Please cite this article in press as: Zhang et al., Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis, Plant Communications (2023), https://doi.org/10.1016/j.xplc.2023.100671

Plant Communications
Research article

CellPress
Partner Journal

Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis

Yu Zhang[1,17], Yuan Fu[1,15,16,17], Wenfei Xian[1,17], Xiuli Li[1,17], Yong Feng[1], Fengjiao Bu[1], Yan Shi[1], Shiyu Chen[1], Robin van Velzen[2], Kai Battenberg[3], Alison M. Berry[3], Marco G. Salgado[4], Hui Liu[5], Tingshuang Yi[5], Pascale Fournier[6], Nicole Alloisio[6], Petar Pujic[6], Hasna Boubakri[6], M. Eric Schranz[2], Pierre-Marc Delaux[7], Gane Ka-Shu Wong[8], Valerie Hocher[9], Sergio Svistoonoff[9], Hassen Gherbi[9], Ertao Wang[10], Wouter Kohlen[11], Luis G. Wall[12], Martin Parniske[13], Katharina Pawlowski[4], Philippe Normand[6], Jeffrey J. Doyle[14,*] and Shifeng Cheng[1,*]

[1]Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, China

[2]Biosystematics Group, Department of Plant Sciences, Wageningen University, 6708PB Wageningen, the Netherlands

[3]Department of Plant Sciences, University of California, Davis, Davis, CA 95616, USA

[4]Department of Ecology, Environment and Plant Sciences, Stockholm University, 106 91 Stockholm, Sweden

[5]Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Lanhei Road, Kunming 650201, China

[6]Université de Lyon, Université Lyon 1, CNRS, UMR5557, Ecologie Microbienne, INRA, UMR 1418, 43 bd du 11 novembre 1918, 69622 Villeurbanne, France

[7]Laboratoire de Recherche en Sciences Végétales (LRSV), Université de Toulouse, CNRS, UPS, 24 chemin de Borde Rouge, Auzeville, BP42617, 31326 Castanet Tolosan, France

[8]Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada

[9]French National Research Institute for Sustainable Development (IRD), UMR LSTM (IRD/CIRAD/INRAe/Montpellier University/Supagro)- Campus International Baillarguet, TA A-82/J, 34398 Montpellier Cedex 5, France

[10]National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, SIBS, Chinese Academy of Sciences, Shanghai, China

[11]Laboratory of Molecular Biology, Department of Plant Sciences, Wageningen University, 6708PB Wageningen, the Netherlands

[12]Laboratory of Biochemistry, Microbiology and Soil Biological Interactions, Department of Science and Technology, National University of Quilmes, CONICET, Bernal, Argentina

[13]Faculty of Biology, Genetics, LMU Munich, Großhaderner Strasse 2-4, 82152 Martinsried, Germany

[14]School of Integrative Plant Science, Sections of Plant Biology and Plant Breeding & Genetics, Cornell University, Ithaca, NY 14853, USA

[15]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

[16]University of Chinese Academy of Sciences, Beijing 100049, China

[17]These authors contributed equally to this article.

***Correspondence: Jeffrey J. Doyle (jjd5@cornell.edu), Shifeng Cheng (chengshifeng@caas.cn)**

https://doi.org/10.1016/j.xplc.2023.100671

ABSTRACT

Plant root-nodule symbiosis (RNS) with mutualistic nitrogen-fixing bacteria is restricted to a single clade of angiosperms, the Nitrogen-Fixing Nodulation Clade (NFNC), and is best understood in the legume family. Nodulating species share many commonalities, explained either by divergence from a common ancestor over 100 million years ago or by convergence following independent origins over that same time period. Regardless, comparative analyses of diverse nodulation syndromes can provide insights into constraints on nodulation—what must be acquired or cannot be lost for a functional symbiosis—and the latitude for

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

variation in the symbiosis. However, much remains to be learned about nodulation, especially outside of legumes. Here, we employed a large-scale phylogenomic analysis across 88 species, complemented by 151 RNA-seq libraries, to elucidate the evolution of RNS. Our phylogenomic analyses further emphasize the uniqueness of the transcription factor NIN as a master regulator of nodulation and identify key mutations that affect its function across the NFNC. Comparative transcriptomic assessment revealed nodule-specific upregulated genes across diverse nodulating plants, while also identifying nodule-specific and nitrogen-response genes. Approximately 70% of symbiosis-related genes are highly conserved in the four representative species, whereas defense-related and host-range restriction genes tend to be lineage specific. Our study also identified over 900 000 conserved non-coding elements (CNEs), over 300 000 of which are unique to sampled NFNC species. NFNC-specific CNEs are enriched with the active H3K9ac mark and are correlated with accessible chromatin regions, thus representing a pool of candidate regulatory elements for genes involved in RNS. Collectively, our results provide novel insights into the evolution of nodulation and lay a foundation for engineering of RNS traits in agriculturally important crops.

Key words: nitrogen-fixing root-nodule symbiosis, two competing hypotheses, phylogenomics, phylotranscriptomics, conserved non-coding elements, convergence, deep homology

Zhang Y., Fu Y., Xian W., Li X., Feng Y., Bu F., Shi Y., Chen S., van Velzen R., Battenberg K., Berry A.M., Salgado M.G., Liu H., Yi T., Fournier P., Alloisio N., Pujic P., Boubakri H., Schranz M.E., Delaux P.-M., Wong G.K.-S., Hocher V., Svistoonoff S., Gherbi H., Wang E., Kohlen W., Wall L.G., Parniske M., Pawlowski K., Normand P., Doyle J.J., and Cheng S. (2023). Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis. Plant Comm. **4**, 100671.

INTRODUCTION

Nitrogen is one of the main nutrient elements indispensable for plant growth and development, but it is not directly accessible to plants without the help of nitrogenase-containing nitrogen-fixing bacteria (Geurts and Xiao, 2016; Mathesius, 2022). Some plants can obtain ammonium effectively by accommodating nitrogen-fixing bacteria in a specialized root organ, the symbiotic nodule, allowing them to grow even in nitrogen-poor soils. How nodulation evolved is a fascinating question in its own right, but the ability of nodulating plants to acquire nitrogen without exogenous fertilizer makes understanding how plants recruited and assembled the diverse components required for functioning nodules a topic of agronomic, economic, and ecological importance.

The best-known nodulating species belong to the legume family (Fabaceae; e.g., soybean, pea, alfalfa) in the flowering plant order Fabales, but nodulation also occurs in three other orders (Fagales, Rosales, Cucurbitales). There is rich genetic, phenotypic, and eco-adaptive diversity among nodulating plants across the four orders, including biogeographic distribution, nodule ontogeny, infection mode, and formation of intracellular endosymbiosis across the different lineages (Shen et al., 2020). Most notable is the diversity of the microsymbionts: legumes (Fabales) and *Parasponia* (Cannabaceae, Rosales) associate with a phylogenetically diverse group of Gram-negative nitrogen-fixing soil bacteria collectively called rhizobia (Sprent et al., 2017; Ardley and Sprent, 2021), whereas the remaining nodulating species from Fagales, Rosales, and Cucurbitales engage with actinobacteria of the genus *Frankia* and are termed actinorhizal plants. This diversity, coupled with the fact that the four orders were distantly related in pre-phylogenetic classification systems, suggested that there could be many paths to nodulation.

A major result of early molecular phylogenetic studies was the placement of these four orders in a monophyletic "Nitrogen Fixing Nodulation Clade" (NFNC) within the large Rosid clade of angiosperms (Soltis et al., 1995). This led Soltis et al. (1995) to hypothesize that a "predisposition" for nodulation evolved in the most recent common ancestor (MRCA) of the NFNC over 100 million years ago that was either the nodulation "synnovation" (Donoghue and Sanderson, 2015) itself (single origin hypothesis; Scenario I) or an unknown precursor trait that conferred a propensity for a nodulation "syndrome" (Sinnott-Armstrong et al., 2022) to evolve independently and convergently several different times, in some cases many millions of years after the NFNC MRCA (multiple origins model; Scenario II). If a constraining predisposition trait could be identified, Scenario II would provide more hope for engineering the full nodulation syndrome in non-NFNC species.

Nodulating species are found in only 10 of the 28 NFNC plant families and are rare in most of these families; even in the Leguminosae, although in some large clades nearly all species nodulate, much of the phylogenetic diversity of the legume family is non-nodulating (Doyle, 2011). Therefore, Scenario I requires massive parallel losses of nodulation across the NFNC, and because of this, both intuitive reasoning and formal modeling studies have long favored the precursor/multiple origins Scenario II model (Doyle, 2011; Werner et al., 2014; Li et al., 2015; Battenberg et al., 2018; Kates et al., 2022). Recently, however, acceptance of Scenario I has been driven by two studies that reported genomic evidence consistent with non-nodulating NFNC species having lost the ability to nodulate rather than lacking it fundamentally (Griesmann et al., 2018; van Velzen et al., 2018). Moreover, the current absence of nodulation in many unrelated species can be explained by global reduction in the benefit of nitrogen relative to the cost of carbon as atmospheric CO_2 levels have decreased over the last ~100 million years (van Velzen et al., 2019). Acceptance of Scenario I by many in the nodulation community is also fueled by the fact that, after nearly three decades of intense searching, the precursor trait required by Scenario II has yet to be identified, although some candidates have been suggested (e.g., Soyano and Hayashi, 2014; Miri et al., 2016; Mergaert et al., 2020).

Root-nodule symbiosis (RNS) is not a single trait, but a complex association involving a number of integrated but independent genetic processes, including intracellular recognition and signaling (Fournier et al., 2018), nodule organogenesis, nitrogen responses, trophic exchanges, and bacteria accommodation (Soyano and Hayashi, 2014; Mergaert et al., 2020; Soyano et al., 2021; Mathesius, 2022). In both scenarios, key components of the syndrome were recruited from other phenomena (e.g., Doyle, 1994, 2016), notably mycorrhizal signaling (Markmann and Parniske, 2009; Wang et al., 2022), but also other pre-existing processes, such as elements of the lateral root development program (Soyano et al., 2021). In Scenario I, the diversity of nodulation is due to divergence of homologous traits over more than 100 million years. By contrast, nodulation diversity in Scenario II is a product of independent origins, and similarities in non-homologous nodules and the processes by which they form are due to convergence, including convergent recruitment of the same traits, leading to "deep homology" (Shubin et al., 2009). Regardless, however, commonalities shared by unrelated nodulating species represent potential constraints on the process of nodulation—elements that are required for nodulation and are present in all nodulation symbioses either because they are features inherited from the only originator of nodulation or because they had to be recruited convergently to build an effective nodulation symbiosis.

With a clearer understanding of the comparative biology of nodulation across the NFNC, it may be possible to identify the features that make the NFNC unique among plants in generating numerous lineages containing species that fix nitrogen in symbiotic partnership with bacteria that would otherwise be recognized as enemies (Parniske, 2018). Greater knowledge of non-nodulating NFNC taxa would also be useful (e.g., Tokumoto et al., 2020), particularly lineages in which nodulation has clearly been lost (Billault-Penneteau et al., 2019), as these can provide a baseline for assessing genome evolution in species whose absence of nodulation could be due to either loss or primary absence.

In this study, we have filled in some key gaps with three new genomes and 19 transcriptomes and performed phylogenomic and phylotranscriptomic analyses within the NFNC. We explored genomic variation, gene expression changes, and regulatory sequences that are candidates for driving the origin(s) and diversification of plant nodulation.

RESULTS

New genome and transcriptome sequences from NFNC species

Actinorhizal plants have been underrepresented relative to Fabaceae in availability of genomic and transcriptomic data, which is unfortunate given their phylogenetic diversity. To fill in this gap, we added three plant species to our previous sampling (Griesmann et al., 2018): (1) a non-nodulating member of Fagales (*Fagus sylvatica*; also recently sequenced by Mishra et al., 2018, 2021), (2) the non-nodulating *Dryas octopetala* to complement nodulating *Dryas drummondii* (Billault-Penneteau et al., 2019) and form a contrasting comparison pair within the same Rosales genus, and (3) the nodulating plant *Purshia tridentata* (also from the Rosales; Figure 1 and supplemental Table 2).

We sequenced these genomes using traditional shotgun short-read sequencing technologies with hierarchical DNA libraries of varied insert size (see m'aterials and m'ethod's), resulting in 361, 48, and 130 Gb of sequencing data for *D. octopetala* (estimated genome size 257 Mb), *F. sylvatica* (497 Mb), and *P. tridentata* (244 Mb). The two tree species are highly heterozygous, leading to relatively fragmented genome assemblies, even though more than 95% of BUSCO core genes are covered by the assemblies (supplemental Table 2). We obtained 28 191 (*D. octopetala*), 23 155 (*P. tridentata*), and 35 140 (*F. sylvatica*) annotated gene models from the assembled genomes of the three species. All 227 canonical legume symbiosis-related genes (Roy et al., 2020; van Velzen et al., 2018) (supplemental Table 3) were identified in the three genomes by sequence alignment and comparison of target genes. However, the key nodulation regulator *NIN* is a pseudogene in the non-nodulating *D. octopetala*, having experienced a deletion of 69 nucleotides (21 amino acids) relative to the intact gene in the congeneric nodulator, *D. drummondii* (supplemental Figure 1). This contrasts with the other non-nodulator, *F. sylvatica*, in which *NIN* is intact, as in other non-nodulating Fagales.

We also sequenced new transcriptomes from 20 phylodiverse nodulating and non-nodulating NFNC species (supplemental Table 14; supplemental Figure 7), emphasizing root and nodule samples. We failed with some (non-model) species, particularly in nodule sample collection, because of difficulty in tissue culturing followed by RNA extraction (supplemental Table 14; supplemental Figure 7). However, we generated high-quality RNA-seq datasets from at least one representative nodulating species from each of the four orders in the NFNC, including *Casuarina glauca* (Fagales), *Datisca glomerata* (Cucurbitales), *P. tridentata* (Rosales), and *Medicago truncatula* (Fabales) (Figure 4A). We also built a complete treatment collection involving growth conditions with and without exogenous nitrate and with and without inoculation with symbiotic bacteria (B+/N−, B−/N+, B+/N+, B−/N−) in mature root tissues of two closely related comparison pairs: *Alnus* and *Betula* from Fagales (Figure 4C) and *Datisca* and *Begonia* from Cucurbitales (supplemental Figure 8). Pairwise correlations indicated high-quality and consistent RNA-seq datasets within and between species (supplemental Figure 7B).

