Research Report Open Access

# Comparative Genomic Analysis of Nitrogen Fixation Genes in Major Legume Crops

Yuping Huang, Yunxia Chen, Hangming Lin
Tropical Legume Research Center, Hainan Institute of Tropical Agricultural Resources, Sanya, 572025, Hainan, China
Corresponding email: hangming.lin@hitar.org

Legume Genomics and Genetics, 2026 Vol.17, No.1 doi: 10.5376/lgg.2026.17.0002
Received: 10 Feb., 2026
Accepted: 12 Feb., 2026
Published: 22 Mar., 2026

**Preferred citation for this article**:
Huang Y.P., Chen Y.X., and Lin H.M., 2026, Comparative genomic analysis of nitrogen fixation genes in major legume crops, Legume Genomics and Genetics, 17(1): 14-31 (doi: 10.5376/lgg.2026.17.0002)

**Abstract** Leguminous plants are capable of achieving biological nitrogen fixation through symbiotic relationships with rhizobia, thereby holding significant ecological and economic value in agricultural production. With the completion of genome sequencing for various legume crops, the systematic elucidation of the structural characteristics, evolutionary patterns, and functional divergence of nitrogen fixation-related genes-utilizing comparative genomics approaches-has emerged as a key direction in current plant molecular biology and crop genetic improvement research. Focusing on major legume crops such as soybean, pea, peanut, and alfalfa, this review systematically summarizes the molecular mechanisms underlying nitrogen fixation in legumes and the key genes regulating this process. Particular emphasis is placed on analyzing recent progress in identifying genes associated with nitrogen fixation signal recognition, nodule development regulation, and nitrogen metabolism. Building upon this foundation and integrating findings from comparative genomics studies, a comprehensive analysis is conducted regarding the evolutionary characteristics of nitrogen fixation-related genes in major legume crops, examining aspects such as gene family structure, sequence conservation, gene duplication, and phylogenetic relationships. Furthermore, using the NIN gene family-a key regulator of nitrogen fixation in soybean-as a case study, the review explores its evolutionary patterns across different legume species and its potential value for improving nitrogen fixation efficiency. Finally, this paper discusses the future prospects for applying comparative genomics in the mining of nitrogen fixation genes, the development of molecular markers, and molecular design breeding, while also offering perspectives on future directions for multi-omics integration research. The findings indicate that comparative genomics provides a crucial theoretical foundation for unraveling the mechanisms of nitrogen fixation in legumes and facilitating their genetic improvement, thereby holding significant importance for advancing the development of green agriculture.

**Keywords** Legume crops; Biological nitrogen fixation; Comparative genomics; Nodule formation; Nitrogen fixation genes

## 1 Introduction

Legumes occupy a pivotal position in global agriculture because they simultaneously provide high-quality plant protein and enrich soils through biological nitrogen fixation (BNF). Nitrogen is frequently the most limiting nutrient for crop production, and its industrial supply via the Haber–Bosch process is energy intensive and environmentally costly, contributing to greenhouse gas emissions, nitrate leaching, and eutrophication (Mahmud et al., 2020; Ladha et al., 2022). In contrast, symbiotic BNF in legumes relies on specialized root nodules that host nitrogen-fixing rhizobia, converting atmospheric dinitrogen into ammonia that can be assimilated by the plant (Lindström and Mousavi, 2019; Mahmud et al., 2020). Global assessments indicate that legume-based BNF supplies large quantities of nitrogen to agroecosystems, with typical fixation rates of 20–200 kg N $ha^{-1}$ $year^{-1}$ and average proportions of plant nitrogen derived from the atmosphere (Ndfa) around 60%–80% in many systems. These inputs not only satisfy a substantial part of the legumes' own nitrogen demand but also improve soil fertility and nitrogen availability for subsequent or companion crops, thereby reducing dependence on synthetic fertilizers and supporting ecological intensification of cereal-based systems (Jithesh et al., 2024). As climate change, soil degradation, and the need to lower external inputs intensify, enhancing the efficiency and stability of nitrogen fixation in major legume crops has become central to sustainable agricultural development (Ladha et al., 2022; De Lima et al., 2024).

The importance of nitrogen-fixing legumes has been recognized empirically for millennia, long before the mechanisms of nodulation and BNF were understood (Lindström and Mousavi, 2019). Systematic research into legume nitrogen fixation began with classical agronomy and microbiology, which documented the yield benefits and soil fertility gains associated with legumes in rotations, green manures, and intercrops, as well as the discovery of rhizobia and root nodules as the sites of symbiotic nitrogen reduction (Lindström and Mousavi, 2019). Over the twentieth century, physiological and ecological studies characterized environmental controls on fixation, including soil mineral nitrogen, water availability, temperature, and cropping system design (Ladha et al., 2022). The advent of molecular biology, genomics, and advanced isotopic methods in the late twentieth and early twenty-first centuries markedly accelerated progress. Quantitative reviews and meta-analyses have now synthesized hundreds of field and pot studies to evaluate Ndfa across legume species, climates, and management regimes, highlighting both the large potential contribution of BNF and its strong dependence on species traits and environmental conditions (Palmero et al., 2022; Yang et al., 2024). In parallel, the genetics of symbiotic nitrogen fixation (SNF) have been intensively dissected, and forward-and reverse-genetic approaches have uncovered nearly 200 plant genes required for nodulation and SNF in model and crop legumes, illuminating processes ranging from early signal perception to nodule senescence (Pankievicz et al., 2019; Roy et al., 2019; Goyal et al., 2021). These advances underpin breeding efforts to improve nitrogen fixation in crops such as soybean, chickpea, pea, and faba bean, and have spurred attempts to transfer nitrogen-fixing capabilities to cereals through synthetic biology and microbiome engineering (Mahmud et al., 2020; Qiao et al., 2024).

