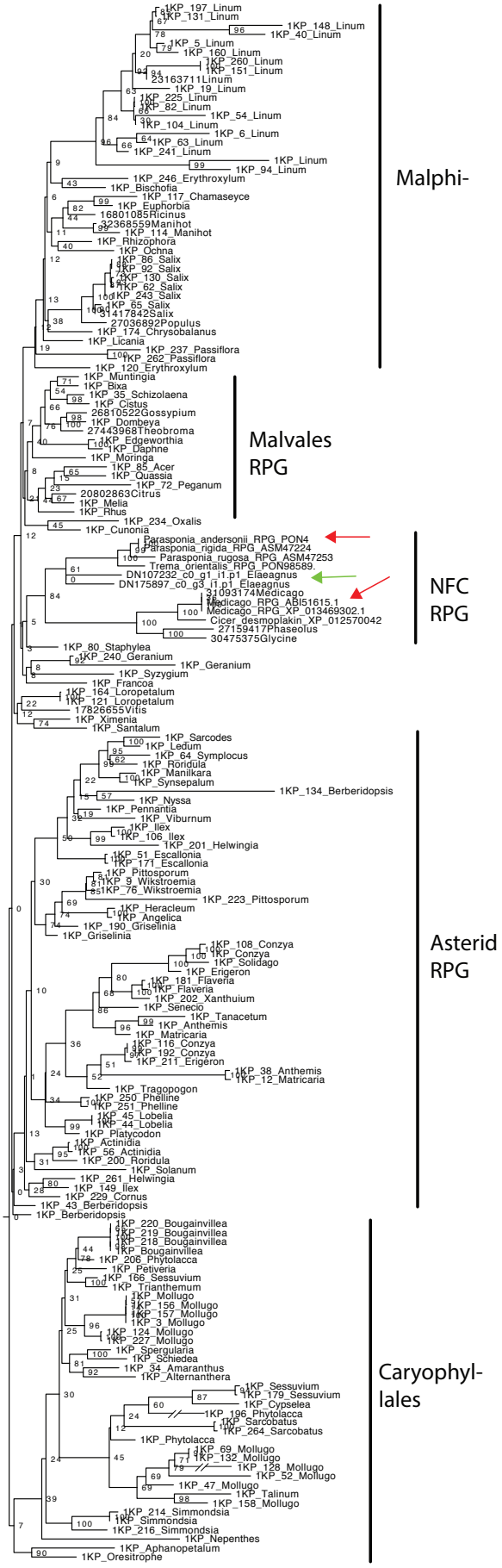


**Figure 4.13A** : Gene phylogeny of plant *CYCLOPS* homologs. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, down-loaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