No evidence for convergence of individual genes

Proteins recruited for nodulation might evolve new, nodulation-specific amino acids. In Scenario II, such changes could occur convergently, such that nodulating species would share sites not found in non-nodulating relatives. We implemented a test that Parker et al. (2013) used to show convergence at a small number of amino acids in hundreds of genes across the genomes of echolocating mammals (bats and dolphins). We included genomes of 30 nodulating species and 50 closely related non-nodulating species within the NFNC, as well as 8 outgroup species (Figure 1 and supplemental Table 1). The overwhelming majority of tested orthologous genes (4412/4413), including all 54 symbiosis-related

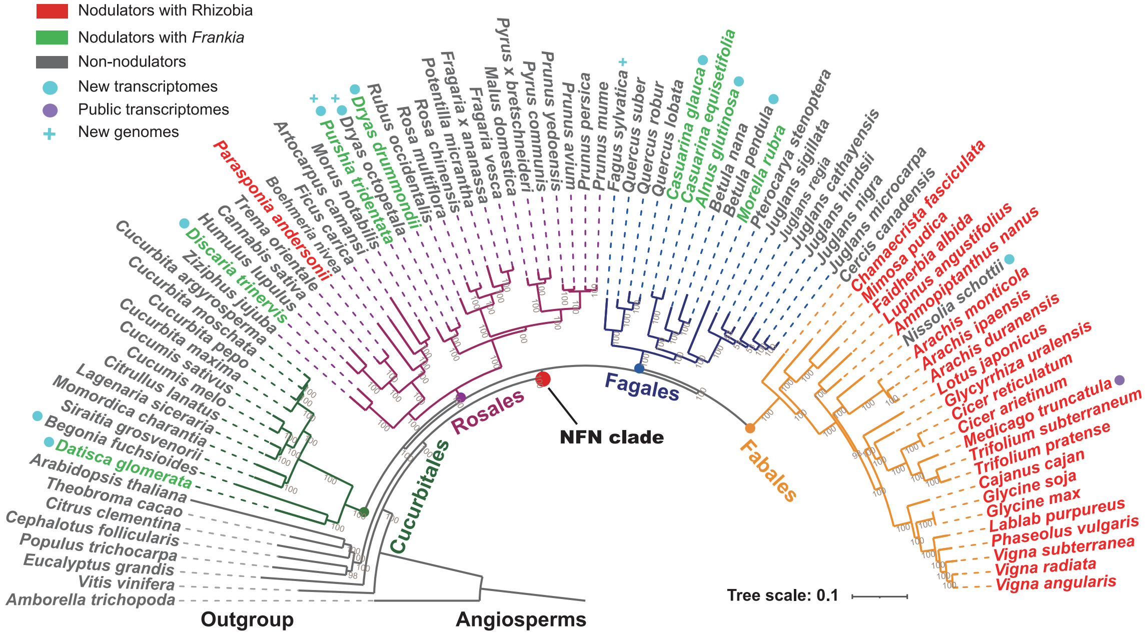

Figure 1. Genome and transcriptome datasets used in this study summarized in a phylogenetic framework

The 88 genomes and representative transcriptomes used in this study are shown in a phylogenetic tree. The phylogenetic tree presented here for the 88 species, including 3 newly sequenced species, was reconstructed using a concatenated super matrix (47 473 amino acids) of 824 one-to-one orthologs present in all species. The tree was built with the IQ-TREE ML algorithm (Nguyen et al., 2015) using the JTT+F+R6 model with 10 000 bootstraps.

genes, generated gene trees consistent with the phylogenetic pattern of the species tree, as expected for orthologous genes (Figure 2 and supplemental Table 4). Evidence for convergent evolution was sought by comparing support of each locus for this topology vs. a topology in which nodulating species were forced to form a monophyletic group. No loci with a significant convergent nodulation signal were found. The method of Parker et al. (2013) is known to overpredict convergence (Thomas and Hahn, 2015), suggesting that our conclusion of a lack of convergence is robust.

Gene family expansion

The origin of a complex novel phenomenon such as nodulation could involve the expansion of gene families; in the case of independent origins, expansion could involve some of the same families (Merényi et al., 2020). This was tested previously on a smaller subset of species (Griesmann et al., 2018; Zhao et al., 2021), revealing evidence of expansion that did not provide definitive support for either single- or independent-gain scenarios. We therefore performed an association test on the significance of gene copy numbers between nodulating and non-nodulating species on a larger scale (see m·aterials and m·ethods) (supplemental Tables 5–7). For candidate orthologous groups detected (53 657 OGs) across 88 species with gene annotations, we found that 96 gene families had experienced expansion in nodulating species compared with non-nodulating species (*t*-test $p < 0.01$, difference in average copy number >1, supplemental Tables 6 and 7), and almost all of them were order-specific gene expansions that showed no convergence among different nodulating lineages. Furthermore, of these 96 expanded gene families, 8 encoded proteins were involved in the symbiosis (Figure 3A and supplemental Table 7) from legumes, which could be explained by previous observations that complex ancestral polyploidization of the legumes may have led to duplication of symbiosis genes (Li et al., 2013; Koenen et al., 2021; Zhao et al., 2021). Expansion of different gene families in diverse nodulating lineages could be due to refinement of nodulation after a single origin (Scenario I) or to independent origins of nodulation (Scenario II); more complex scenarios involving loss of family members could also be envisioned, particularly under Scenario I.

We next sought to identify orthologs that had experienced either loss or pseudogenization in non-nodulating species. To avoid misinterpretation of the presence/absence variation results due to variable quality of genome data derived from different studies (in sequencing, assembly, annotation, or alignment), we used a set of cutoffs to define ortholog presence in nodulating species (50%–100% present, 6 conditions) and ortholog absence in non-nodulating species (50%–100% absent, 6 conditions), resulting in 36 conditions (supplemental Table 12). We ultimately defined multiple losses of orthologous groups in non-nodulating species as those present in at least 70% of nodulating species and absent in at least 60% of non-nodulating species, resulting in the identification of 461 orthologs, including previously reported genes (*NIN* and *RPG*) and three additional symbiosis genes (*IAG12*, *CNGC15a*, *BZF*) through a gene-trait presence/absence variation association study (Figure 3A and supplemental Table 11). These genes play important roles in early signaling (*CNGC15a*), rhizobia/*Frankia* infection (*RPG*, *IAG12*), and nodule organogenesis (*NIN*, *BZF*).

Sub-/neofunctionalization of the key nodulation-related gene *NIN* across the NFNC

We performed genome-wide searches for underlying genomic novelties from the protein-coding orthologous groups that were specific to the NFNC, then evaluated their association with nodulation. We integrated phylogenetic analysis for each gene family

Please cite this article in press as: Zhang et al., Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis, Plant Communications (2023), https://doi.org/10.1016/j.xplc.2023.100671

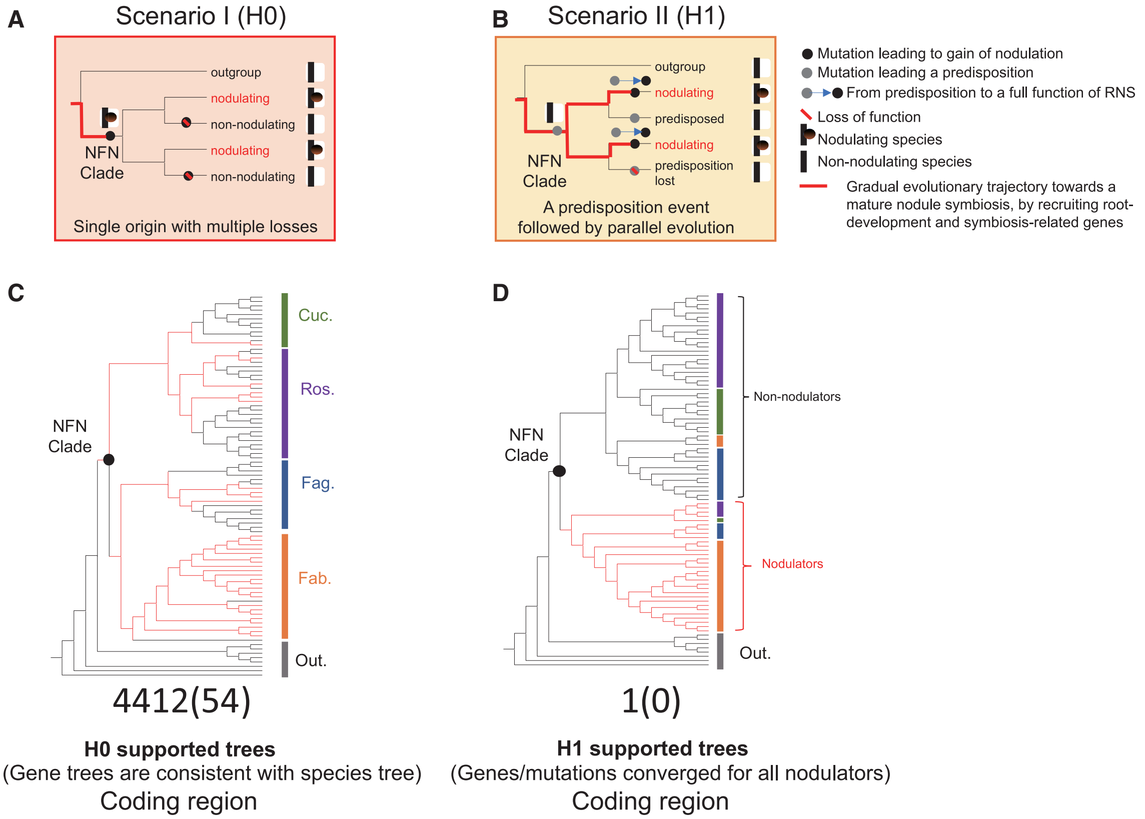
A
Scenario I (H0)
outgroup
nodulating
non-nodulating
nodulating
non-nodulating
NFN Clade
Single origin with multiple losses
B
Scenario II (H1)
outgroup
nodulating
predisposed
nodulating
predisposition lost
NFN Clade
A predisposition event followed by parallel evolution
Mutation leading to gain of nodulation
Mutation leading a predisposition
From predisposition to a full function of RNS
Loss of function
Nodulating species
Non-nodulating species
Gradual evolutionary trajectory towards a mature nodule symbiosis, by recruiting root-development and symbiosis-related genes
C
NFN Clade
Cuc.
Ros.
Fag.
Fab.
Out.
4412(54)
H0 supported trees
(Gene trees are consistent with species tree)
Coding region
D
NFN Clade
Non-nodulators
Nodulators
Out.
1(0)
H1 supported trees
(Genes/mutations converged for all nodulators)
Coding region

Figure 2. Phylogenetic inference of two evolutionary scenarios for the origin of RNS, based on thousands of gene orthogroups: two evolutionary scenarios were established

(A) Scenario I. Single origin with multiple independent losses. This model predicts that novel gene(s) or mutations, which were sufficient to enable establishment of a functional RNS, occurred at (exclusively in the ancestral state) or before (previously, via a long stepwise evolutionary journey) the common ancestor of the NFN clade.

(B) Scenario II. Predisposition followed by parallel evolution. A common predisposing mutation (genomic innovation) took place in the common ancestor of the NFN clade, followed by a (series of) decisive novel secondary mutations that permitted a functional nodule symbiosis to arise in parallel in different lineages.

(C and D) A tree topology inference method and an alignment method were used to identify which gene families followed Scenario I (left) or Scenario II (right) proposed here and to detect convergence signals (no convergence signal was treated as H0, as indicated in Scenario I) in orthologous sequences of nodulating species from different lineages. A total of 4412 genes support H0, 54 of which are symbiosis genes. A single, non-symbiosis gene supports H1.

and a Reciprocal Best Hit (RBH) search for each family member (supplemental Tables 8–12) to detect rapidly evolving genes, gene sub-clusters, or sequence neo-functionalization potentially specific to the common ancestor of the NFNC. This led to identification of 37 orthologous gene families or sub-families that are absent in all outgroup taxa but present in at least 60% of NFNC species (supplemental Tables 8–10). Among the 37 orthologous gene clusters, *NIN* stands out as the only one whose involvement in nodule symbiosis has been functionally validated (supplemental Table 10). The 37 candidate genes were further evaluated individually for significant correlations between the number of species with the gene and whether those species nodulate. Again, *NIN* showed the most significant association with the nodulating phenotype (Figure 3C and supplemental Table 13). We reconstructed the phylogenetic tree for all detected NIN and NIN-like protein (NLP) family members in the genome and transcriptome datasets (including 1KP transcriptomes) (Figure 3B) and obtained a tree that was largely consistent with a recent legume-focused study (Zhao et al., 2021) and other recent studies (e.g., Wu et al., 2022). The origin of the conserved *NLP*-like gene family, whose members serve as nitrate-responsive master regulators, can be traced back to the common ancestor of green plants. The ancestral gene subsequently experienced at least three duplication events that resulted in four *NLP* subgroups. Consistent with previous observations (Liu and Bisseling, 2020), a duplication event within *NLP* group 3 (*NLP3*) occurred early in the divergence of eudicots (Figure 3C, Duplication 3) and produced two paralogous clades, one of which includes *NIN* (in NFNC members) and its orthologs (in other eudicots).