Despite this progress, the genetic architecture of nitrogen fixation in major legume crops remains incompletely understood, particularly in the context of complex field environments and diverse rhizobial communities. Symbiotic efficiency is determined by coordinated interactions between plant and bacterial genomes, and by the composition of the broader rhizosphere microbiome, which includes “helper” bacteria that can enhance BNF through metabolic support such as vitamin production (Qiao et al., 2024; Yang et al., 2024). On the microbial side, comparative studies have revealed extensive genomic diversity among diazotrophic bacteria, including a wide range of genera capable of nitrogen fixation and large variation in the organization and content of nodulation (nod) and nitrogen fixation (nif, fix) genes (De Lima et al., 2024; Zhong et al., 2024). On the plant side, legume genomes encode elaborate signaling pathways, transcriptional networks, and metabolic systems dedicated to establishing and maintaining effective symbiosis (Pankievicz et al., 2019; Roy et al., 2019). However, many of these components have been characterized in a few model species, and relatively less is known about how nitrogen fixation-related genes are organized, diversified, and functionally constrained across major grain and oilseed legumes cultivated worldwide. Given that natural variation in BNF is substantial within and among legume species, understanding the genomic basis of this variation is essential for targeted improvement through molecular breeding and genome editing (Neda, 2021).

In this context, comparative genomics provides a powerful framework for advancing nitrogen fixation research in legume crops. High-quality reference genomes and pan-genomes have now been generated for multiple legumes and their associated rhizobia, enabling systematic comparisons of gene family structures, sequence conservation, and presence–absence variation for key symbiotic functions (De Lima et al., 2024; Zhong et al., 2024). Comparative genomic analyses of Bradyrhizobium, for example, have shown that its pan-genome contains a small set of core housekeeping genes and a vast accessory genome enriched in nodulation, nitrogen fixation, and secretion system genes, which segregate into distinct genetic profiles linked to symbiotic capacity and host range (Zhong et al., 2024). Similar cross-species comparisons in legumes can identify conserved “core” *SNF* genes as well as lineage-specific expansions and neofunctionalization events in receptors, transcription factors, transporters, and metabolic enzymes that support nodulation and nitrogen assimilation (Pankievicz et al., 2019; De Lima et al., 2024). By integrating evolutionary analyses of gene duplication and divergence, phylogenetic reconstruction of nitrogen fixation-related gene families, and patterns of sequence conservation with emerging functional data, comparative genomics can reveal how nitrogen fixation pathways have diversified among major legume crops, pinpoint candidate genes and regulatory elements underlying superior fixation phenotypes, and guide marker development and genomic selection for enhanced BNF (Roy et al., 2019; De Lima et al., 2024). The present study, “Comparative Genomic Analysis of Nitrogen Fixation Genes in Major Legume Crops,” builds on this foundation to systematically characterize the genomic landscape of nitrogen fixation across key legumes, with the ultimate aim of informing molecular design breeding and contributing to sustainable, low-input agricultural systems.

## 2 Molecular Mechanisms of Nitrogen Fixation in Legumes

### 2.1 Signal recognition processes in nodule formation

The initiation of legume–rhizobium symbiosis depends on a finely tuned molecular dialogue that allows plants to distinguish compatible nitrogen-fixing partners from the multitude of soil microbes. Root-exuded flavonoids stimulate rhizobia to express nodulation (nod) genes and synthesize lipochitooligosaccharide Nod factors, which act as key symbiotic signals guiding early infection events. In host roots, Nod factors are perceived mainly by LysM-domain receptor-like kinases at the epidermis, which trigger intracellular signaling cascades shared with other endosymbioses, including the common symbiosis pathway (SYM) initially characterized in model legumes (Roy et al., 2019). This pathway involves plasma membrane receptors, nuclear-localized ion channels, and a calcium/calmodulin-dependent protein kinase that decodes characteristic nuclear calcium oscillations to activate nodulation-specific transcriptional programs. Comparative studies show that although the core signaling components are broadly conserved among legumes, gene family expansions, promoter variation, and allelic diversity in receptor and downstream signaling genes contribute to differences in host range, infection mode, and responsiveness to environmental conditions in major crops such as soybean, common bean, and pea (Sharma et al., 2020; Tsyganov and Tsyganova, 2020).

Beyond Nod factors, additional bacterial cues and secretion systems further refine host recognition and specificity. Rhizobia possess multiple protein secretion systems, including type III, IV, and VI systems, that deliver effector proteins influencing infection success, host range, and nodule number (Nelson and Sadowsky, 2015). These secretion systems, traditionally associated with pathogenicity, are increasingly recognized as central modulators of symbiotic compatibility and can either promote or block nodulation depending on host genotype. Plants in turn integrate symbiotic signaling with innate immunity, using pattern-recognition receptors and downstream defense pathways to prevent colonization by non-beneficial microbes while permitting entry of compatible rhizobia (Shumilina et al., 2023). Recent work has revealed that rhizobial tRNA-derived small RNAs (tRFs) can act as mobile signals taken up by host cells, where they hijack the RNA interference machinery to modulate expression of plant genes involved in nodule initiation and development, thereby promoting nodulation (Ren et al., 2019). These multilayered recognition and signaling processes demonstrate that early stages of nodule formation arise from a co-evolved network of small molecules, receptors, secretion systems, and regulatory RNAs, the genes for which show both strong conservation and lineage-specific diversification across legume genomes.

### 2.2 Nodule development and symbiotic regulatory networks

Following successful signal perception, legumes coordinate rhizobial infection in epidermal/root hair cells with organogenesis in the cortex to produce nodules that house nitrogen-fixing bacteroids (Lepetit and Brouquisse, 2023). Infection threads guide bacteria through root hairs into inner tissues, while cortical cell divisions generate nodule primordia that ultimately give rise to determinate or indeterminate nodules, differing in meristem persistence and zonation (Lepetit and Brouquisse, 2023). Genetic and genomic studies over the last two decades have identified nearly 200 plant genes required for symbiotic nitrogen fixation, including receptors, ion channels, transcription factors, transporters, and enzymes that collectively orchestrate infection, nodule morphogenesis, and bacteroid accommodation (Roy et al., 2019; Tsyganov and Tsyganova, 2020). Central to this regulatory network is the transcription factor NODULE INCEPTION (NIN), which integrates upstream symbiotic and nitrogen-status signals to initiate nodule organogenesis and regulate infection thread progression (Qiao et al., 2023). Recent work demonstrated that proteolytic processing of NIN by a signal peptidase complex yields a C-terminal fragment that specifically activates genes required for symbiosome development and the transition to a nitrogen-fixing state, revealing a conserved mechanism linking early development to functional maturation.