Denotes condensed long branch

0.3 Substitutions/Site

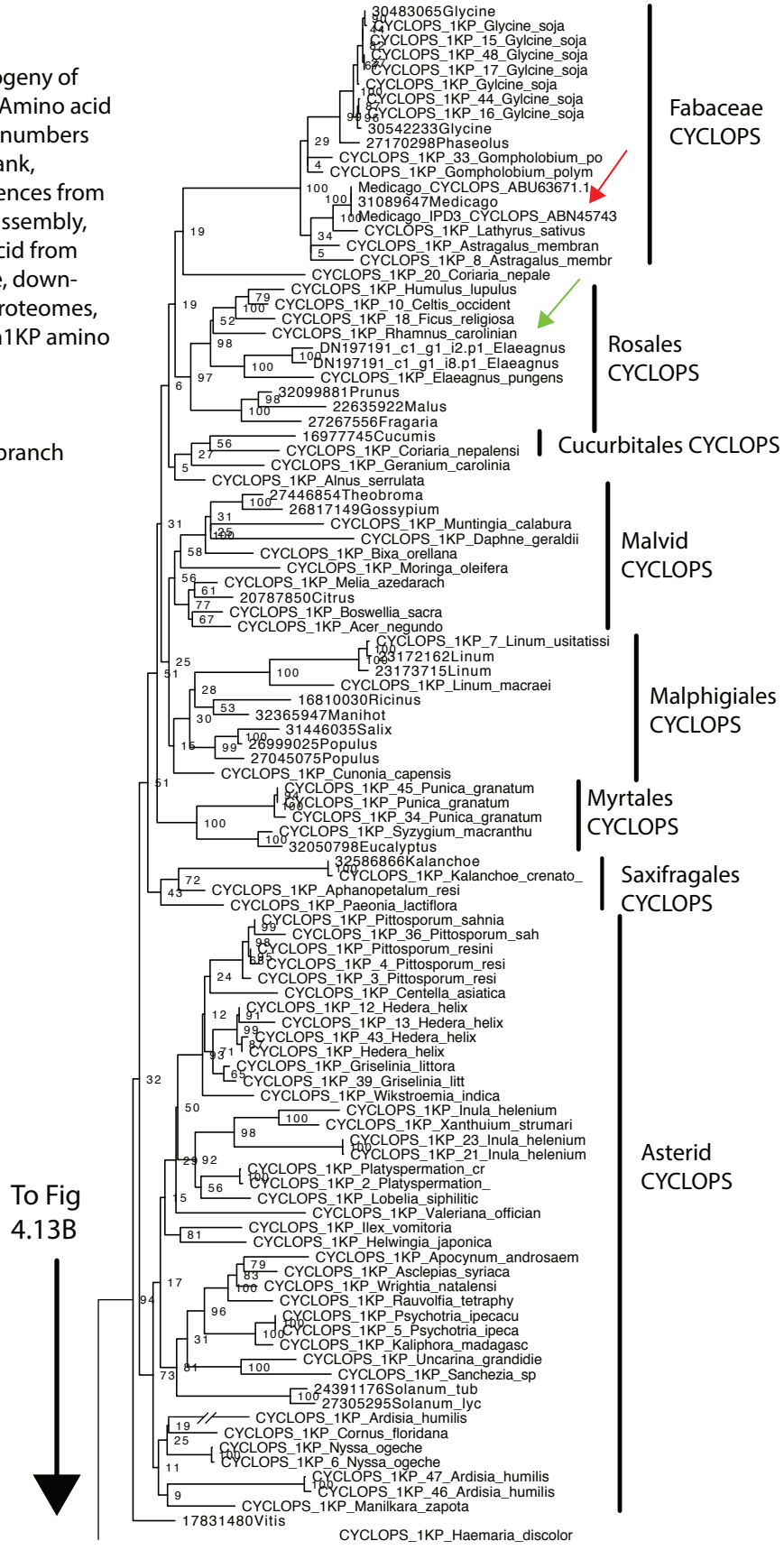
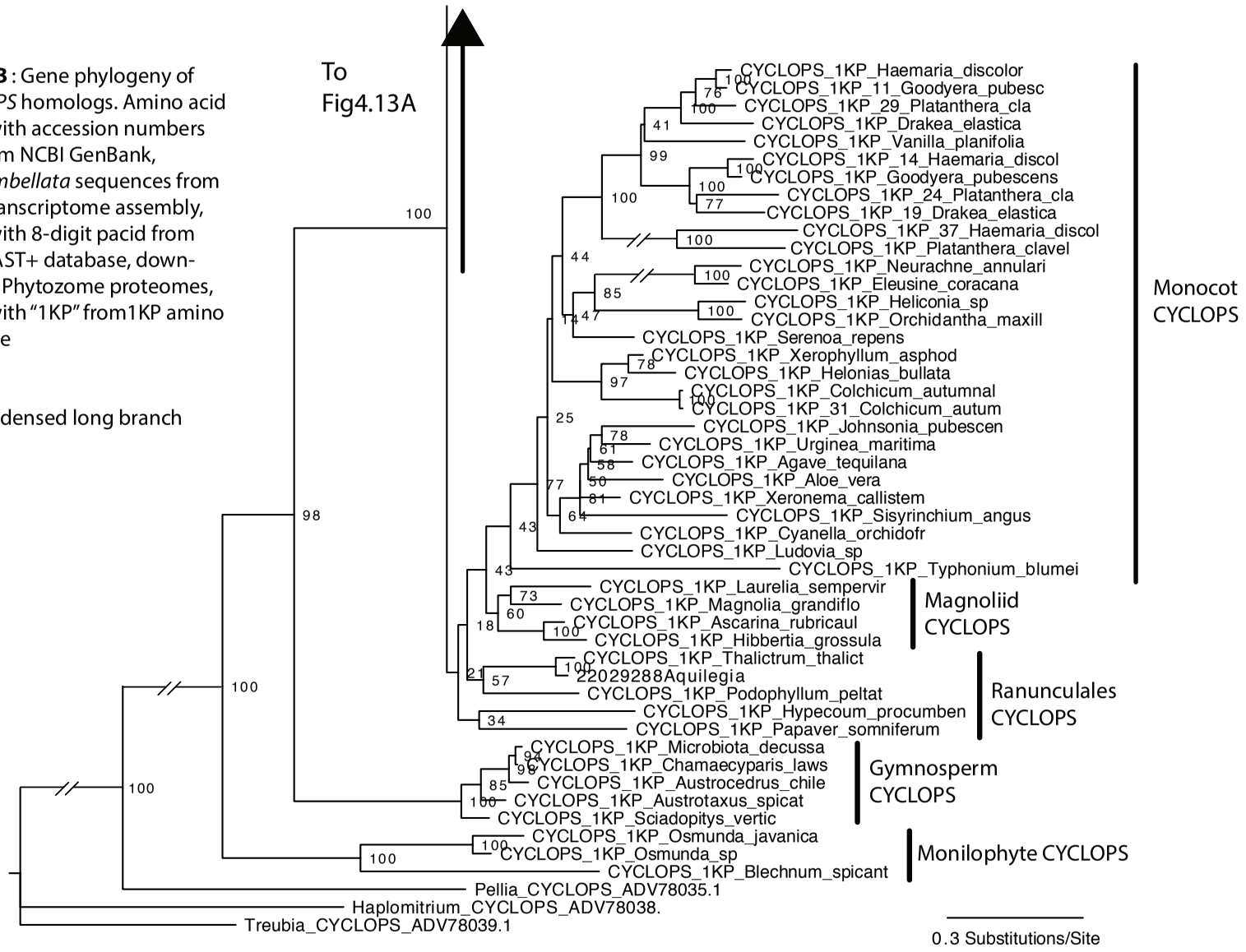