To identify sequence changes that might have led to the functional transition from the as-yet-unknown ancestral function of *NLP* to the key nodulation roles of *NIN*, whether once or convergently, we identified nonsynonymous mutations in *NIN* that might have led to neo-functionalization specific to the NFNC. Two changes in *NIN* were identified as having occurred in the common ancestor of the NFNC. The first was the loss or likely inactivation of the nitrate-sensing motif in the nitrate-responsive domain (NRD) of *NIN* that is conserved in non-NFNC species: all

Please cite this article in press as: Zhang et al., Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis, Plant Communications (2023), https://doi.org/10.1016/j.xplc.2023.100671

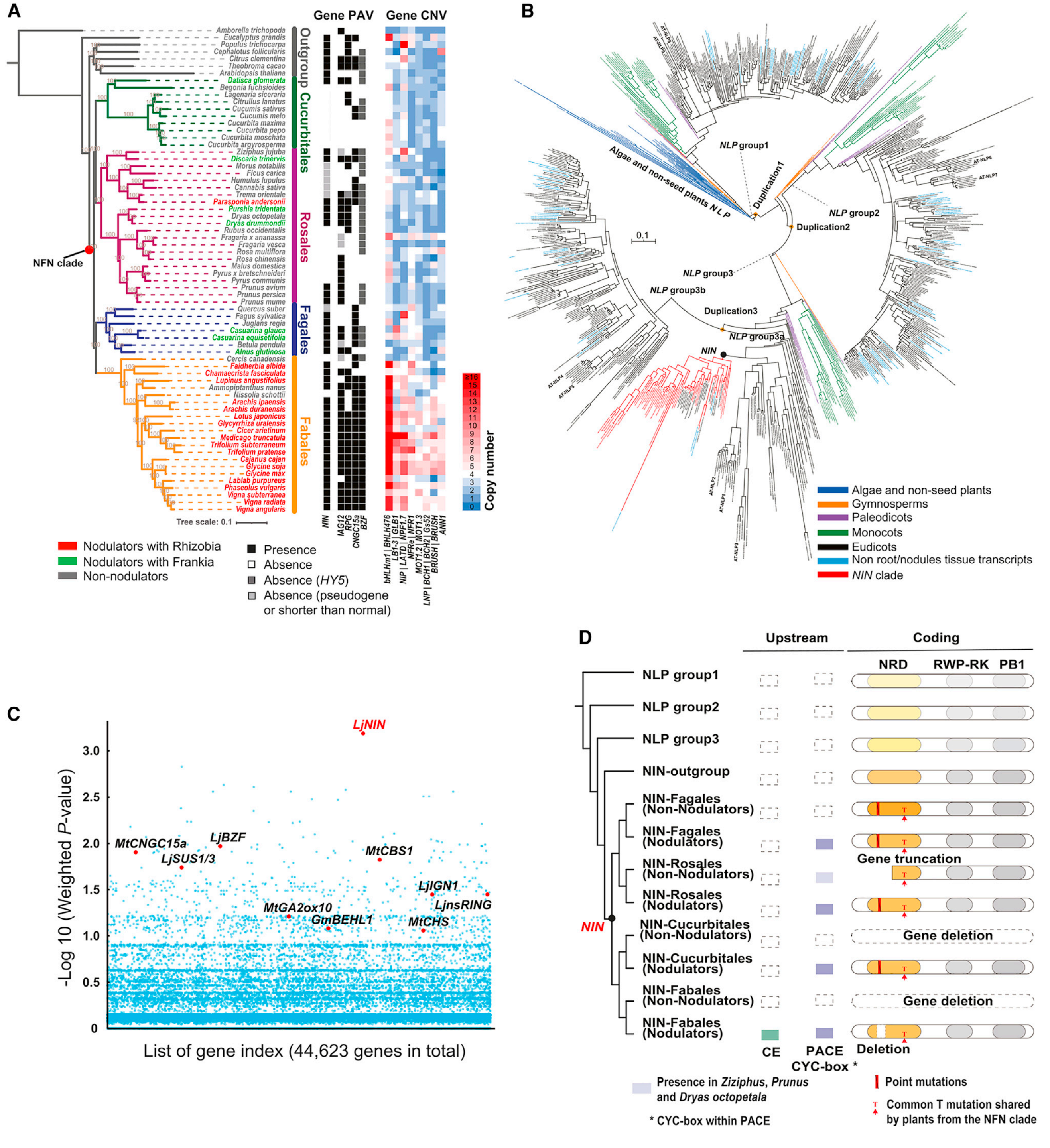

Figure 3. Presence and absence of symbiosis-related genes in the NFN clade, and a phylogenetic and structural analysis of the NIN and NIN-like gene family

(A) Occasional loss of symbiosis genes in a one-to-one ortholog detected by phylogenetic analysis in non-nodulating species and cases of gene families with putative contraction and expansion generated with OrthoFinder.

(B) ML tree of the NIN-like protein family built from a multiple sequence alignment of NIN/NLP proteins using IQ-TREE (JTT+I+G4, 1000 bootstraps).

(C) Significance test of the association between gene presence/absence of the identified orthologous groups and nodulation status among nodulating, non-nodulating, and outgroup species. Fisher's exact test was used to infer the association between presence of a target gene and symbiosis status.

(D) Structural conservation and variation in proteins and upstream sequences among different subgroups of NINs and NLPs, with evolutionary changes in the NRD region highlighted.

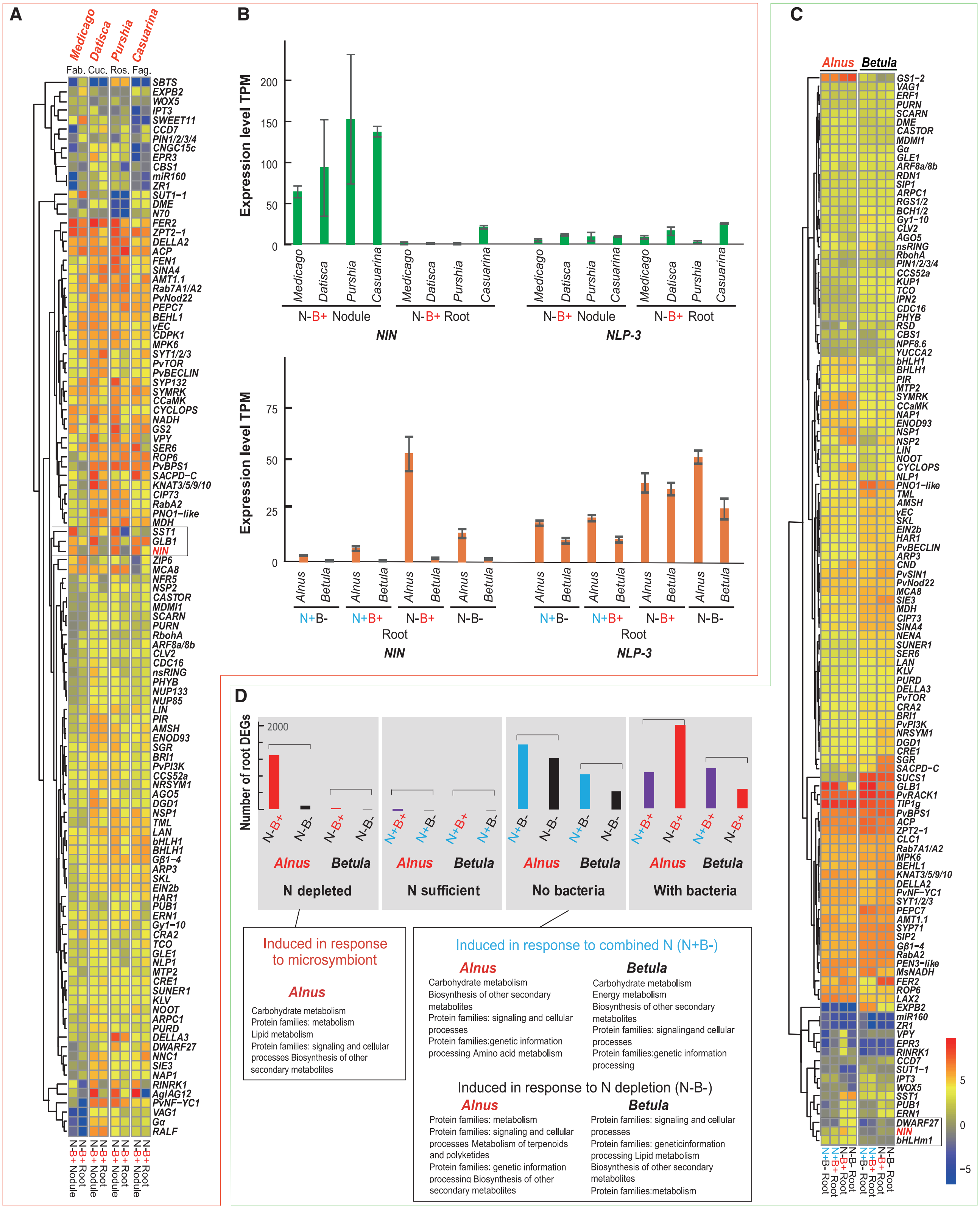

Figure 4. Differentially expressed genes that responded to nitrate treatment and/or inoculation with N_2-fixing bacteria within and between nodulating and non-nodulating plants in the NFN clade

(A) Gene expression changes in roots and nodules in response to inoculation with N_2-fixing bacteria (“N−B+” treatment) between representative nodulating species (in red) within the NFN clade (left). Hierarchical clustering of gene expression profiles from roots/nodules was performed with the R

(legend continued on next page)

Please cite this article in press as: Zhang et al., Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis, Plant Communications (2023), https://doi.org/10.1016/j.xplc.2023.100671

actinorhizal nodulating species examined here, as well as *Parasponia andersonii*, carry independent point mutations or small deletions in this motif in *NIN* compared with the NLP3 orthologs in the outgroup, and a larger deletion of this motif occurs in legume *NIN*s (Suzuki et al., 2013) (Figure 3D and supplemental Figure 3). In *Arabidopsis*, phosphorylation of a serine residue (S205) within this nitrate-sensing motif is indispensable for relocation of the protein encoded by *AtNLP7* from the cytoplasm to the nucleus and subsequent activation of downstream nitrate-responsive genes (Liu and Bisseling, 2020). By contrast, in *Medicago*, *NIN* has lost the ability to sense nitrate and localizes directly to the nucleus. The other NFNC-specific mutation occurred in consensus position 363 in the NRD domain of NIN: an amino acid transition to threonine (363T event) (supplemental Figure 3). This 363T site is the only NIN-specific mutation we found that occurred exclusively in the MRCA of NFNC. We then performed a complementation experiment with Ljnin-2 mutant plants using different LjNIN variants, including threonine to alanine and threonine to aspartic acid substitutions. Nodulation phenotypes were observed, and average nodules per plant were counted at 21 days post inoculation with *Mesorhizobium loti* MAFF303099. However, average nodules per plant did not differ significantly between the complemented mutants and wild-type plants, which suggests that the 363T site alone may not determine nodulation ability (supplemental Figure 4). Much more remains to be learned about the structure and function of NIN in both nodulating and non-nodulating species.

NIN orthologs were highly expressed almost exclusively in nodules of different nodulating species and in roots of *Alnus* when inoculated with N_2-fixing bacteria under nitrogen-depleted conditions (Figure 4B), whereas NLP3 orthologs were relatively highly expressed in nitrogen-depleted roots. Several genes were co-expressed with *NIN*: e.g., *GLB1* and *SST1* in nodules (Figure 4A) and *bHLHm1* and *DWARF27* in roots (Figure 4C).

Transcriptomics of symbiosis-related genes in non-nodulating plants

Because symbiosis is strongly inhibited by nitrate, we performed DEG analysis on roots of nodulating/non-nodulating species pairs from two orders, *Datisca* and *Begonia* from Cucurbitales and *Alnus* and *Betula* from Fagales, under N-limited and N-replete conditions without *Frankia* inoculation in order to study the regulation and activity of symbiosis-related genes in each genome (supplemental Tables 15–18). In total, 222, 150, 1227, and 479 genes were upregulated under N-limited conditions, and 714, 213, 1536, and 863 genes were upregulated during growth in the presence of nitrate in roots of *Datisca*, *Begonia*, *Alnus*, and *Betula*, respectively (Figure 4D) (supplemental Tables 26 and 27). We searched for genes that were upregulated in roots of nodulating plants but not non-nodulating plants. Genes from 19 gene families, including *NPF8.6, SST1, and LOG1,* were specifically expressed in roots of *Datisca* and *Alnus* under N-limited conditions (supplemental Figure 8) but not in those of related non-nodulating species. The nitrate transporter family protein LjNPF8.6 controls the N-fixation activity of *Lotus japonicus* nodules (Valkov et al., 2017). The sulfate transporter SST1 is crucial for symbiotic nitrogen fixation in *L. japonicus* nodules (Krusell et al., 2005). LOG1 is required for cytokinin biosynthesis and homeostasis of *M. truncatula* nodule development while also playing a negative role in lateral root development (Mortier et al., 2014).

Comparative transcriptomic analysis reveals strong conservation of an enhanced RNS gene expression network

Hundreds of protein-coding genes play roles during nodulation in *Medicago* (supplemental Table 3) (Roy et al., 2020), *Parasponia* (van Velzen et al., 2018), and diverse actinorhizal nodulators (Battenberg et al., 2018; Diédhiou et al., 2014). However, much remains to be learned about responses to nitrate or N_2-fixing bacteria across different NFNC lineages. We selected at least one nodulating species from each order of the NFN clade and compared differentially expressed genes (DEGs) between nodule and root samples of *C. glauca* (Fagales), *D. glomerata* (Cucurbitales), *P. tridentata* (Rosales), and *M. truncatula* (Fabales). There were 3088, 1854, 2517, and 4416 upregulated genes in nodules of *Casuarina*, *Datisca*, *Purshia*, and *Medicago*, respectively, but fewer (2459, 1183, 1763, 2161, and 3411) were detected in root samples (Figure 5D). We next examined the distribution of nodule-enhanced genes on phylogenetic trees of gene families of symbiosis genes. Five genes (*NIN*, *NF-YA1*, *NOOT*, *MCA8*, *NADH*) were restricted to a monophyletic orthologous gene clade (supplemental Figure 6), whereas the other genes had more complex expression profiles and showed lineage-specific upregulation of different paralogs. This result suggests that the functions of these nodules may rely on different gene sets with lineage-specific adaptations. Our analysis revealed that a minimum of six symbiosis-related genes (*CHIT5*, *NF-YA1*, *CP6*, *NFH1*, *NIN*, *RSD*), which cover almost all stages of nodulation, were consistently upregulated across all selected nodulating plant genomes (Figure 5D).

We then performed a genome-wide systematic evaluation of the connection between the conservation of orthologous groups and the conservation of gene expression level across lineages. From the four representative nodulating species described above, we classified genes as follows: shared by all four orders; shared by species from three orders; shared by species from two orders where one species was from either Fabales or Fagales and the other was from either Cucurbitales or Rosales; shared by species from two sister orders; or unique to a single species (Figure 5A). Genes shared by three to four orders or by species

package hclust. Gene expression values (TPM) were calculated and compared on the basis of one-to-one orthologous symbiosis genes across species. N, nitrogen; B, N_2-fixing bacteria; +, with; −, without.

(B) Gene expression levels (TPM) of *NIN* and *NLP-3* in nodules and roots of selected plant species under different growth conditions.

(C) Detailed comparison between the nodulating plant *Alnus* and its closely related comparator species *Betula* in Fagales under various treatments (N+B−, N+B+, N−B+, and N−). Gene expression values (TPM) were calculated and compared on the basis of one-to-one orthologous symbiosis genes across species. N, nitrogen; B, N_2-fixing bacteria; +, with; −, without.