Nodule development is further shaped by systemic and local regulatory networks that balance the high carbon cost of symbiosis with whole-plant nitrogen demand. Autoregulation of nodulation (AON) is a root–shoot–root signaling loop in which early nodulation induces CLE peptide production in roots; these peptides travel to the shoot and are perceived by CLAVATA1-like leucine-rich repeat receptor kinases (e.g., HAR1/SUNN/NARK), which generate a systemic inhibitory signal that limits further nodule formation (Ferguson et al., 2018; Chaulagain and Frugoli, 2021). More recent studies have revealed a complementary positive systemic pathway involving CEP peptides and the CRA2 receptor, as well as a miR2111/TML module, together forming a balanced regulatory system that allows legumes to adjust nodule numbers according to nitrogen availability and energy status (Li et al., 2022; Yanlin et al., 2024). Nutrient-dependent regulation adds another layer of control: nitrate and other mineral nutrients modulate the expression and activity of NIN, NIN-like proteins (NLPs), and multiple hormone pathways, thereby integrating soil nitrogen signals with symbiotic development and nodule senescence (Qiao et al., 2023; Yanlin et al., 2024). Comparative and functional genomic analyses across legumes indicate that key components of AON, nutrient signaling, and developmental control reside in gene families with species-specific expansions and allelic diversity, providing a genomic basis for natural variation in nodule number, size, and persistence in major legume crops (Tsyganov and Tsyganova, 2020; Li et al., 2022; Yanlin et al., 2024).

### 2.3 Molecular Mechanisms Related to Nitrogen Fixation and Metabolism

Once nodules are established and rhizobia differentiate into bacteroids within plant-derived symbiosomes, effective nitrogen fixation depends on a highly integrated set of metabolic and regulatory mechanisms that couple bacterial nitrogenase activity to plant carbon supply and nitrogen assimilation. In the nodule infection and fixation zones, bacteroids express nif and fix gene clusters encoding the nitrogenase complex and associated functions for electron transfer and respiration, enabling the reduction of atmospheric $N_2$ to ammonia under microaerobic conditions (Schulte et al., 2021; Lepetit and Brouquisse, 2023). Plant leghemoglobins, encoded by multigene families, buffer free oxygen to maintain a low-oxygen environment compatible with nitrogenase while supporting the high respiratory demand of bacteroids; their absence or disruption leads to impaired fixation and early nodule senescence (Chaulagain and Frugoli, 2021). Metabolic modelling and ^13C-based flux analysis indicate that bacteroids rely primarily on plant-supplied dicarboxylates such as malate as energy and carbon sources, and that oxygen limitation constrains the decarboxylating arm of the tricarboxylic acid cycle, favoring ammonia release over assimilation (Schulte et al., 2021). This metabolic configuration, shaped by both plant and bacterial genomes, promotes efficient export of reduced nitrogen to the host, a defining feature of rhizobium–legume symbioses.

On the plant side, ammonia derived from bacteroids is rapidly assimilated through the glutamine synthetase/glutamate synthase (GS–GOGAT) pathway and further incorporated into amino acids or ureides, which are then transported to shoots (Schwember et al., 2019). Nodule metabolism is tightly coordinated with whole-plant nitrogen and carbon status, involving regulation of N and C metabolic enzymes, oxygen flux, redox balance, and responses to phosphorus and other nutrients (Lepetit and Brouquisse, 2023; Yanlin et al., 2024). Systemic changes in sucrose allocation, oxidative pentose phosphate pathway activity, and redox status appear to serve as integrators of plant N demand and nodule functioning, influencing both nitrogenase activity and nodule senescence (Schwember et al., 2019; Lepetit and Brouquisse, 2023). High-throughput transcriptomic and other “omics” approaches have identified extensive transcriptional reprogramming of nodule tissues in response to nitrate addition and abiotic stresses, including differential expression of nodulins, nodule-specific cysteine-rich peptides, transporters, and regulatory proteins (Schwember et al., 2019). MicroRNAs and other small RNAs are emerging as important regulators linking nitrogen fixation, stress responses, and developmental transitions, complementing transcription factor networks such as those centered on NAC proteins that govern N-induced nodule senescence (Schwember et al., 2019; Qiao et al., 2023). Comparative genomic analyses across legumes, combined with functional genomics, are beginning to reveal how gene family diversification, regulatory sequence evolution, and network rewiring in key metabolic and regulatory genes underlie species-specific strategies of nitrogen transport (amides vs. ureides), oxygen management, and stress resilience in nodules (Sharma et al., 2020). These insights provide a mechanistic and genomic foundation for designing legume cultivars and symbiotic combinations with enhanced nitrogen fixation efficiency and stability under variable field conditions.

## 3 Identification of Nitrogen Fixation-Related Genes in Major Legume Crops

### 3.1 Genes involved in nitrogen fixation signal recognition

Comparative genetic analyses across model and crop legumes have identified a core set of plant genes required for recognition of rhizobial signals and initiation of nodulation. More than 190 plant genes essential for symbiotic nitrogen fixation (SNF) have been catalogued in Medicago truncatula, Lotus japonicus, soybean (*Glycine max*) and common bean (Phaseolus vulgaris), many of which encode receptors and signaling proteins that mediate perception of Nod factors and downstream transduction of symbiotic signals (Chaulagain and Frugoli, 2021). LysM-domain receptor-like kinases that recognize rhizobial lipochitooligosaccharides, co-receptors such as SYMRK/DMI2, nuclear cation channels, and the calcium/calmodulin-dependent protein kinase CCaMK form the backbone of the common symbiosis pathway, and orthologous genes have been identified in all major legume crops examined (Mahmud et al., 2020). Downstream transcription factors, including NODULE INCEPTION (NIN), NIN-LIKE PROTEINS (NLPs), GRAS proteins (NSP1/NSP2), and NF-Y subunits, decode these signals to activate infection and organogenesis programs (Soyano et al., 2019). Integration of large RNA-seq datasets in soybean has shown that many of these genes belong to expanded gene families shaped by whole-genome duplication, and that nodulation-related paralogs often exhibit strong nodule- or root-specific expression, indicating functional specialization for symbiotic signaling (Cardoso-Silva et al., 2025).