**Figure 4.13B** : Gene phylogeny of plant *CYCLOPS* homologs. Amino acid sequences with accession numbers retrieved from NCBI GenBank, *Elaeagnus umbellata* sequences from translated transcriptome assembly, sequences with 8-digit pacid from our local BLAST+ database, down-loaded from Phytozome proteomes, sequences with "1KP" from1KP amino acid database

Denotes condensed long branch



# Chapter 5

## Conclusions and Synthesis

By symbiotically coupling plants, which are photosynthetic but unable to fix nitrogen, with heterotrophic, nitrogen-fixing bacteria, nodulation drives nutrient cycling in terrestrial ecosystems worldwide (Smil, 1999). In addition to its stoichiometric impact, nodulation is a key innovation driving the diversity and ecological dominance of the Fabaceae, the 3rd largest plant family (McKey, 1994), as well as other nodulating rosid angiosperms. Legumes play a critical role in human agriculture by maintaining soil fertility in crop rotation schemes and providing dietary protein for humans directly and indirectly as animal feed (Galloway *et al.,* 1995). Understanding the evolution of nodulation has important practical implications for determining which crop species could be genetically engineered to nodulate and how (Markmann & Parniske, 2009; Charpentier & Oldroyd, 2010). This could reduce the use of environmentally costly synthetic nitrogen fertilizers, and improve human diet with increased protein (Charpentier & Oldroyd, 2010).

Aside from its environmental and agricultural importance, nodulation represents a fascinating example of the repeated evolution of a complex, symbiotic organ. The second chapter of this dissertation reviews the nodulation literature through the lens of homology and presents the symbiosis as a well-described case study in deep homology, evolutionary tinkering, and the ortholog conjecture. This chapter elaborates on and

sharpens several foundational concepts concerning the evolution of nodulation. First, the concept of an NFC-specific genetic endowment or "predisposition" to nodulate was introduced by Soltis *et al.* in 1995, but this idea has remained vague (van Velzen *et al.,* 2018), particularly since the best-understood genetic component of nodulation is the CSSP, which is present in all land plants and not specific to the NFC. In this chapter, I define the requirements of this predisposition in terms of an increase in the "bandwidth" of the single-copy orthologs in the CSSP, the ability to discriminate between AM and nodulation signals, highlighting the genes *CCAMK* and *SYMRK* as candidates for this increased bandwidth. Other non-CSSP genes may also play a role in this ability to discriminate between AM fungi and nodulating bacteria, and induce different downstream developmental cascades. Here, gene duplication and neofunctionalization or subfunctionalization may play a central role in a genetic "predisposition" to nodulate (Vanneste *et al.,* 2013).