(D) Summary of differentially expressed genes (DEGs) in roots from different treatments with nitrate and *Frankia inoculation*. The bars represent the number of upregulated genes under a given growth condition compared with the other conditions. Blue, supplied with 5 mM KNO_3; red, inoculated with *Frankia*; purple, supplied with 5 mM KNO_3 and inoculated with *Frankia*; gray, supplied with neither KNO_3 nor *Frankia*.

Please cite this article in press as: Zhang et al., Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis, Plant Communications (2023), https://doi.org/10.1016/j.xplc.2023.100671

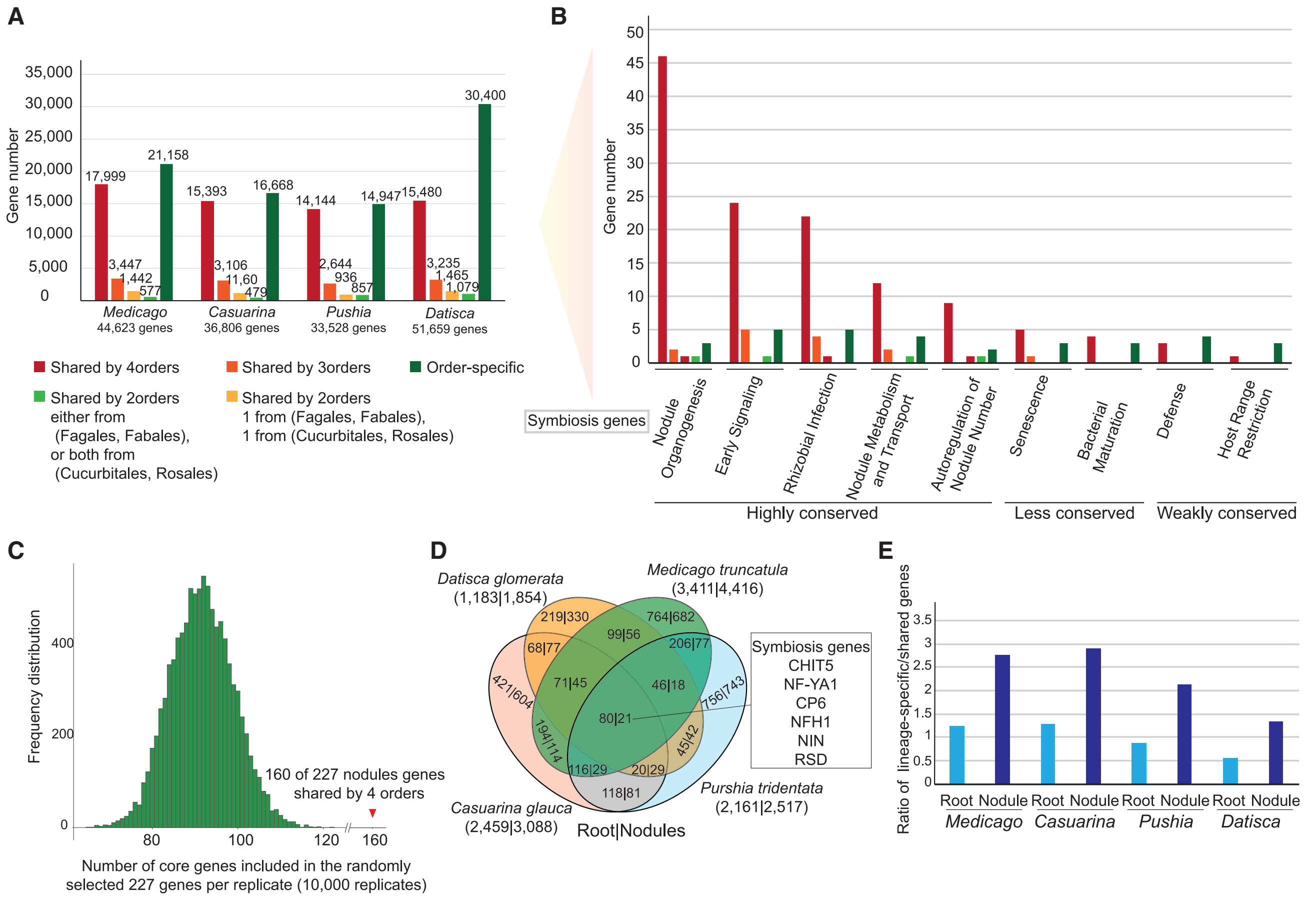

Figure 5. Comparative transcriptomic and genomic analysis

(A) Identification and catalog of shared and lineage-specific genes among four representative nodulating plants in the context of ((Fagales, Fabales), (Cucurbitales, Rosales)). Five categories were analyzed: shared by all four species; shared by three species; shared by two orders, one from (Fagales, Fabales) and one from (Cucurbitales, Rosales); shared by two orders, either both from (Fagales, Fabales) or both from (Cucurbitales, Rosales); and lineage/order-specific genes. The first three categories are inferred to be present in the most common ancestor of the NFNC. The latter two categories arose later than the NFNC MRCA or were lost from either an ancestor of two orders or the descendants of a single order.

(B) Evolutionary characterization of genes involved in different aspects of nodulation symbioses as defined in Roy et al. (2020) (Figure 2). Genes are categorized as in **(A).** Highly conserved categories are those for which the ratio of core/lineage-specific genes is ≥4.0, less conserved categories show a lower core/lineage-specific ratio, and weakly conserved categories are those for which lineage-specific genes outnumber core genes.

(C) Frequency distribution of the number of core genes included in 227 randomly selected genes (10 000 replicates); 160 out of the 227 nodulation-related genes are shared by species in 4 orders.

(D) Distribution of upregulated orthologous groups in nodules or roots from four selected nodulating plants. Numbers in Venn diagrams represent the number of orthologous groups. Numbers below species names represent the total number of nodule-upregulated genes. Numbers on the left side of vertical bars represent root upregulation data, and numbers on the right side represent nodule upregulation data.

(E) The ratio of lineage-specific/shared genes, where "shared" = genes inferred to have been present in the MRCA of the NFNC. Data from the Venn diagram in **(D)**, number of DEG+ genes shared among different species.

representing non-sister orders are most parsimoniously inferred to have been present in the NFNC MCRA. Genes found in species representing two sister orders do not provide evidence for presence in the NFNC MRCA; they could equally parsimoniously have been present in the NFNC ancestor and lost in the ancestor of the other two orders or have originated in the ancestor of the sister orders with the gene. Genes found in only a single species are inferred to have arisen at some point in the lineage leading from the species backward to the ancestor of its order.

We next determined to which of these categories the 227 *Medicago* nodule symbiosis-related target genes belonged (Roy et al., 2020). For each of the nodulation processes (e.g., early signaling, nodule organogenesis) into which Roy et al. (2020) divided these genes, we calculated the ratio of core/lineage-specific genes. Most categories were dominated by highly conserved genes (Figure 5B), even though lineage-specific genes comprise a high percentage of each of the four genomes (Figure 5A). This was particularly true for genes involved in nodule organogenesis, followed by early signaling, rhizobial infection, nodule metabolism and transport, and autoregulation of nodulation. Total gene numbers were much smaller for other categories, but lineage-specific genes dominated the defense and host-range restriction categories (Figure 5B). Overall, more than 70% (160) of the 227 nodulation-related genes (supplemental Table 2) were highly conserved genes,

Please cite this article in press as: Zhang et al., Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis, Plant Communications (2023), https://doi.org/10.1016/j.xplc.2023.100671

significantly more than expected based on randomly selected sets of 227 genes from the *Medicago* gene set (Figure 5C).

We then categorized DEGs between nodules and roots and between different lineages (supplemental Tables 14–22). For each species, the ratio of lineage-specific genes likely to have been present in the NFNC MRCA was higher for nodule DEG+ genes than for either nodule DEG− or non-DEG genes (supplemental Table 28 and Figure 5E). Moreover, in all four species, this ratio was approximately double for nodule DEG+ genes than for root DEG+ genes (supplemental Table 28 and Figure 5E). Taken together, these results suggest that recently evolved genes play a greater role in nodulation than in root biological processes. DEG− and non-DEG lineage-specific genes represent a baseline for genes that have arisen in each order and are not associated with nodulation, and the excess of DEG+ lineage-specific genes therefore suggests that many of these genes could have evolved through involvement in nodulation.

Conserved non-coding elements associated with nodulation

Recent studies have identified *cis*-regulatory elements (CREs) that play important roles in the spatiotemporal expression and regulation of symbiosis-related genes (Liu et al., 2019; Soyano et al., 2019). Here, we performed extensive identification and char acterization of conserved non-coding elements (CNEs) across 88 phylodiverse genomes by combining two pipelines: (1) reference-based Lastz-ChainNet-Roast followed by CNSpipeline (Liang et al., 2018) and (2) reference-free Progressive Cactus followed by PhastCons (Hubisz et al., 2011). Our approach differed from that of the recent study of Pereira et al., 2022 by integrating the two pipelines, combining all 88 species for whole-genome alignments followed by CNE identification (supplemental Figure 2), thereby reducing false positives incorrectly defined as "conservation."

In addition to *M. truncatula*, we selected eight representative genomes comprising one non-nodulator and one nodulator from each of the four orders, as well as one outgroup species (*Populus trichocarpa*) (supplemental Figure 12), and we anchored all the identified CNEs to the *Medicago* genome for comparison. In total, we predicted 931 454 high-quality putative CNEs (≥5 bp; corresponding to 4.16% of the *Medicago* genome) (Figure 6A and supplemental Table 29), many more than detected by Pereira et al., 2022 (supplemental Figure 13), who reported only 6729 CNEs. Furthermore, 84.49% of RNS-specific CNEs and 84.60% of NFN-specific CNEs detected by Pereira et al., 2022 were also present in our study, including the five experimentally validated CNEs in *cis*-regulatory regions (supplemental Figure 13), suggesting a high-quality CNE dataset as well as many new CNE candidates first discovered in our study (supplemental Figures 2 and 13). As expected, a large proportion (78.4%) of CNEs were located within 20 kb upstream of the 5′ UTR, within 20 kb downstream of the 3′ UTR, or in introns (Figure 6A and supplemental Table 29). Many were transposon-related sequences, indicating a rich source of CNEs derived from TEs (supplemental Table 34). Among the 931 454 CNEs, 331 153 are NFNC specific, and 15 101 of these are RNS specific. Notably, 3788 out of the 7651 nodulation-related CNEs associated with the 232 symbiosis-related genes of Roy et al. (2020) (227 protein coding genes plus 5 miRNA genes) are NFNC specific (supplemental Tables 30–33). This number is significantly larger than that obtained for randomly selected sets of 232 genes (Figure 6B), consistent with the hypothesis that at least some CNEs play roles in the RNS (Pereira et al., 2022), a trait confined to the NFNC.

To further evaluate the quality and function of the identified CNEs, we compared each category of CNEs with the PLACE and plantPAN 3.0 databases (supplemental Figure 2B). We calculated the enrichment of motifs based on a Z-score for each CNE, retrieved their hits from the databases, and found that 52.94% of the CNEs had at least one database hit, suggesting the potential functional roles of these CNEs. Notably, one nodulation-related motif, "AAAGAT," originally identified in the soybean leghemoglobin *lbc3* promoter (Stougaard et al., 1990; Fehlberg et al., 2005), ranked in the top 10 sorting by the Z-score based on the CNEs that were conserved in all nodulating plants (supplemental Figure 2B). At least 4468 CNEs contained the "AAAGAT" motif located around symbiosis genes, suggesting that an amplification of this motif in nodulating plants might have contributed to emergence of the RNS (supplemental Figure 2B).

The distribution of CNEs in the *Medicago* genome was compared (for all 931 454 CNEs and for only the 331 153 NFNC-specific CNEs) with various features associated with transcriptional regulation (the two heterochromatic repressive histone marks H3K9me2 and H3K27me1, the active mark H3K9ac, and open chromatin as assayed by ATAC-seq) (Figures 6C, 6D, and supplemental Figure 9). CNEs were enriched at H3K9ac marks around transcriptional start and end sites in both nodules and roots, potentially promoting the expression of their associated gene (supplemental Figure 10). Interestingly, we found that a significant majority of NFNC-specific CNEs overlapped and were enriched with the active marker H3K9ac and active ATAC-seq peaks (Figure 6D), suggesting that some of the NFNC-specific CNEs are CREs and play roles in initiating expression of genes involved in RNS.

Two conserved CREs were detected in our study. One was a remote *cis*-regulatory region located 20 kb upstream of the translation start site of *NIN* (supplemental Figure 10). This is consistent with the fact that the putative cytokinin-responsive elements within this region are required for triggering of *NIN* expression in the pericycle by cytokinin and are indispensable for nodule organogenesis (Liu et al., 2019). Interestingly, the other legume-specific CRE was a NIN-binding site located in the intron of *LBD16a* (supplemental Figure 11). *LBD16*a is involved in both lateral root formation and nodule org anogenesis, only the second of which is dependent on NIN (Soyano et al., 2019).