Beyond the canonical SYM pathway, additional signal recognition genes and regulatory modules have been uncovered in crop legumes using forward genetics, expression profiling, and functional genomics. In soybean, members of the CCR4–NOT complex, particularly GmNOT4-1, were identified as regulators of Nod factor signaling: both overexpression and knockdown of GmNOT4-1 reduced nodule number and suppressed transcription of Nod factor pathway genes, highlighting dosage-sensitive post-transcriptional control of early symbiotic signaling (Zheng et al., 2023). Studies in common bean identified SRS/STY transcription factors (PvSRS) whose expression is induced throughout symbiosis; PvSRS10 is transcriptionally activated by NF-Y and by a MADS-box factor, situating it within the NIN–NF-Y cascade that translates perception of rhizobial signals into developmental responses (Ayra et al., 2025). Comparative reviews emphasize that many of these signal recognition components are conserved across legumes but also display lineage-specific expansions and allelic variation, linked to differences in host range, infection strategy, and environmental responsiveness in crop species (Chaulagain and Frugoli, 2021). This expanding catalog of signal perception and transduction genes, derived from model and crop legumes, provides a genomic framework for engineering or enhancing signal recognition capacity in major legume crops and, potentially, in non-legume hosts.

### 3.2 Genes regulating nodule development

Genetic dissection of nodulation has revealed a large suite of genes that regulate the transition from early signaling to nodule organogenesis and morphogenesis. Central among these is NIN, a master transcription factor whose loss abolishes nodule formation; NIN orthologs in multiple legumes have been shown to control cortical cell divisions, infection thread progression, and later steps leading to the nitrogen-fixing state (Feng et al., 2021). Recent work demonstrated that NIN is proteolytically processed by a signal peptidase complex to release a C-terminal fragment with DNA-binding activity that specifically activates genes required for symbiosome development and nitrogen fixation, and similar processing was detected in several legumes, indicating a conserved mechanism controlling the developmental transition to functional nodules (Feng et al., 2021). Downstream of NIN, NF-Y transcription factor complexes, GRAS proteins, ERF/AP2 factors, and additional TF families (including SRS/STY and MADS-domain proteins in common bean) regulate expression of genes for cell cycle control, infection thread guidance, and nodule meristem activity (Ayra et al., 2025). High-resolution single-nucleus transcriptomic atlases in soybean have further identified novel regulators of nodulation, such as GmbHLH93, GmSCL1, and the cytokinin receptor GmCRE1; knockout of GmCRE1 led to severe defects in the nitrogen fixation zone, loss of leghemoglobin expression, and almost complete loss of biological nitrogen fixation, underscoring its central role in nodule development and identity.

Nodule development is strongly shaped by systemic regulatory networks that adjust nodule number and activity to plant nitrogen status, and many key genes in these pathways have now been identified in major legumes. Autoregulation of nodulation (AON) involves root-derived CLE peptides induced by early nodulation, shoot-localized CLAVATA1-like receptors (HAR1/SUNN/NARK), and downstream signaling components that limit further nodule initiation (Chaulagain and Frugoli, 2021; Zorin et al., 2022). A complementary positive systemic pathway based on CEP peptides and the CRA2 receptor, and a miR2111–TML module, together create a balance between promotion and suppression of nodulation in response to soil nitrogen and whole-plant demand (Chaulagain and Frugoli, 2021; Cardoso-Silva et al., 2025). Comparative and transcriptomic studies in forage crops such as red clover and pea have highlighted additional developmental genes, including large families of nodule-specific cysteine-rich (NCR) peptides and PLAT-domain proteins, whose copy number and expression variation correlate with differences in nitrogen fixation efficiency and nodule traits (Figure 1) (Dinkins et al., 2022; Vlk et al., 2022). Single-cell and cell-type–specific transcriptomics in Lotus japonicus nodules identified hundreds of differentially expressed genes between infected and uninfected cells, including transcription factors and transporters with previously unknown roles, greatly expanding the set of candidates implicated in spatially patterned gene networks that sustain nodule development (Wang et al., 2022). Collectively, these findings delineate a complex, multi-layered regulatory landscape in which conserved core regulators such as NIN are complemented by lineage-specific TFs, peptides, and signaling genes that fine-tune nodule organogenesis across legume crops.