The role of gene duplication in the evolutionary origin of nodulation is a second major concept elaborated in this review chapter. Some duplications of nodulation genes are specific to lineages within the NFC, such as the duplications of the LysM-RK *NFR1* and the origin of leghemoglobins in legumes (Gopalasubranamiam *et al.,* 2008; De Mita *et al.,* 2014). In other cases different nodulating lineages recruited different ancient paralogs, such as the differential recruitment of subtilase paralogs in the Fabales and Fagales (Taylor & Qiu, 2017).This differential recruitment of paralogs has several implications, aside from providing further evidence for the nonhomology of nodules (Doyle, 1994). Differential recruitment of paralogs could account for differences in nodulation in different lineages, and the phylogenetic distribution of different paralogs

has implications for what specific variation of nodulation is likely to arise in different lineages.

The discussion of how the evolutionary history of individual genes recruited for nodulation in this chapter provides concrete, detailed examples of deep homology and evolutionary tinkering in the independent origins of nodules the theoretical understanding of the evolution of nodulation. This chapter presents the most comprehensive and up-to-date synthesis of the literature concerning the evolution of nodulation, and applies theories on the origins of nodules to case studies of the independent recruitment of individual genes.

The third chapter demonstrates differential recruitment of deeply divergent paralogous gene lineages in the independent origins of nodulation, showing that the subtilases recruited for nodulation in the Fagales and Fabales are in paralogous gene lineages that diverged long before the origin of the NFC. In light of the transcriptional conservation of these subtilases in nodulation in these two lineages (Svistoonoff *et al.,* 2003; Svistoonoff *et al.,* 2004), this pattern of differential recruitment of paralogous subtilases for convergent function counters the ortholog conjecture, that orthologous genes are more likely to be recruited for similar functions (Kondrashov *et al.,* 2002; Gabaldon & Koonin, 2013). Further, we show through synteny analysis that *SBTM1* and *SBTM3,* which are induced only during AM, likely arose from *SBTM4* (mediating both nodulation and AM) during the whole genome duplication near the origin of the papilionoid legumes (Cannon *et al.,* 2010). This pattern of subfunctionalization of paralogous genes following the whole genome duplication has been proposed to account for the widespread and sophisticated nodulation in the Papilionoideae (Vanneste *et al.,*

2014). Independent recruitment of paralogous gene lineages adds a nuance to the homologous genetic basis of nodules, since different nodulating lineages have different, lineage-specific paralogs available to them.

The fourth chapter of this dissertation is the first study exploring the transcriptomics of nodulation in *Elaeagnus umbellata*. *E. umbellata* is an actinorhizal shrub in the Elaeagnaceae, an intercellularly-infected actinorhizal lineage for which nodulation has not been closely examined on a genetic level. We found multiple genes involved in IT formation in other lineages to be differentially expressed upon *Frankia* inoculation in *E. umbellata*, suggesting that there may be a common genetic basis to infection in lineages that form ITs and lineages with intercellular infection. Further, our transcriptome assembly recovered many of the *E. umbellata* genes homologous to those required for nodulation in other lineages. This mirrors findings in other nodulating lineages representing independent evolutionary origins of the symbiosis, showing deep homology in the origins of nodulation (Hocher *et al.,* 2011; Demina *et al.,* 2013).

We examined the evolutionary history of 13 of these genes using extensive taxonomic sampling, by incorporating data from fully sequenced genomes and the 1KP project as well as sequences from our *E. umbellata* assembly and from the nodulation literature. This wide sampling allowed us to identify patterns of duplication and loss in these nodulation genes. For example, we found that the *CASTOR* paralog originated in the gymnosperms rather than the angiosperms, as previously thought (Delaux *et al.,* 2013), and that a paralog of *MtMCA8,* a gene required for calcium spiking in *Medicago truncatula* (Capoen *et al.,* 2011), was specifically lost in the legumes (Fig. 4.8).

Our assembled *E. umbellata* transcript was orthologous to the gene required for nodulation in other lineages for 12 of the 13 nodulation genes we examined, replicating previous findings of deep homology between nodules in different lineages. However, we found that our assembled *E. umbellata* subtilase gene is orthologous to legume *SBTM4*, and not the *CG12* subtilase lineage required for nodulation in *Alnus glutinosa* and *Casuarina glauca* (Fagales). These findings again show that the evolutionary origins of nodulation involved differential recruitment of divergent paralogs for convergent functions. The divergence of paralogous genes involved in nodulation in different lineages could have functional consequences, and may help account for developmental and morphological differences in non-homologous nodules. If these paralogs are functionally equivalent, on the other hand, that is an important contradicition of the ortholog conjecture, which is already under doubt (Gabaldon & Koonin, 2013).