DISCUSSION

These newly sequenced genomes and transcriptomes help fill important gaps in the NFNC. New genomes from *F. sylvatica*, *D. octopetala*, and *P. tridentata* not only enhance the phylogenetic diversity of available data but also provide related nodulating/non-nodulating pairs for more robust comparative analyses. The fact that all 227 canonical legume symbiosis-related genes were found in the genomes of all three nodulating species reaffirms the fundamental role of these genes in nodulation. The discovery of a pseudogenized *NIN* gene in non-nodulating *D.*

Please cite this article in press as: Zhang et al., Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis, Plant Communications (2023), https://doi.org/10.1016/j.xplc.2023.100671

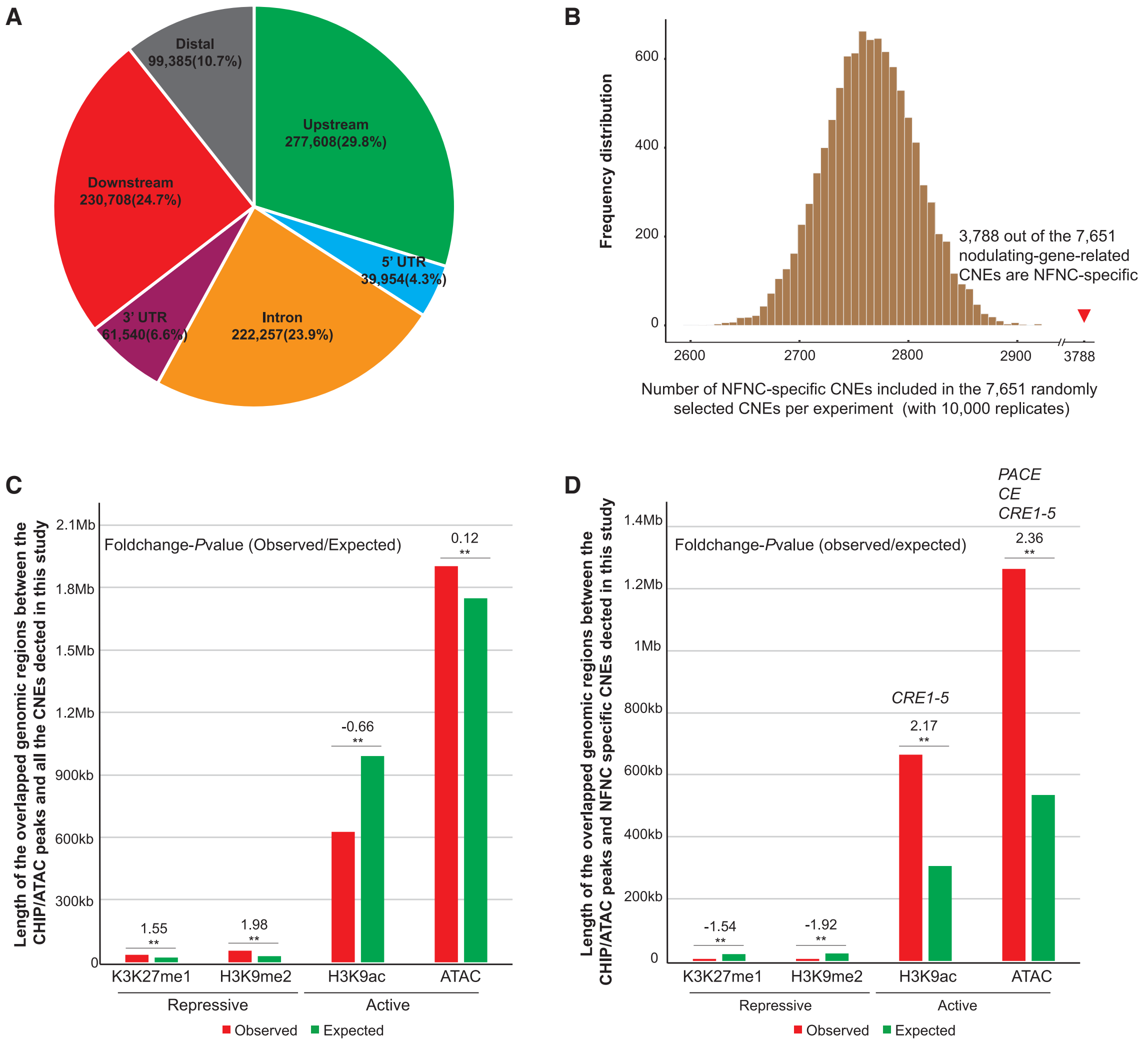

Figure 6. Identification, catalog, and evolutionary and functional analyses of CNEs in the NFNC clade

(A) Summary statistics on the distribution of CNEs identified by whole-genome alignments across 88 genomes (see materials and methods) using the *M. truncatula* genome as a reference. Distal region, 20 kb away from transcription start site; upstream, 20 kb upstream of the gene; downstream, 20 kb upstream of the transcription stop site.

(B) Frequency distribution of the number of NFNC-specific CNEs out of 7651 randomly selected CNEs for each simulated experiment (10 000 replicates in total). The red triangle indicates the number of NFNC-specific CNEs out of 7651 nodulation gene–related CNEs.

(C) Comparison between genomic regions with histone markers/ATAC peaks and all CNEs detected in this study. Significance was calculated with GAT (**$P < 0.001$). The ChIP/ATAC dataset was downloaded from https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR. Expected: frequency distribution of the length of genomic regions overlapping with the ChIP/ATAC peaks in a simulation experiment in which we randomly selected the same number of CNEs from the whole genome (repeated 10 000 times).

(D) The same analysis as in **(C)**, but for NFNC-specific CNEs only. As controls, those CNEs with functional validations, i.e., the functional NFNC-specific CNEs PACE, CE, and CRE1-5, were overlapped with ATAC peaks, and CRE1-5 was also overlapped with H3K9ac peaks.

octopetala also underscores its role as one of the few genes whose function appears to be unique to nodulation in the NFNC (Griesmann et al., 2018; van Velzen et al., 2018). The successful sequencing of new transcriptomes from 20 representative nodulating and non-nodulating species across four orders in the NFNC provides a solid foundation for future studies to explore the diversity and similarity of nodulation at the molecular level.

For many years, it has been accepted that nodulation is a complex phenomenon, many of whose components have been recruited from pre-existing processes that are widely shared among nodulating and non-nodulating plants both within and outside the single angiosperm clade, NFNC, in which nodulating species occur. What is unique about nodulation is therefore the assembly of these disparate components into a functional association in which bacteria that would otherwise be pathogenic are attracted, their entry past highly effective organ- and cell-level defenses is facilitated, existing developmental programs are co-opted to build a home for them, nutrition is provided, and the nitrogen-containing fruits of their labor are harvested and transported into the plant. How did this assembly process occur? Were there intermediate states, analogous, for example, to the origin of key features of avian wings prior to their use in flying (Uno and Hirasawa, 2023). If so, might some extant non-nodulating species be proto-nodulators? Finding

Please cite this article in press as: Zhang et al., Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis, Plant Communications (2023), https://doi.org/10.1016/j.xplc.2023.100671

examples of non-nodulating species that have some but not all components of nodulation would be an exciting development in efforts to engineer symbiotic nitrogen fixation outside of the NFNC and argues for investment in research on non-nodulating relatives of nodulators across the NFNC. Such plants would represent steps toward nodulation under the multiple-origins Scenario II. Alternatively, in Scenario I, they would represent plants that had lost the full capacity for nodule production but in which cessation of nodulation had not led to loss of all nodulation-related features. The latter case would be of particular interest if loss were avoided because some selective benefit were still provided by the retained features.

Development of tests that distinguish between homology and convergence of nodulation is very difficult owing to the complexity of deep homology, itself an evolving concept in evolutionary theory, with recent emphasis on combinatorial "character integration mechanisms" as the generators of evolutionary novelty (DiFrisco et al., 2022). As such, it is possible that there are no genes or GRNs truly unique to nodulation. In any case, "unique" is a relative term when so much of gene evolution involves duplication and neo- or subfunctionalization and given the complex and often overlapping roles that individual genes perform, such that a high percentage of the genome is expressed even in single cell types (e.g., Coate et al., 2020). Here, we have increased genomic and transcriptomic sampling of phylogenetically diverse nodulating and non-nodulating taxa to identify shared and divergent elements of nodulation. Even after our efforts to identify other key genes, the transcription factor *NIN* remains the best candidate for a gene specifically associated with nodulation. As our results show, much remains to be learned about its structure and function across the NFNC, and achieving a greater understanding of this key gene will be important regardless of whether nodulation evolved once or multiple times.

A major debate in evolutionary biology concerns the relative contributions to evolutionary innovation of novel regulatory sequences vs. protein coding genes (Stern, 2000; Carroll, 2005, 2008). In the case of nodulation, specific CRE changes in *LjLBD16* and *MtSCR* have been shown to be required for legume–rhizobial symbiosis (Dong et al., 2021). Recruitment of genes for nodulation likely involved the evolution of new regulatory sequences associated with genes that also retained their original functions, making characterization of such sequences critical to understanding not only the process of recruitment—including potentially distinguishing between the two evolutionary scenarios (Doyle, 2016)—but also how gene regulatory processes are modified when nodulation is added to existing developmental programs. CNEs include regulatory elements (e.g., Schmitz et al., 2022), and a recent study identified many such elements associated with nodulation and validated one such element experimentally (Pereira et al., 2022).

Our novel pipeline, which combines the merits of Lastz-based and Cactus-based pipelines, identified a much larger set of NFNC-specific and RNS-specific CNEs, many of which are associated with symbiosis-related genes. This provides a vast pool of candidates in the search for nodulation-associated regulatory elements (Pereira et al., 2022), including those currently associated with nodulation and those that could provide fossil evidence of former nodulation ability in non-nodulating species (Doyle, 2016). However, CNE identification and characterization is still a challenging task, requiring additional high-quality chromosome-level genome assemblies and annotation, as well as functional validation for the identified CNEs in the context of regulatory genomics (e.g., Pereira et al., 2022).

We found particularly striking the significant overlap and enrichment of NFNC-specific CNEs with the active marker H3K9ac and ATAC-seq peaks, indicative of open chromatin regions. This association implies that a portion of these NFNC-specific CNEs may be actively involved in initiating the transcription of genes related to RNS. This adds an additional layer of complexity to our understanding of the regulatory mechanisms that drive RNS, emphasizing the role that these CREs may play in modulating gene expression in this context (Liu et al., 2019). Future research may aim to dissect the exact nature of this regulation and the specific genes that are targeted, providing deeper insights into the regulatory landscape of RNS.

The phylogenetic diversity of nodulating species provides an opportunity to explore the many different solutions these lineages have evolved for attracting and housing bacteria. The question of whether the various modules were assembled once or many times, as fascinating as it is, pales in significance compared with determining in detail how nodulation can produce the same result in taxa some of which diverged over 100 million years ago, involving diverse bacteria housed in structures that are highly divergent despite developmental commonalities. If Scenario I is correct, then differences in how nodulation occurs in such lineages provide information on the robustness of an ancestral symbiosis. If Scenario II is correct, then convergent similarities represent the requirements for establishing a nodulation symbiosis *de novo*. In either case, there is a clear need for additional phylogenomic and phylotranscriptomic sampling and deep comparative biology analysis of protein-coding genes, gene expression, and CNEs across the entire NFNC, as proposed in The Legume Nodulation and NFNC Phylogenomics v2.0 Project (https://www.legumedata.org/beanbag/68/issue-68-legume-genome-sequencing-consortium).

METHODS

Plants and bacteria

Seeds of *Alnus glutinosa* (harvested from a tree growing on the bank of the Rhône River in Lyon, France) and *Betula pendula* (Vilmorin, La Ménitré, France) were sown, left to germinate, and grown for 6 weeks in a sterile soil/vermiculite substrate (1:1, v/v) in a greenhouse with a 16-h light/8-h dark regime and temperatures of 21°C (light) and 16°C (dark). Seedlings were transferred to 500 ml of Fåhraeus medium (Fåhraeus, 1957) with or without 5 mM KNO_3 in opaque plastic pots (8 seedlings per pot) and grown for 4 weeks before inoculation (or not). Inoculation was performed using syringed 18-day-old *Frankia alni* ACN14a culture in BAP-PCM medium that contained 5 mM NH_4Cl (pH 6.2) (Schwencke, 1991). There were therefore four treatment groups: (1) seedlings with 5 mM KNO_3 and *Frankia*, (2) seedlings with 5 mM KNO_3 without *Frankia*, (3) seedlings without KNO_3 and with *Frankia*, (4) seedlings without KNO_3 and without *Frankia*. After inoculation, plants were grown for 22 days. Then, roots—nodulated or not—were harvested and frozen in liquid nitrogen.

Cuttings of *Begonia fuchsioides* were obtained from the Nymphenburg Botanical Garden in Munich (Germany) in 2015 and grown in a growth cabinet under low light (two fluorescent lights removed from the cabinet) at 16-h light (18.5°C)/8-h dark (12°C). They were grown in a 1:1 (v/v) mixture of sand (grain size 1–1.2 mm)/Stender Vermehrungssubstrat A 210 (Stender AG, Schermbeck, Germany). Plants were watered regularly with deionized water and once a week with $^1/_4$ Hoagland's solution (Hoagland

Please cite this article in press as: Zhang et al., Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis, Plant Communications (2023), https://doi.org/10.1016/j.xplc.2023.100671

and Arnon, 1950)—either without nitrogen (minus N samples) or with 10 mM KNO_3 (plus N samples). Material from cuttings was harvested after approximately 3 months of growth and shock-frozen in liquid nitrogen. Roots and leaves were frozen separately. Entire root systems were harvested, except for the top ~1 cm, which showed secondary growth and lignification.

C. glauca *Sieb*

Ex Spreng seeds were purchased from the Australian Tree Seed Centre (CSIRO, Australia) and grown as described by Auguy et al. (2011). The compatible bacterial strain *Frankia casuarinae* CcI3 (Zhang et al., 1984) was used to inoculate *C. glauca* plantlets as described previously (Franche et al., 1997). Seedlings were transferred to a soil/vermiculite substrate (4:1, v/v) in a greenhouse under natural light at temperatures between 25°C and 30°C. After 1 month, seedlings were transferred to pots containing 500 ml of a modified Broughton and Dillworth (BD) medium (Broughton and Dilworth, 1971) supplemented with nitrogen (5 mM KNO_3) and cultivated in a growth chamber under the following conditions: 25°C, average 45% humidity, 16-h photoperiod, and photosynthetically active radiation (PAR) of 150 μmol m^{-2} s^{-1}. After 3 weeks, plants were starved of nitrogen for 1 week before inoculation with the symbiotic bacteria. Plants were inoculated with 10 ml of a concentrated *F. casuarinae* CcI3 suspension at a density corresponding to ~25 μg ml^{-1} of protein (Franche et al., 1997). After 2 h of contact, plants were placed in pots containing 490 ml of BD medium without nitrogen and 10 ml of the CcI3 suspension. Nodule initiation was monitored twice per week. Six conditions were sampled: leaves, non-inoculated roots with or without KNO_3, inoculated roots (2, 4, or 8 days after inoculation), and 3-week-old nodules. All samples were immediately frozen in liquid nitrogen.