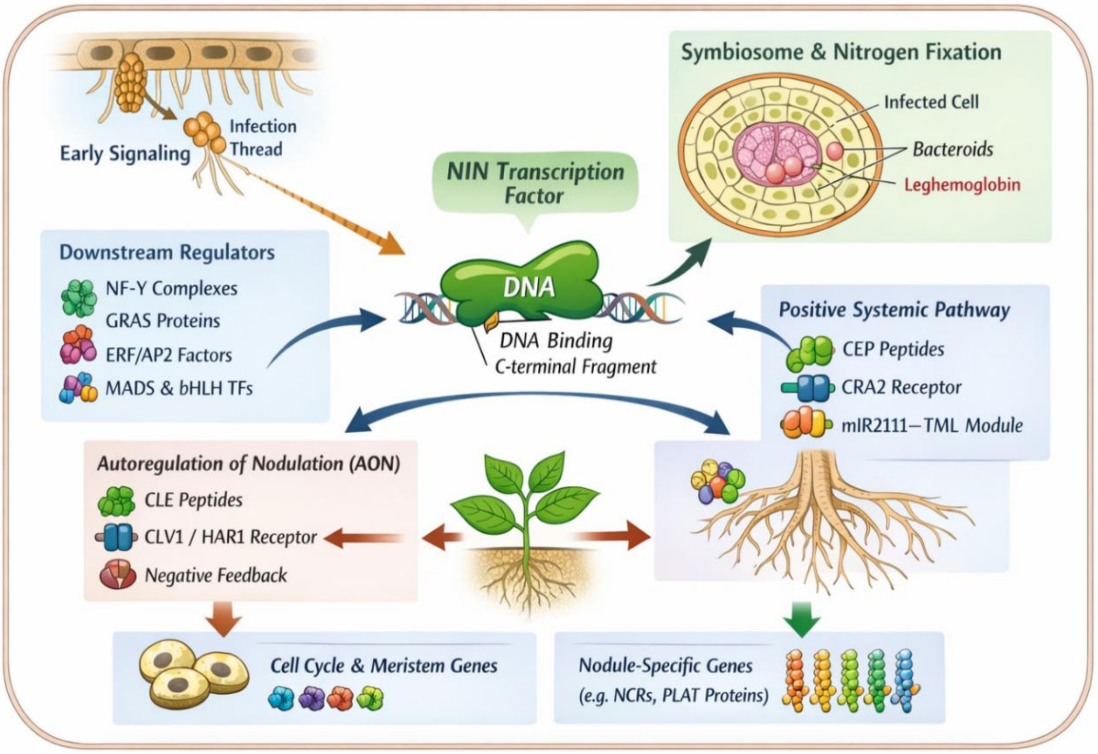

Figure 1 Genes regulating nodule development (Adopted from Dinkins et al., 2022)

### 3.3 Functional genes for nitrogen fixation and metabolism

In mature nodules, suites of functional genes from both plant and bacterial partners underpin nitrogenase activity, oxygen homeostasis, and nitrogen assimilation and export. On the plant side, leghemoglobin genes encode oxygen-binding hemoproteins that create a microaerobic environment compatible with nitrogenase while sustaining high respiration; nodule-specific expression and gene family expansion of leghemoglobins have been documented in multiple legumes, and disruption of these genes impairs nitrogen fixation (Roy et al., 2019). Single cell-type transcriptomics in L. *japonicus* revealed strong enrichment of genes for heme biosynthesis, leghemoglobins, and related pathways in infected cells, and functional tests showed that altering expression of haem biosynthesis regulators affected haem content, nodule coloration, and nitrogen fixation capacity. Enzymes of the carbonic anhydrase–PEPC–malate dehydrogenase pathway, which provides dicarboxylates (especially malate) as carbon and energy sources for bacteroids, are major determinants of nodule carbon metabolism, and their expression and activity respond to whole-plant nitrogen and phosphorus status (Schwember et al., 2019).

Transcriptomic analyses across species indicate coordinated upregulation of genes involved in sucrose metabolism, the oxidative pentose phosphate pathway, and redox homeostasis in nodules, reflecting the tight coupling of carbon supply, oxygen flux, and nitrogenase regulation (Feng et al., 2021).

Functional genomics has also identified a wide array of genes involved in ammonia assimilation, nitrogen transport, and nodule cell differentiation that collectively determine nitrogen fixation efficiency. Genes encoding glutamine synthetase, glutamate synthase, asparagine synthetase, ureide biosynthesis enzymes, and multiple transporter families (for ammonium, amino acids, ureides, and other metabolites) are highly expressed in nodules and display cell-type–specific expression patterns in soybean and L. *japonicus* (Feng et al., 2021; Wang et al., 2022). Single-nucleus and cell-type–resolved datasets in soybean mapped key steps of the ureide synthesis pathway to distinct nodule cell types, revealing spatial compartmentalization of nitrogen metabolism genes during fixation. In addition, large and variable families of NCR peptides in IRLC legumes such as pea and red clover play central roles in bacteroid differentiation, persistence, and, likely, nitrogen fixation efficiency; genomic surveys have identified hundreds of NCR genes clustered in legume genomes, with species-specific expansions and diverse expression patterns, suggesting rapid evolution of these effectors (Zorin et al., 2022). Transcriptomic comparison of red clover genotypes with contrasting biological nitrogen fixation efficiency identified nearly 500 differentially expressed genes, with significant enrichment of tandem- and dispersedly duplicated genes, underscoring the contribution of gene family expansion and regulatory divergence to fixation performance. Finally, regulatory genes, including NAC-type transcription factors controlling nitrate-induced nodule senescence, miRNAs modulating nodule metabolism, and numerous nodule-enriched transcription factors and transporters identified through high-throughput “omics” approaches, form higher-order networks that coordinate nitrogenase activity with plant nutritional and environmental cues (Schwember et al., 2019; Qiao et al., 2023). Together, these functional and regulatory gene sets illustrate how comparative genomics and transcriptomics across legumes are revealing both conserved and species-specific modules that can be targeted to enhance nitrogen fixation and nitrogen-use efficiency in major legume crops.

## 5 Transcriptional Regulation and Functional Validation of Nitrogen Fixation-Related Genes

### 5.1 Transcriptional regulatory networks

Transcriptional regulation of symbiotic nitrogen fixation (SNF) in legumes is organized in hierarchical networks that integrate developmental, nutritional, and environmental signals. Central transcription factors such as NODULE INCEPTION (NIN) and NIN-LIKE PROTEINS act at early stages to couple Nod factor signaling with infection and nodule organogenesis, while downstream networks modulate nodule maturation, nitrogenase activity, and senescence (Qiao et al., 2023). Large-scale expression and co-expression analyses in soybean and other legumes have identified hundreds of transcription factors correlated with SNF traits, including components of the circadian clock (e.g., GmLHY1b), stress-responsive factors, and regulators of lipid and defense signaling, highlighting the polygenic and distributed nature of SNF control (Li et al., 2025). Weighted gene co-expression and independent component analyses have further delineated modules enriched in transcription factors that track variation in nodule number, nodule weight, and nitrogen fixation efficiency, many of which map to domestication sweeps, implying selection on regulatory circuitry during crop improvement (Qiao et al., 2023). In addition, transcriptional networks governing responses to fluctuating mineral nitrogen have been partially resolved, with NAC-type transcription factors (soybean SNAPs; Lotus LjNAC094) identified as master regulators of nitrate-induced nodule senescence, linking external nitrogen status to coordinated reprogramming of senescence-associated and metabolic genes in nodules (Haskett et al., 2025).