This dissertation presents nodulation as a case study of deep homology and evolutionary tinkering. Examination of the evolutionary history of the subtilase gene family in land plants found differential recruitment of deeply diverged paralogs in the independent evolution of nodulation. Exploration of the transcriptomics of nodulation in *E. umbellata* adds another independent example of these processes, and our transcriptome assembly recovered orthologs of multiple CSSP genes involved in nodulation in other lineages. These findings contribute to our understanding of the evolutionary origins of nodules, a complex, symbiotic organ that has shaped the world.

**References:**


**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*(7), p.e11630.

**Capoen, W., Sun, J., Wysham, D., Otegui, M.S., Venkateshwaran, M., Hirsch, S., Miwa, H., Downie, J.A., Morris, R.J., Ané, J.M. and Oldroyd, G.E.,** 2011. Nuclear membranes control symbiotic calcium signaling of legumes. *Proceedings of the National Academy of Sciences*, *108*(34), pp.14348-14353.

**Charpentier, M. and Oldroyd, G.,** 2010. How close are we to nitrogen-fixing cereals?. *Current opinion in plant biology*, *13*(5), pp.556-564.

**Delaux, P.M., Séjalon-Delmas, N., Bécard, G. and Ané, J.M.,** 2013. Evolution of the plant–microbe symbiotic 'toolkit'. *Trends in plant science*, *18*(6), pp.298-304.

**Demina IV, Persson T, Santos P, Plaszczyca M, Pawlowski K. 2013.** Comparison of the nodule vs. root transcriptome of the actinorhizal plant *Datisca glomerata*: actinorhizal nodules contain a specific class of defensins. *PloS one* **8(8):** e72442.

**Gabaldón, T., and E. V. Koonin.** 2013 . Functional and evolutionary implications of gene orthology. *Nature Reviews. Genetics* 14 : 360 – 366.

**Galloway, J.N., Schlesinger, W.H., Levy, H., Michaels, A. and Schnoor, J.L.,** 1995. Nitrogen fixation: Anthropogenic enhancement-environmental response. *Global biogeochemical cycles*, *9*(2), pp.235-252.

**Hocher V, Alloisio N, Auguy F, Founier P, Doumas P, Pujic P, Gherbi H.** 2011. Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant Physiology* Online publication.

**Kondrashov, F.A., Rogozin, I.B., Wolf, Y.I. and Koonin, E.V.,** 2002. Selection in the evolution of gene duplications. *Genome biology*, *3*(2), pp.research0008-1.

**Markmann, K. and Parniske, M.,** 2009. Evolution of root endosymbiosis with bacteria: How novel are nodules?. *Trends in plant science*, *14*(2), pp.77-86.

**McKey, D.,** 1994. Legumes and nitrogen: the evolutionary ecology of a nitrogen-demanding lifestyle. *Advances in legume systematics*, *5*, pp.211-228.

**Raven, P.H., Evert, R.F. and Eichhorn, S.E.,** 2005. *Biology of plants*. Macmillan.

**Smil, V.,** 1999. Nitrogen in crop production: An account of global flows. *Global biogeochemical cycles*, *13*(2), pp.647-662.

**Svistoonoff, S., Laplaze, L., Auguy, F., Runions, J., Duponnois, R., Haseloff, J., Franche, C. and Bogusz, D.,** 2003. cg12 expression is specifically linked to infection of root hairs and cortical cells during Casuarina glauca and Allocasuarina verticillata actinorhizal nodule development. *Molecular plant-microbe interactions*, *16*(7), pp.600-607.

**Svistoonoff, S., Laplaze, L., Liang, J., Ribeiro, A., Gouveia, M.C., Auguy, F., Fevereiro, P., Franche, C. and Bogusz, D.,** 2004. Infection-related activation of the cg12 promoter is conserved between actinorhizal and legume-rhizobia root nodule symbiosis. *Plant physiology*, *136*(2), pp.3191-3197.

**van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., Santuari, L., Cao, Q., Sharma, T., Shen, D. and Roswanjaya, Y.,** 2018. Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proceedings of the National Academy of Sciences*, p.201721395.

**Vanneste K, Maere S, Van de Peer Y. 2014.** Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **369(1648):** 20130353.