D. glomerata *(C. Presl) Baill*

Seeds originating from plants growing at Gates Canyon in Vacaville, CA, were brought to Europe in 1995 by Katharina Pawlowski and propagated in greenhouses ever since. For roots, plants were grown in axenic culture. Seeds were surface sterilized by incubation in 25% H_2SO_4 for 30 s, followed by two washes with sterile deionized water. They were then incubated for 5 min in 2.5% NaOCl and washed six times with sterile deionized water. Seeds were transferred to vertical Petri dishes with $^1/_4$ Hoagland's solution with 10 mM KNO_3 and 1% agar or $^1/_4$ Hoagland's solution without N and 1% agar. Roots were harvested after 7 weeks of cultivation. For nodulation, seeds were transferred to pots with germination soil (Såjord, Weibull Trädgard AB, Hammenhög, Sweden) covered by sand (1.2–2 mm quartz; Radasand AB, Lidköping, Sweden) in the greenhouse. Greenhouse conditions were 13-h light (23°C)/11-h dark (19°C) and 200 μmol m^{-2} s^{-1} PAR. When seeds had germinated, seedlings were transferred to small pots with germination soil. For infection with *Candidatus* Frankia datiscae Dg1 (Persson et al., 2011), plantlets of 10-cm height were transferred to larger pots (diameter 15 cm) containing a 1:1 (v/v) mixture of sand (0–2 mm quartz; Rådasand AB) and germination soil (bottom third), sand (middle third), and a 1:1 (v/v) mixture of sand and germination soil (top third). Inoculum was applied to the root system during transfer in the form of *D. glomerata* nodules, freshly harvested from an older inoculated plant and crushed in deionized water with a mortar and pestle. Inoculated plants were watered with $^1/_4$ Hoagland's solution without N once per week and otherwise with deionized water. Nodules were harvested 6–12 weeks after infection. For leaf production, *D. glomerata* seeds were germinated after 2 weeks of vernalization and then transferred to pots and grown in the greenhouse at LMU Munich (18°C/12°C, 16-h/8-h day/night cycles, 150 μmol m^{-2} s^{-1} PAR). The growth substrate was A210 (perlite, with 0.5 kg/m^3 of 14-10-18 NPK and sphagnum peat H3-H5 [pH 6.2]) from Stender AG. Plants were watered regularly with deionized water and once a week with $^1/_4$ Hoagland's solution with 10 mM KNO_3. Leaves were harvested after 6–12 weeks of growth.

P. tridentata seedlings were purchased in 2012 from Cornflower Farms Native Nursery in Elk Grove, CA, where they had been grown in soil mix with slow-release nitrogen fertilizer. The nursery soil mix was replaced with a pasteurized soil mix, sand:fir bark:peat moss:perlite. Seedlings were maintained on deionized water before and after inoculation. One week after transplant, the seedlings were inoculated with an aqueous suspension of rhizosphere soil that had been excavated from mature shrubs of *Ceanothus velutinus* at Sagehen Creek Field Station, Truckee, CA, and stored at 4°C. Nodulation was observed at 95 days post-inoculation. Thereafter, the nodulated *P. tridentata* plants were maintained as stock plants in greenhouse conditions at the University of California, Davis, CA, and were irrigated with deionized water, except for two supplements of Hoagland's solution: 10 ml of $^1/_2$ strength Hoagland's solution/5-l container (23/09/2014; 26/02/2015). The $^1/_2$ strength Hoagland's solution in the greenhouse contained 150 mg N/l. Thus the 10-ml supplement was equivalent to 1.5 mg N/5-l container. In July 2015, to test whether recent application of nitrogen affected nodule-lobe or root-tip gene expression, half the plants from which nodules and roots were collected were given two supplements of Hoagland's solution, 20 ml of $^1/_2$ strength Hoagland's per 5 l container; the other half of the plants used for sampling did not receive any supplement. Twenty milliliters of the greenhouse $^1/_2$ strength Hoagland's solution (150 mg/l N) is equivalent to 3 mg N/5-l container. Samples collected from the *P. tridentata* root systems consisted of mature nodule lobe tips (i.e., the portion of individual perennial nodule lobes from the current growing season that had reached their full extent for the season) and root tips (<4 cm length). Nodule lobe tips and root tips were rinsed in sterile deionized water, immediately frozen in liquid nitrogen, and stored at −80°C.

RNA isolation

For *A. glutinosa* and *B. pendula*, RNA was extracted as described by Alloisio et al. (2010) using the RNAeasy Plant Mini Kit (QIAGEN) and on-column DNA digestion with the RNase-free DNase set (QIAGEN). To remove any remaining DNA contamination, a second DNase treatment was performed with RQ1 RNase-free DNase (Promega, Charbonnières-les-Bains, France), followed by RNA clean-up using the RNeasy Plant Mini Kit. Purity, concentration, and quality of RNA samples were checked using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Courtaboeuf, France) or an Ultrospec 3300 Pro spectrophotometer (Amersham Biosciences, Buckinghamshire, UK) and agarose gel electrophoresis.

For *C. glauca*, two conditions were sampled within three biological replicates: 21-day-old nodules and roots were sampled from inoculated and non-inoculated plants (controls), respectively. Total RNA was purified by ultracentrifugation (Hocher et al., 2006). Residual DNA was removed from RNA samples using the Turbo DNA-free kit (Ambion) and quantified using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific). We used a pool of 28 plants for each time point. The integrity of the RNA samples was assessed using a Bioanalyzer 2100 according to the manufacturer's instructions (Agilent, Santa Clara, CA). mRNA libraries were constructed for each condition, and sequencing was performed at the MGX platform (Montpellier Genomix, Institut de Genomique Fonctionnelle, Montpellier France). The RNA libraries were constructed using the TruSeq stranded mRNA library construction kit (Illumina). Quantitative and qualitative analyses of the libraries were performed with an Agilent DNA 1000 chip and qPCR (Applied Biosystems 7500, SYBR Green). RNA was sequenced using the Illumina SBS (sequencing by synthesis) technique on a HiSeq 2000 instrument in single-read 100-nt mode. Image analysis, base calling, and quality filtering were performed using Illumina software.

For *D. glomerata* and *B. fuchsioides*, RNA isolations were performed using the Sigma Spectrum Plant Total RNA extraction kit (Sigma-Aldrich, St. Louis, MO); Polyclar AT (Serva, Heidelberg, Germany) was added to the extraction buffer (2%, w/v). On-column DNA digestion was performed with the RNase-free DNase set (QIAGEN), followed by treatment with Ambion TURBO DNase (Thermo Fisher Scientific) after isolation. RNA quality was determined using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific).

RNA was extracted from *P. tridentata samples* using either a QIAGEN RNeasy Plant Mini-Kit or, for a subset of samples, the Spectrum Plant

Please cite this article in press as: Zhang et al., Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis, Plant Communications (2023), https://doi.org/10.1016/j.xplc.2023.100671

Total RNA Kit, followed by treatment with the Ambion Turbo RNase-free DNase kit (Thermo Fisher Scientific) after isolation. The Spectrum Total RNA kit was used to enable BGI to test whether *Frankia* transcripts could be detected in nodule tissue, in addition to measurement of plant gene expression.

Genome sequencing, assembly, and annotation

Plant cultivation, DNA extractions, and RNA extractions were performed as described in Griesmann et al., 2018. The genomes of *P. tridentata*, *D. octopetala*, and *F. sylvatica* were sequenced using Illumina technology (HiSeq 2000 and HiSeq 4000). A hierarchical DNA library strategy was employed, which included multiple paired-end libraries with insert sizes ranging from 170 to 800 bp and mate-pair libraries using large DNA fragments with insert sizes of 2–20 kb. The reads were filtered using SOAPfilter (Luo et al., 2015) following a strict quality control protocol. The genome size was estimated using 17-mer analysis (Luo et al., 2015).

Whole-genome assembly was performed using SOAPdenovo2 (Luo et al., 2015), ALLPATHS-LG (Luo et al., 2015), and Platanus (Kajitani et al., 2014). Repetitive elements were identified and analyzed using RepeatMasker (Bergman and Quesneville, 2007) and RepeatModeler (Flynn et al., 2020).

The MAKER-P pipeline (Campbell et al., 2014) was used for gene annotation by integrating multiple annotation resources. The genome annotation revealed 23 155, 28 191, and 35 140 genes in the genomes of *Purshia*, *Dryas*, and *Fagus*, respectively. Gene functional prediction and assignment were performed using InterProScan (Jones et al., 2014) and homology searches against the Swiss-Prot (https://www.ebi.ac.uk/uniprot) and KEGG databases (https://www.kegg.jp/blastkoala/).

Evolutionary analysis

Analysis of RBHs and orthologous/paralogous groups

The genome sequences and annotations of the selected plants in this study were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/) and other plant genome websites (supplemental Table 1). We used two strategies to elucidate the orthology of all genes: RBH identification and orthologous/paralogous clustering. First, we identified RBHs by performing all-vs-all BLAST searches of the plant proteomes against the *M. truncatula* proteome and then performing reverse-BLAST searches of the *Medicago* proteome against the other plant proteomes. We analyzed the presence and absence of the orthologs of RBHs for each species. We compared the number of species carrying orthologs of target reference proteins between nodulators, non-nodulators, and outgroups using Fisher's exact test to infer the association between the presence of target genes and nodulation. We calculated p using the formula:

$$p = \sqrt{\left(\left(\frac{\binom{M+N}{M}\binom{m+n}{m}}{\binom{M+N+m+n}{M+m}}\right)^2 + \left(\frac{\binom{M+O}{M}\binom{m+o}{m}}{\binom{M+O+m+o}{M+m}}\right)^2\right) \Big/ 2};$$

Here, M, N, and O are the numbers of nodulators, non-nodulators, and outgroup species, respectively, that carry orthologs of target proteins; m, n, and o are the number of nodulators, non-nodulators, and outgroup species, respectively, that do not carry orthologs of target proteins. Second, we performed all-vs-all BLAST searches of all plant proteomes and used OrthoFinder to identify the orthologous/paralogous groups (Emms and Kelly, 2019). We calculated the gene copy number of each group in each species. We then compared the copy numbers of protein family members from nodulating (legume or non-legume) species, non-nodulating species, and outgroups using a *t*-test. If the difference in copy number revealed by OrthoFinder between two groups of interest (i.e., nodulating species vs. non-nodulating species) for a gene family was larger than 1 and the *t*-test value was smaller than 0.01, that family was considered to be an expanded gene family in nodulating species. To confirm the orthology of families revealed by these two strategies, we constructed the phylogenetic tree of each target gene. We aligned the proteins encoded by target genes from each species using MAFFT (Katoh et al., 2002), and then constructed maximum-likelihood phylogenetic trees with the best-fit model selected by IQ-TREE and 1000 bootstraps (Nguyen et al., 2015).

Identification of convergent loci

We used two combinatorial approaches to detect convergence signals within sequences in nodulating species: a "tree topology inference" method and an "alignment" method.

For the "tree topology inference" method, we used a pipeline based on maximum likelihood phylogenetic reconstruction according to the method of Parker et al. (Parker et al., 2013; Thomas and Hahn, 2015) with minor modifications. We aligned the CDSs of each orthologous protein and measured the fit of the alignment (site-wise log-likelihood support; SSLS) to the known species tree (H0) and an alternative topology in which nodulating species (H1) formed a monophyletic clade. In brief, the pipeline was as follows: (1) the one-to-one orthologs present in all nodulators or non-nodulators were aligned using MAFFT; (2) the log-likelihoods of the phylogenies (H0, H1) were calculated for every site in the alignment using RAXML (v8.2.12) (Stamatakis, 2014) with the parameters "-f g -m GTRGAMMA"; the resulting log-likelihoods of H0 and H1 for each site were subtracted to obtain ΔSSLS (ΔSSLSi = lnLi,H0 – lnLi,H1, where lnLi,H0 and lnLi,H1 denote the log-likelihood of the ith site under H0 and H1, respectively); the sequence convergence of each gene was quantified by taking the mean of ΔSSLS at each site; and (3) 10 000 random trees were generated to simulate the null distribution of ΔSSLS. ΔSSLS of a significant convergent gene (supporting H1) should be smaller than the left-tail 0.01 probability of the null distribution.

H0 tree: ((((((((((Alnus_glutinosa,(Betula_pendula, Betula_nana)),(Casuarina_equisetifolia, Casuarina_glauca)),((((Juglans_regia, Juglans_sigillata),(Juglans_cathayensis,(Juglans_hindsii,(Juglans_microcarpa, Juglans_nigra)))), Pterocarya_stenoptera),Morella_rubra)),(Fagus_sylvatica,(Quercus_suber,(Quercus_lobata, Quercus_robur)))),((((Ammopiptanthus_nanus, Lupinus_angustifolius),((((Arachis_duranensis, Arachis_ipaensis),Arachis_monticola), Nissolia_schottii),((Cajanus_cajan,((Glycine_max,Glycine_soja),(Lablab_purpureus,(Phaseolus_vulgaris,((Vigna_angularis, Vigna_radiata),Vigna_subterranea))))),((((Cicer_arietinum, Cicer_reticulatum),(Medicago_truncatula,(Trifolium_pratense, Trifolium_subterraneum))),Glycyrrhiza_uralensis),Lotus_japonicus)))),(Chamaecrista_fasciculata,(Faidherbia_albida, Mimosa_pudica))),Cercis_canadensis)),(((Begonia_fuchsioides, Datisca_glomerata),((((Citrullus_lanatus, Lagenaria_siceraria),(Cucumis_melo, Cucumis_sativus)),(((Cucurbita_argyrosperma, Cucurbita_moschata),Cucurbita_pepo),Cucurbita_maxima)),(Momordica_charantia, Siraitia_grosvenorii)))),(((((Cannabis_sativa, Humulus_lupulus),(Parasponia_andersonii, Trema_orientale)),((Ficus_carica,(Morus_notabilis, Artocarpus_camansi)),Boehmeria_nivea)),(Discaria_trinervis, Ziziphus_jujuba)),(((Dryas_drummondii, Dryas_octopetala),Purshia_tride)))ta),(((((Fragaria_))sca,Fragaria_x_ananassa),potentilla_micra)tha),(Rosa_chinensis, Rosa_multiflora)),Rubus_occidentalis),((Malus_domestica,(Pyrus_communis, Pyrus_x_bretschneideri)),((Prunus_avium, Prunus_yedoensis),(Prunus_mume, Prunus_persica))))))))),(((Arabidopsis_thaliana, Theobroma_cacao),Citrus_clementina),(Cephalotus_follicularis, Populus_trichocarpa))),Eucalyptus_grandis),Vitis_vinifera),Amborella_trichopoda);