Downstream branches of these networks often rely on interconnected transcription factor cascades and feedback loops involving both protein-coding genes and non-coding RNAs. In common bean, the SRS/STY family transcription factor PvSRS10 is transcriptionally activated by NF-Y complexes and by the MADS-box factor PvFUL-like, positioning it within the NIN–NF-Y cascade that conveys early symbiotic signals into developmental responses; protein interaction predictions suggest that other PvSRS members interface with auxin signaling components, providing a mechanistic link between transcriptional control and hormone-mediated nodule morphogenesis (Ayra et al., 2025). MicroRNA-centered modules add additional layers of control: in common bean and soybean, the miR156–SPL and miR172–AP2 circuits form an interconnected cascade, where SPL transcription factors and MADS-domain proteins directly activate MIR172c transcription, and miR172c in turn modulates AP2-like repressors of nodulation. These miRNA–TF feed-forward loops fine-tune nodule number and function and are responsive to nitrogen supply and developmental stage (De Lima et al., 2022). Integrative genomic and transcriptomic analyses across legumes suggest that many nodulation-related TF families (NAC, GRAS, NF-Y, MADS, SRS, WRKY) have undergone lineage-specific expansions, with paralogs showing nodule-enriched and condition-specific expression patterns, consistent with subfunctionalization and specialization within the broader SNF regulatory landscape (Qiao et al., 2016).

### 5.2 Epigenetic regulatory mechanisms

Epigenetic regulation is emerging as a key dimension shaping transcriptional programs during nodulation and nitrogen fixation in major legume crops. Studies in peanut provide one of the first high-resolution views of three-dimensional chromatin organization in roots and nodules, revealing that approximately 2% of the genome switches from a repressive B compartment to an active A compartment in nodules, accompanied by extensive remodeling of topologically associated domains and increased long-range cis-interactions at symbiosis-related loci. Integration of Hi-C with ATAC-seq and histone modification profiling has identified dynamic local open chromatin regions and histone mark changes around nodulation genes, including a chromatin loop mediated by the repressive H3K27me3 mark that appears to regulate NIN expression, and another loop that highlights the nodule-expressed gene *AhMsrA* as a positive regulator of nodulation. DNA methylation also contributes to fine-tuning symbiotic gene activity: in Medicago truncatula, the DNA demethylase DEMETER is essential for proper nodule development, regulating expression of genes involved in nodule differentiation, while nodule-specific cysteine-rich (NCR) peptide genes show extensive hypomethylation correlated with high expression and active histone marks (Wang et al., 2025).

In common bean and soybean, DNA methylation, histone modifications, and small RNAs together form a multilayered regulatory system that influences both nodule initiation and lifespan. Adenine methylation patterns in rhizobial genomes change during symbiosis, and rhizobial tRNA-derived small RNA fragments are transferred into plant cells, where they exploit the host RNA interference machinery to cleave specific mRNA targets and thereby increase nodule number. On the plant side, histone lysine methyltransferases have been implicated in modulating nodule number and bacterial colonization, while numerous nodulation-associated microRNAs (including, but not limited to, miR156 and miR172) regulate transcription factors and transporters important for nodule development and nitrogen metabolism (De Lima et al., 2022). Integrative chromatin-accessibility and expression studies in soybean diversity panels have shown that expression quantitative trait loci for many SNF-related genes are enriched within open chromatin regions in nodules, providing population-level evidence that variation in cis-regulatory elements and local chromatin state underlies natural diversity in SNF performance (Li et al., 2025). Together, these findings indicate that epigenetic remodeling of chromatin architecture, DNA methylation landscapes, histone marks, and small RNA profiles is tightly intertwined with transcription factor networks in controlling where, when, and how strongly nitrogen fixation genes are expressed in legume crops.

### 5.3 Functional gene validation techniques

The rapid growth in candidate nitrogen fixation-related genes identified by comparative genomics and expression profiling has spurred extensive use of functional validation approaches in legumes. Classical forward genetics, often using fast-neutron or chemical mutagenesis, remains a powerful tool: in Medicago truncatula, fast-neutron mutagenesis combined with whole-genome sequencing enabled the identification of a symbiotic nitrogen fixation-defective mutant harboring a large deletion on chromosome 3, providing direct functional links between lost genomic segments and SNF phenotypes (Shen et al., 2023). Transcriptomic analysis of chromosome segment substitution lines in peanut with severely impaired nitrogen fixation similarly uncovered differential expression of key symbiotic regulators (including NIN, EFD and SNF-related transporters) and pinpointed FEN1-encoding a homocitrate synthase essential for nitrogenase activity-as a strong candidate gene underlying the defective phenotype (Nzepang et al., 2025). In red clover, RNA-seq of genotypes with contrasting biological nitrogen fixation efficiency cataloged 491 differentially expressed genes and revealed enrichment of tandemly and dispersedly duplicated genes, which, together with co-expression network analysis, provided a prioritized list of candidates whose functions can be tested in breeding and biotechnology programs.

Targeted reverse-genetic approaches, particularly CRISPR/Cas-based genome editing, have dramatically accelerated functional dissection of SNF genes in both model and crop legumes. In *Lotus japonicus*, CRISPR/Cas9 was used to efficiently disrupt the symbiosis receptor kinase SYMRK and three leghemoglobin genes (*LjLb1–3*), with biallelic and multi-gene knockouts showing strong nodulation and nitrogen fixation defects, thus validating their essential roles and demonstrating the feasibility of multiplex editing in nodules (Wang et al., 2016). Subsequent work using CRISPR/Cas9-generated triple leghemoglobin mutants combined with physiological, biochemical, ultrastructural, and RNA-seq analyses revealed synergistic functions of leghemoglobins in maintaining optimal nitrogen fixation and preventing premature nodule senescence (Figure 2).