H1 tree: ((((((((((Lupinus_angustifolius, Ammopiptanthus_nanus),(((Arachis_duranensis, Arachis_ipaensis),Arachis_monticola),((Cajanus_cajan,((Glycine_max,Glycine_soja),(Lablab_purpureus,(Phaseolus_vulgaris,((Vigna_angularis, Vigna_radiata),Vigna_subterranea))))),((((Cicer_arietinum, Cicer_reticulatum),(Medicago_truncatula,(Trifolium_pratense, Trifolium_subterraneum))),Glycyrrhiza_uralensis),Lotus_japonicus)))),(Chamaecrista_fasciculata,(Faidherbia_albida, Mimosa_pudica))),(((Casuarina_equisetifolia, Casuarina_glauca),Alnus_glutinosa),Morella_rubra)),(Datisca_glomerata,((Discaria_trinervis, Parasponia_andersonii),(Purshia_tridentata, Dryas_drummondii)))),(((((Betula_pendula, Betula_nana),(((Juglans_regia, Juglans_sigillata),(Juglans_cathayensis,(Juglans_hindsii,(Juglans_microcarpa, Juglans_nigra)))),Pterocarya_stenoptera)),(Fagus_sylvatica,(Quercus_suber,(Quercus_lobata, Quercus_robur)))),(Nissolia_schottii, Cercis_canadensis)),((Begonia_fuchsioides,((((Citrullus_lanatus, Lagenaria_siceraria),(Cucumis_melo, Cucumis_sativus)),(((Cucurbita_argyrosperma, Cucurbita_moschata),Cucurbita_pepo),Cucurbita_maxima)),(Momordica_charantia, Siraitia_grosvenorii))),(((((Cannabis_sativa, Humulus_lupulus),Trema_orientale),((Ficus_carica,(Morus_notabilis, Artocarpus_camansi)),Boehmeria_nivea)),Ziziphus_jujuba),(Dryas_octopetala,(((((Fragaria_vesca, Fragaria_x_ananassa),Potentilla_micrantha),(Rosa_chinensis, Rosa_multiflora)),Rubus_occidentalis),((Malus_domestica,(Pyrus_communis, Pyrus_x_bretschneideri)),((Prunus_avium, Prunus_yedoensis),(Prunus_mume, Prunus_persica)))))))),(((Arabidopsis_thaliana, Theobroma_cacao),Citrus_clementina),(Cephalotus_follicularis, Populus_trichocarpa))),Eucalyptus_grandis),Vitis_vinifera),Amborella_trichopoda);

Identification of CNEs

We identified CNEs by comparing the *M. truncatula* reference genome with 87 query genomes (supplemental Figures 2 and 13) (Hubisz et al., 2011; Haudry et al., 2013; Liang et al., 2018; Sackton et al., 2019; Armstrong et al., 2020). First, we annotated simple repeats using Tantan (Frith, 2011) to find orthologous sequences more accurately. Multiple sequence alignments of whole genomes were generated separately using two methods: Lastz-ChainNet-Roast (Liang et al., 2018) and Progressive Cactus (Armstrong et al., 2020). For the Lastz-ChainNet-Roast method, each query genome was aligned to the *Medicago* genome using LASTz (v1.04.00) (Harris, 2007) with parameters ''–ambiguous = iupac –chain –notransition H = 2000 Y = 3000 L = 3000 K = 2200 –format = axt –gfextend''; the alignments for each query were linked into longer chains using axtChain, chainPreNet, and chainNet with default parameters (Kent et al., 2003). For each query, when two alignments overlapped in the *Medicago* genome, the overlapping part of the shorter one was removed by single_cov2 (http://www.bx.psu.edu/∼cathy/toast-roast.tmp/README.toast-roast.html). Next, we linked all the pairwise alignments of each query genome and the *Medicago* reference genome according to the topology of the species tree using ROAST (v3; http://www.bx.psu.edu/∼cathy/toast-roast.tmp/README.toast-roast.html). For the Progressive Cactus method, the phylogenetic tree of one-to-one orthologs was used as a guide tree, and the high-quality assemblies (scaffold N50 $\geq$1 Mb and contig N50 $\geq$ 20 kb) were marked with asterisks (*). The reference-free alignment in HAL format was exported as a maf alignment using different species (*M. truncatula*, *Ammopiptanthus nanus*, *Discaria trinervis*, *Ziziphus jujuba*, *D. glomerata*, *Lagenaria siceraria*, *C. glauca*, *F. sylvatica*, *Populus trichocarpa*) as references. Maf alignments from the two methods using *M. truncatula* as a reference were combined. Only alignments with length $\geq$5 bp and rows $\geq$10 were retained. PhastCons (Hubisz et al., 2011) and CNSpipeline were used separately to identify conserved elements (CEs). For PhastCons, we used default parameters to identify CEs. For CNSpipeline, we calculated the conserved score for each alignment: score = number of aligned sequences/total number of species. Alignments with score $\geq$0.9 were selected as conserved CEs in each species. We then merged the CEs generated by the two CE identification methods (merged overlapping CEs and remaining specific CEs). CEs that only overlapped with noncoding regions were referred to as CNEs. CNE presence/absence was analyzed as described above for gene presence/absence. The overlapping CNEs between nodulating and non-nodulating species were calculated using a series of Python scripts. The significance of CNE motif enrichment was calculated by Z score = $\frac{x-\mu}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2}}$, where x_i is the number of CNEs that matched a specific $\geq$6-nt motif in the PLACE or plantPAN3.0 database (http://plantpan.itps.ncku.edu.tw/) ($x_i \geq$ 2), N is the total number of matched motifs, and μ is the average of number of CNE matched motifs. To investigate the enrichment of two repressive histone marks (H3K27me1 and H3K9me2), one activating mark (H3K9ac), and ATAC-seq data in CNEs, we used GAT (Heger et al., 2013) to compare the genome loci of CHIP marks and CNEs.

Hairy root transformation

The CE region (Liu et al., 2019), 5-kb promoter, and full-length CDS of wild-type *LjNIN* and its variants were cloned into the pUB-GFP vector to obtain pUB-GFP LjNINproCE-LjNIN-wt, pUB-GFP LjNINproCE-LjNIN-TA, and pUB-GFP LjNINproCE-LjNIN-TD, respectively. *A. rhizogenes* strain LBA1334 cells carrying pUB-GFP LjNINproCE-LjNIN-wt, pUB-GFP LjNINproCE-LjNIN-TA, pUB-GFP LjNINproCE-LjNIN-TD, and the empty vector were used to induce hairy root formation in *Ljnin-2* mutant plants (*L. japonicus*) using a previously described procedure (Yuan et al., 2012). Phenotypes of transgenic hairy roots were screened and photographed 21 days after inoculation with *M. loti* MAFF303099. Transgenic hairy roots expressing the empty vector (pUG-GFP) were used as a negative control. Mean nodule number calculations and Student's *t*-tests were performed using R.

RNA-seq analysis

We designed a systematic meta-analysis of transcriptomes from a variety of species in the NFNC. We obtained 151 RNA-seq libraries across 20 phylodiverse species within the NFNC, 7 of which were highlighted for comparison across tissues (roots/nodules) and treatments (with or without nitrate and/or compatible *Frankia*). These included 4 actinorhizal nodulating plant species (*A. glutinosa*, *D. glomerata*, *C. glauca*, *P. tridentata*) and 2 non-nodulating species (*B. pendula*, *B. fuchsioides*) (Figure 4A, supplemental Figure 7, and supplemental Tables 14–27). We added 9 transcriptomes of *M. truncatula* from NCBI for comparative analysis.

Low-quality raw RNA-seq reads were filtered using SOAPfilter (Luo et al., 2015). For species with sequenced genomes, we mapped the filtered reads to the reference genome using HISAT2 (Kim et al., 2019). Reference-guided transcript assembly for each library was performed using StringTie (Pertea et al., 2015). Assembled transcripts from each library were merged, and transcripts were re-annotated on the genome using gffcompare. For species without sequenced genomes, transcripts were assembled *de novo* using Trinity (Grabherr et al., 2011). We calculated the expression level (TPM) of each gene using Salmon (Patro et al., 2017). Pairwise correlations between transcriptomes were calculated using gene expression levels (estimated as TPM) of 3987 one-to-one orthologs across the 21 species (including the public dataset for Medicago) (supplemental Figure 7). Genes that were differentially expressed in different treatments (+N–B, +N+B, −N+B, −N−B) were identified with DESeq2 (adjusted p value <0.05 and log2FoldChange >2) (Love et al., 2014). Ortholog/paralog groups of DEGs that were upregulated in the same tissue/treatment in different species were identified with OrthoFinder. Symbiosis genes were mapped to each group, and variation among species was analyzed with a series of perl scripts. The DEGs and their encoded proteins were functionally annotated using InterProScan, Swiss-Prot, BLAST, and KEGG (https://www.kegg.jp/blastkoala/).

DATA AND CODE AVAILABILITY

The raw RNA-seq and DNA sequencing data, as well as the new genome assemblies and annotations, have been deposited at the CNGB Sequence Archive (CNSA) of the China National Gene Bank DataBase (CNGBdb) under accession number CNP 0004055. Multiple whole-genome alignment files (Cactus) have been uploaded to Zenodo (https://zenodo.org/record/5798193#.ZCGPt-zP30p).

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

We thank Douglas Soltis and Pamela Soltis for providing suggestions. S.C. is supported by the National Natural Science Foundation of China (32022006), the Program for Guangdong "ZhuJiang" Innovation Teams (2019ZT08N628), the Agricultural Science and Technology Innovation Program (ASTIP) (CAAS-XTCX2016001), and the special funds for science technology innovation and industrial development of Shenzhen Dapeng New District (PT202101-01). Y.Z. is supported by the National Natural Science Foundation of China (32070250), the Natural Science Foundation of Guangdong Province (2020A1515011030), and the open research project of "Cross-Cooperative Team" of the Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences. Y.Feng is supported by the Natural Science Foundation of Guangdong Province (2022A1515110240). For *C. glauca* RNA-seq data, financial support was provided by IRD and the Agence Nationale de la Recherche (Project SESAM, 2010 BLAN 1708 01).

AUTHOR CONTRIBUTIONS

S.C. designed and oversaw the study. S.C., J.J.D., and Y.Z. wrote the manuscript. W.X., Y.Fu, Y.Z., and X.L., analyzed the data. A.M.B. provided samples for *P. tridentata*. P.-M.D. provided samples of *Mimosa pudica*. V.H. provided samples of *C. glauca*. M.P. provided samples of *D. drummondii* and *D. octopetala*. K.P. provided samples of *B. fuchsioides* and *D. glomerata*. P.P., N.A., P.F., H.B., and P.N. provided DNA from *A. glutinosa*, *Betula nana*, and *F. sylvatica* and contributed their RNA-seq experiments.

ACKNOWLEDGMENTS

No conflict of interest declared.

Received: June 25, 2023
Revised: July 10, 2023
Accepted: August 3, 2023

REFERENCES

Alloisio, N., Queiroux, C., Fournier, P., Pujic, P., Normand, P., Vallenet, D., Médigue, C., Yamaura, M., Kakoi, K., and Kucho, K.-i. (2010). The Frankia alni symbiotic transcriptome. Mol. Plant Microbe Interact. **23**:593–607.

Ardley, J., and Sprent, J. (2021). Evolution and biogeography of actinorhizal plants and legumes: a comparison. J. Ecol. **109**:1098–1121.

Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I.T., Novak, A.M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., et al. (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature **587**:246–251.

Auguy, F., Abdel-Lateif, K., Doumas, P., Badin, P., Guerin, V., Bogusz, D., and Hocher, V. (2011). Activation of the isoflavonoid pathway in actinorhizal symbioses. Funct. Plant Biol. **38**:690–696.

Battenberg, K., Potter, D., Tabuloc, C.A., Chiu, J.C., and Berry, A.M. (2018). Comparative Transcriptomic analysis of two actinorhizal plants and the legume Medicago truncatula supports the homology of root nodule symbioses and is congruent with a two-step process of evolution in the nitrogen-fixing clade of angiosperms. Front. Plant Sci. **9**:1256.

Bergman, C.M., and Quesneville, H. (2007). Discovering and detecting transposable elements in genome sequences. Briefings Bioinf. **8**:382–392.

Billault-Penneteau, B., Sandré, A., Folgmann, J., Parniske, M., and Pawlowski, K. (2019). Dryas as a Model for Studying the Root Symbioses of the Rosaceae. Front. Plant Sci. **10**:661.

Broughton, W.J., and Dilworth, M.J. (1971). Control of leghaemoglobin synthesis in snake beans. Biochem. J. **125**:1075–1080. https://doi.org/10.1042/bj1251075.

Campbell, M.S., Law, M., Holt, C., Stein, J.C., Moghe, G.D., Hufnagel, D.E., Lei, J., Achawanantakun, R., Jiao, D., Lawrence, C.J., et al. (2014). MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. Plant Physiol. **164**:513–524.

Carroll, S.B. (2005). Evolution at two levels: on genes and form. PLoS Biol. **3**:e245.

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell **134**:25–36.

Coate, J.E., Farmer, A.D., Schiefelbein, J.W., and Doyle, J.J. (2020). Expression partitioning of duplicate genes at single cell resolution in Arabidopsis roots. Frontiers in Genetics **11**:596150.

Diédhiou, I., Tromas, A., Cissoko, M., et al. (2014). Identification of potential transcriptional regulators of actinorhizal symbioses in Casuarina glauca and Alnus glutinosa. BMC Plant Biol **14**:1–3.

DiFrisco, J., Wagner, G.P., and Love, A.C. (2022). Reframing research on evolutionary novelty and co-option: character identity mechanisms versus deep homology. Seminars in Cell & Developmental Biology (Elsevier).

Dong, W., Zhu, Y., Chang, H., Wang, C., Yang, J., Shi, J., Gao, J., Yang, W., Lan, L., Wang, Y., et al. (2021). An SHR–SCR module specifies legume cortical cell fate to enable nodulation. Nature **589**:586–590.

Donoghue, M.J., and Sanderson, M.J. (2015). Confluence, synnovation, and depauperons in plant diversification. New Phytol. **207**:260–274.

Doyle, J.J. (1994). Phylogeny of the legume family - an approach to understanding the origins of nodulation. Annu. Rev. Ecol. Systemat. **25**:325–349.

Doyle, J.J. (2011). Phylogenetic perspectives on the origins of nodulation. Mol. Plant Microbe Interact. **24**:1289–1295.

Doyle, J.J. (2016). Chasing unicorns: Nodulation origins and the paradox of novelty. Am. J. Bot. **103**:1865–1868. https://doi.org/10.3732/ajb.1600260.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. **20**:238.

Fåhraeus, G. (1957). The infection of clover root hairs by nodule bacteria studied by a simple glass slide technique. Microbiology **16**:374–381.

Fehlberg, V., Vieweg, M.F., Dohmann, E.M.N., Hohnjec, N., Pühler, A., Perlick, A.M., and Küster, H. (2005). The promoter of the leghaemoglobin gene VfLb29: functional analysis and identification of modules necessary for its activation in the infected cells of root nodules and in the arbuscule-containing cells of mycorrhizal roots. J. Exp. Bot. **56**:799–806.

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. USA **117**:9451–9457.

Fournier, J., Imanishi, L., Chabaud, M., Abdou-Pavy, I., Genre, A., Brichet, L., Lascano, H.R., Muñoz, N., Vayssières, A., Pirolles, E., et al. (2018). Cell remodeling and subtilase gene expression in the actinorhizal plant Discaria trinervis highlight host orchestration of intercellular Frankia colonization. New Phytol. **219**:1018–1030.

Franche, C., Diouf, D., Le, Q., Bogusz, D., N'diaye, A., Gherbi, H., Gobé, C., and Duhoux, E. (1997). Genetic transformation of the actinorhizal tree Allocasuarina verticillata by Agrobacterium tumefaciens. Plant J. **11**:897–904.

Frith, M.C. (2011). A new repeat-masking method enables specific detection of homologous sequences. Nucleic Acids Res. **39**:e23.

Geurts, R., Xiao, T.T., and Reinhold-Hurek, B. (2016). What does it take to evolve a nitrogen-fixing endosymbiosis? Trends Plant Sci. **21**:199–208.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. **29**:644–652.

Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M.B., Billault-Penneteau, B., Lauressergues, D., Keller, J., Imanishi, L., et al. (2018). Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. Science **361**, eaat1743.

Harris, R.S. (2007). Improved Pairwise Alignment of Genomic DNA (The Pennsylvania State University).

Haudry, A., Platts, A.E., Vello, E., Hoen, D.R., Leclercq, M., Williamson, R.J., Forczek, E., Joly-Lopez, Z., Steffen, J.G., Hazzouri, K.M., et al. (2013). An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. Nat. Genet. **45**:891–898. https://doi.org/10.1038/ng.2684.

Heger, A., Webber, C., Goodson, M., Ponting, C.P., and Lunter, G. (2013). GAT: a simulation framework for testing the association of genomic intervals. Bioinformatics **29**:2046–2048. https://doi.org/10.1093/bioinformatics/btt343.

Hoagland, D.R., and Arnon, D.I. (1950). The Water-Culture Method for Growing Plants without Soil, 347 (Circular. California agricultural experiment station).

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

Hubisz, M.J., Pollard, K.S., and Siepel, A. (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. Briefings Bioinf. **12**:41–51.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics **30**:1236–1240.

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., et al. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. **24**:1384–1395.

Kates, H.R., O'Meara, B.C., LaFrance, R., Stull, G.W., James, E.K., Conde, D., Liu, S., Tian, Q., Yi, T., Kirst, M., et al. (2022). Two shifts in evolutionary lability underlie independent gains and losses of root-nodule symbiosis in a single clade of plants. Preprint at bioRxiv. https://doi.org/10.1101/2022.07.31.502231.

Katoh, K., Misawa, K., Kuma, K.i., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. **30**:3059–3066.

Kent, W.J., Baertsch, R., Hinrichs, A., Miller, W., and Haussler, D. (2003). Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. Proc. Natl. Acad. Sci. USA **100**:11484–11489.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. **37**:907–915.

Koenen, E.J.M., Ojeda, D.I., Bakker, F.T., Wieringa, J.J., Kidner, C., Hardy, O.J., Pennington, R.T., Herendeen, P.S., Bruneau, A., and Hughes, C.E. (2021). The origin of the legumes is a complex paleopolyploid phylogenomic tangle closely associated with the Cretaceous–Paleogene (K–Pg) mass extinction event. Syst. Biol. **70**:508–526.

Krusell, L., Krause, K., Ott, T., Desbrosses, G., Krämer, U., Sato, S., Nakamura, Y., Tabata, S., James, E.K., Sandal, N., et al. (2005). The sulfate transporter SST1 is crucial for symbiotic nitrogen fixation in Lotus japonicus root nodules. Plant Cell **17**:1625–1636.

Li, H.-L., Wang, W., Mortimer, P.E., Li, R.-Q., Li, D.-Z., Hyde, K.D., Xu, J.-C., Soltis, D.E., and Chen, Z.-D. (2015). Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. Sci. Rep. **5**:14023–14028.

Li, Q.-G., Zhang, L., Li, C., Dunwell, J.M., and Zhang, Y.-M. (2013). Comparative genomics suggests that an ancestral polyploidy event leads to enhanced root nodule symbiosis in the Papilionoideae. Mol. Biol. Evol. **30**:2602–2611.

Liang, P., Saqib, H.S.A., Zhang, X., Zhang, L., and Tang, H. (2018). Single-Base Resolution Map of Evolutionary Constraints and Annotation of Conserved Elements across Major Grass Genomes. Genome Biol. Evol. **10**:473–488. https://doi.org/10.1093/gbe/evy006.

Liu, J., and Bisseling, T. (2020). Evolution of NIN and NIN-like Genes in Relation to Nodule Symbiosis. Genes **11**:777.

Liu, J., Rutten, L., Limpens, E., Van Der Molen, T., Van Velzen, R., Chen, R., Chen, Y., Geurts, R., Kohlen, W., Kulikova, O., et al. (2019). A remote cis-regulatory region is required for NIN expression in the pericycle to initiate nodule primordium formation in Medicago truncatula. Plant Cell **31**:68–83.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**:550.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2015). Erratum: SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience **4**:30–31.

Markmann, K., and Parniske, M. (2009). Evolution of root endosymbiosis with bacteria: how novel are nodules? Trends Plant Sci. **14**:77–86.

Mathesius, U. (2022). Are legumes different? Origins and consequences of evolving nitrogen fixing symbioses. J. Plant Physiol. **276**, 153765.

Merényi, Z., Prasanna, A.N., Wang, Z., Kovács, K., Hegedüs, B., Bálint, B., Papp, B., Townsend, J.P., and Nagy, L.G. (2020). Unmatched level of molecular convergence among deeply divergent complex multicellular fungi. Mol. Biol. Evol. **37**:2228–2240.

Mergaert, P., Kereszt, A., and Kondorosi, E. (2020). Gene expression in nitrogen-fixing symbiotic nodule cells in Medicago truncatula and other nodulating plants. Plant Cell **32**:42–68.

Miri, M., Janakirama, P., Held, M., Ross, L., and Szczyglowski, K. (2016). Into the root: how cytokinin controls rhizobial infection. Trends Plant Sci. **21**:178–186.

Mishra, B., Gupta, D.K., Pfenninger, M., Hickler, T., Langer, E., Nam, B., Paule, J., Sharma, R., Ulaszewski, B., Warmbier, J., et al. (2018). A reference genome of the European beech (Fagus sylvatica L.). GigaScience **7**:giy063.

Mishra, B., Ulaszewski, B., Meger, J., Aury, J.-M., Bodénès, C., Lesur-Kupin, I., Pfenninger, M., Da Silva, C., Gupta, D.K., Guichoux, E., et al. (2021). A Chromosome-level genome assembly of the European Beech (Fagus sylvatica) reveals anomalies for organelle DNA integration, repeat Content and distribution of SNPs. Front. Genet. **12**:691058.

Mortier, V., Wasson, A., Jaworek, P., De Keyser, A., Decroos, M., Holsters, M., Tarkowski, P., Mathesius, U., and Goormachtig, S. (2014). Role of LONELY GUY genes in indeterminate nodulation on Medicago truncatula. New Phytol. **202**:582–593.

Nguyen, L.-T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. **32**:268–274.

Please cite this article in press as: Zhang et al., Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis, Plant Communications (2023), https://doi.org/10.1016/j.xplc.2023.100671

Parker, J., Tsagkogeorga, G., Cotton, J.A., Liu, Y., Provero, P., Stupka, E., and Rossiter, S.J. (2013). Genome-wide signatures of convergent evolution in echolocating mammals. Nature **502**:228–231.

Parniske, M. (2018). Uptake of bacteria into living plant cells, the unifying and distinct feature of the nitrogen-fixing root nodule symbiosis. Curr. Opin. Plant Biol. **44**:164–174.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods **14**:417–419.

Pereira, W.J., Knaack, S., Chakraborty, S., et al. (2022). Functional and comparative genomics reveals conserved noncoding sequences in the nitrogen-fixing clade. New Phytol **234**:634–649.

Persson, T., Benson, D.R., Normand, P., Vanden Heuvel, B., Pujic, P., Chertkov, O., Teshima, H., Bruce, D.C., Detter, C., Tapia, R., et al. (2011). Genome sequence of "Candidatus Frankia datiscae" Dg1, the uncultured microsymbiont from nitrogen-fixing root nodules of the dicot Datisca glomerata. J. Bacteriol. **193**:7017–7018.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. **33**:290–295.

Roy, S., Liu, W., Nandety, R.S., et al. (2020). Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. The Plant Cell **32**:15–41.

Sackton, T.B., Grayson, P., Cloutier, A., Hu, Z., Liu, J.S., Wheeler, N.E., Gardner, P.P., Clarke, J.A., Baker, A.J., Clamp, M., et al. (2019). Convergent regulatory evolution and loss of flight in paleognathous birds. Science **364**:74–78.

Schmitz, R.J., Grotewold, E., and Stam, M. (2022). Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. Plant Cell **34**:718–741.

Schwencke, J. (1991). Rapid, exponential growth and increased biomass yield of some Frankia strains in buffered and stirred mineral medium (BAP) with phosphatidyl choline. Nitrogen Fixation Proceedings of the Fifth International Symposium on Nitrogen Fixation with Non-Legumes (Springer).

Shen, D., Xiao, T.T., van Velzen, R., Kulikova, O., Gong, X., Geurts, R., Pawlowski, K., and Bisseling, T. (2020). A homeotic mutation changes legume nodule ontogeny into actinorhizal-type ontogeny. Plant Cell **32**:1868–1885.

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

Sinnott-Armstrong, M.A., Deanna, R., Pretz, C., Liu, S., Harris, J.C., Dunbar-Wallis, A., Smith, S.D., and Wheeler, L.C. (2022). How to approach the study of syndromes in macroevolution and ecology. Ecol. Evol. **12**, e8583.

Soltis, D.E., Soltis, P.S., Morgan, D.R., Swensen, S.M., Mullin, B.C., Dowd, J.M., and Martin, P.G. (1995). Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc. Natl. Acad. Sci. USA **92**:2647–2651.

Soyano, T., and Hayashi, M. (2014). Transcriptional networks leading to symbiotic nodule organogenesis. Curr. Opin. Plant Biol. **20**:146–154.

Soyano, T., Shimoda, Y., Kawaguchi, M., and Hayashi, M. (2019). A shared gene drives lateral root development and root nodule symbiosis pathways in Lotus. Science **366**:1021–1023.

Soyano, T., Liu, M., Kawaguchi, M., and Hayashi, M. (2021). Leguminous nodule symbiosis involves recruitment of factors contributing to lateral root development. Curr. Opin. Plant Biol. **59**, 102000.

Sprent, J.I., Ardley, J., and James, E.K. (2017). Biogeography of nodulated legumes and their nitrogen-fixing symbionts. New Phytol. **215**:40–56.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

Stern, D.L. (2000). Perspective: evolutionary developmental biology and the problem of variation. Evolution **54**:1079–1091.

Stougaard, J., Jørgensen, J.E., Christensen, T., Kühle, A., and Marcker, K.A. (1990). Interdependence and nodule specificity of cis-acting regulatory elements in the soybean leghemoglobin lbc 3 and N23 gene promoters. Mol. Gen. Genet. **220**:353–360.

Suzuki, W., Konishi, M., and Yanagisawa, S. (2013). The evolutionary events necessary for the emergence of symbiotic nitrogen fixation in legumes may involve a loss of nitrate responsiveness of the NIN transcription factor. Plant Signal. Behav. **8**, e25975.

Thomas, G.W.C., and Hahn, M.W. (2015). Determining the null model for detecting adaptive convergence from genomic data: a case study using echolocating mammals. Mol. Biol. Evol. **32**:1232–1236.

Tokumoto, Y., Hashimoto, K., Soyano, T., Aoki, S., Iwasaki, W., Fukuhara, M., Nakagawa, T., Saeki, K., Yokoyama, J., Fujita, H., et al. (2020). Assessment of Polygala paniculata (Polygalaceae) characteristics for evolutionary studies of legume–rhizobia symbiosis. J. Plant Res. **133**:109–122.

Uno, Y., and Hirasawa, T. (2023). Origin of the propatagium in non-avian dinosaurs. Zoological Lett. **9**:4.

Valkov, V.T., Rogato, A., Alves, L.M., Sol, S., Noguero, M., Léran, S., Lacombe, B., and Chiurazzi, M. (2017). The nitrate transporter family protein LjNPF8. 6 controls the N-fixing nodule activity. Plant Physiol. **175**:1269–1282.

van Velzen, R., Doyle, J.J., and Geurts, R. (2019). A resurrected scenario: single gain and massive loss of nitrogen-fixing nodulation. Trends Plant Sci. **24**:49–57.

1. **van Velzen, R., Holmer, R., Bu, F., et al.** (2018). Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. Proc. Natl. Acad. Sci. USA **115**:E4700–E4709.

Wang, D., Dong, W., Murray, J., and Wang, E. (2022). Innovation and appropriation in mycorrhizal and rhizobial symbioses. Plant Cell **34**:1573–1599.

Werner, G.D.A., Cornwell, W.K., Sprent, J.I., Kattge, J., and Kiers, E.T. (2014). A single evolutionary innovation drives the deep evolution of symbiotic N2-fixation in angiosperms. Nat. Commun. **5**:4087–4089.

Wu, Z., Chen, H., Pan, Y., et al. (2022). Genome of Hippophae rhamnoides provides insights into a conserved molecular mechanism in actinorhizal and rhizobial symbioses. New Phytol **235**:276–291.

Yuan, S., Zhu, H., Gou, H., Fu, W., Liu, L., Chen, T., Ke, D., Kang, H., Xie, Q., Hong, Z., et al. (2012). A ubiquitin ligase of symbiosis receptor kinase involved in nodule organogenesis. Plant Physiol. **160**:106–117.

Zhang, Z., Lopez, M.F., and Torrey, J.G. (1984). A comparison of cultural characteristics and infectivity of Frankia isolates from root nodules of Casuarina species. Frankia Symbioses, 79–90.

Zhao, Y., Zhang, R., Jiang, K.-W., Qi, J., Hu, Y., Guo, J., Zhu, R., Zhang, T., Egan, A.N., Yi, T.-S., et al. (2021). Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. Mol. Plant **14**:748–773.