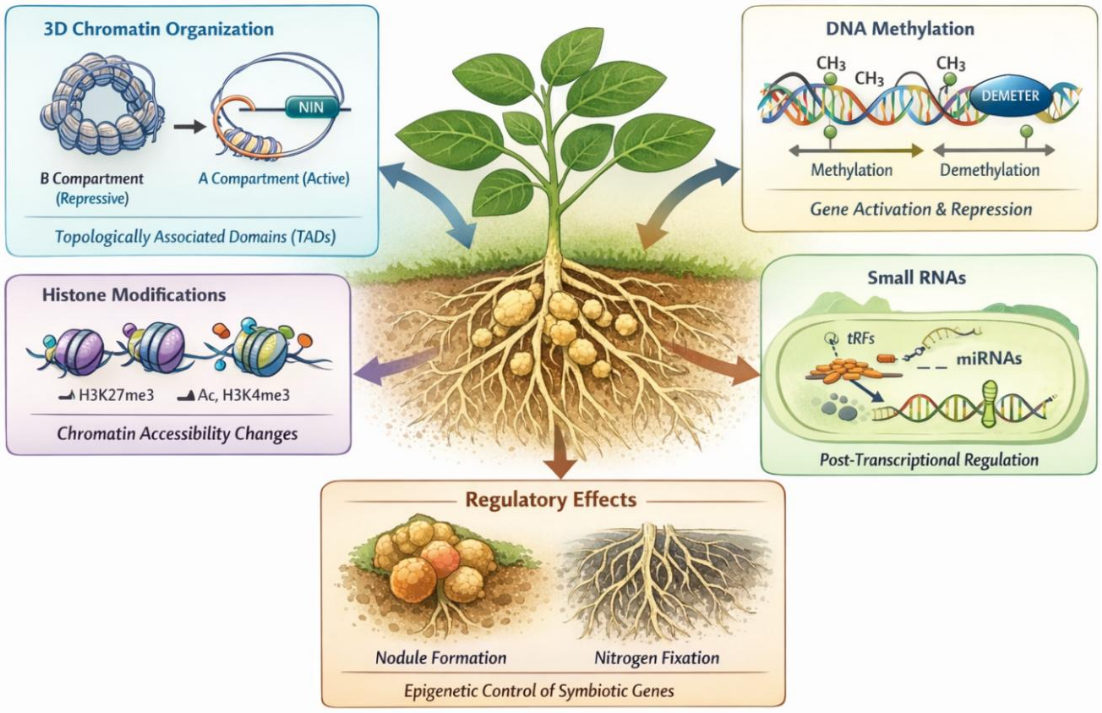

Figure 2 Epigenetic regulation of nitrogen fixation in legumes (Adopted from Wang et al., 2016)

## 6 Case Study: Comparative Genomic Analysis of Key Nitrogen Fixation Genes in Soybean

### 6.1 Comparison of nitrogen fixation genes in soybean and model legumes

Comparative genomic studies across legumes indicate that nitrogen fixation (NF) genes are generally highly conserved, despite extensive polyploidization and genome duplication in some species, including soybean. Analysis of 20 NF-related genes in six legumes (Medicago truncatula, Cicer arietinum, *Lotus japonicus*, Cajanus cajan, Phaseolus vulgaris and Glycine max) showed that orthologs of core nodulation and symbiotic signaling genes are present in all species, with relatively low non-synonymous substitution rates, supporting strong purifying selection on this gene set. Ks-based analyses clearly resolved the ancient genome duplications shared across legumes and the more recent whole-genome duplication (WGD) unique to soybean, which generated numerous NF-related paralogs. This WGD-derived expansion is also evident for several large transcription factor families implicated in nodulation and symbiosis, such as GATA, MYB, bZIP, BES1 and ZF-HD, where soybean contains substantially more members than non-polyploid model legumes (Rizwan et al., 2025). Many of these TFs show tissue-specific or nodule-enriched expression and are co-expressed with canonical symbiosis genes, suggesting that soybean has diversified regulatory components around a relatively conserved NF core (Leng et al., 2025).

Network-level comparisons between soybean and model legumes further reinforce this pattern of conserved core pathways embedded in species-specific regulatory architectures. Cross-species transcriptomic profiling during nitrogen-fixing symbiosis (NFS) and arbuscular mycorrhizal symbiosis (AMS) revealed thousands of genes activated in G. max, M. truncatula and L. *japonicus*, with a conserved “common symbiosis” module but also soybean-specific regulatory circuits (Figure 3) (Wu et al., 2025).

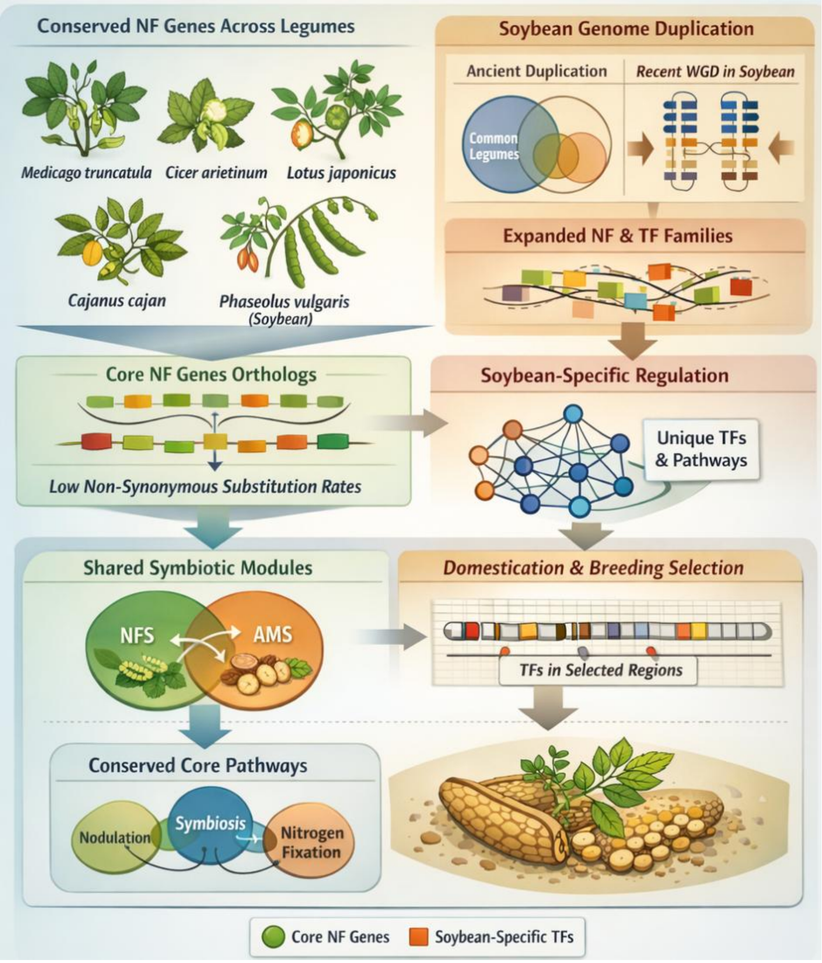

Figure 3 Comparative analysis of nitrogen fixation genes in legumes, highlighting conserved NFgenes across species, soybean genome duplication, shared symbiotic modules, and soybean-specific regulatory networks

Co-expression based gene regulatory networks inferred using machine-learning approaches in G. max and M. truncatula highlighted shared hubs associated with symbiosis, while also identifying soybean-specific regulators, including clock components and defense-related TFs that modulate nodulation and nitrogen fixation (Li et al., 2025). Integrative genomic and transcriptomic analyses in a soybean diversity panel documented that many SNF-associated transcription factors lie within domestication sweeps, in contrast to the more structurally conserved NF core genes, indicating that breeding and domestication in soybean have acted more strongly on regulatory layers than on the underlying symbiotic machinery itself.

### 6.2 Evolutionary analysis of GmNIN and related gene families

Although NIN itself was not the explicit focus of several genome-wide studies, insights into the evolutionary behavior of GmNIN can be inferred from comparative analyses of NF- related genes and of large transcription factor families with documented symbiotic roles. The six-legume comparison of 20 NF- related loci showed that soybean copies of core symbiosis genes generally exhibit low Ka/Ks ratios and often retain multiple paralogs, consistent with strong purifying selection and retention after WGD (Kim et al., 2013). This pattern parallels observations for soybean GATA, bZIP, BES1 and other TF gene families, in which segmental duplication associated with WGD is the primary driver of expansion and most duplicated genes display Ka/Ks < 1, indicating predominantly purifying selection (Gou et al., 2025). For families such as GATA and MYB, integrative phylogenetic and expression analyses across nitrogen-fixing legumes revealed subgroups whose members are specifically co-expressed with NIN and nodulation markers in nodules, pointing to co-evolution of NIN-centered regulatory circuits with lineage-specific TF expansions (Leng et al., 2025; Xia et al., 2025).

At the regulatory-network scale, GmNIN appears embedded within broader transcriptional hierarchies that have diversified through gene family expansion and regulatory rewiring. In soybean, independent component and co-expression analyses across 380 accessions identified modules enriched for transcription factors, including MYB, NAC, GATA, BES1 and clock genes, that are significantly associated with nitrogen fixation phenotypes such as nodule number, nodule weight and nitrogen fixation efficiency (Li et al., 2025). Many of these TFs show evidence of selection during domestication and improvement, whereas the core NF genes they regulate-including GmNIN orthologs-remain more conserved, suggesting that evolutionary tuning of GmNIN function in soybean has occurred largely through changes in dosage, cis-regulation and interaction partners rather than radical coding divergence. Comparative regulatory network analysis between G. max and M. truncatula during AMS and NFS further showed that while NIN-associated modules are shared, soybean has incorporated additional regulators linked to defense and hormonal signaling into these networks, implying that GmNIN operates within an expanded, soybean-specific regulatory context shaped by WGD and subsequent subfunctionalization (Wu et al., 2025).

### 6.3 Implications for molecular breeding to improve nitrogen fixation efficiency in soybean

Insights from comparative and evolutionary genomics of nitrogen fixation genes in soybean have direct implications for molecular breeding. Genome-wide association and mapping studies have revealed that symbiotic nitrogen fixation (SNF) traits-including nodule number, nodule biomass, total N accumulation and percent N derived from the atmosphere-are highly polygenic and environmentally responsive, with many loci of small effect dispersed across the genome. QTLs and TWAS hits for SNF traits are particularly enriched in regions harboring transcription factors, signaling genes and hormone-related regulators rather than structural NF genes, consistent with the notion that variation in regulatory modules around conserved NF cores, such as GmNIN networks, underlies much of the phenotypic diversity in SNF. Integrative analyses combining GWAS, TWAS, eQTLs and ATAC-seq in soybean have further demonstrated that cis-regulatory variants affecting the expression of SNF-related TFs (e.g., GmLHY1b) in open chromatin regions can significantly impact nodulation and nitrogen fixation, providing concrete targets for marker-assisted and genomic selection.

Functional genomic interventions highlight complementary strategies in which knowledge of NF gene evolution can guide precise editing to optimize nodulation and fixation efficiency without compromising yield. Gene editing of systemic nodulation regulators such as CLE receptors (e.g., ric1a/2a alleles) produced soybean lines with moderately increased nodule numbers, improved carbon–nitrogen balance, and higher grain yield and seed protein in multi-year field trials, demonstrating that fine-tuning nodulation control modules can improve both SNF and agronomic performance (Zhong et al., 2024). Similarly, CRISPR/Cas9 editing of the gibberellin receptor gene GmGID1-2 created semi-dwarf soybean lines with enhanced architecture, higher yield and improved nodulation, nitrogenase activity and plant N content, indicating that pleiotropic regulators connecting growth and symbiosis can be leveraged for sustainable yield gains (Tang et al., 2025). Reviews on soybean functional genomics and breeding emphasize that integrating such targeted editing of regulatory hubs (e.g., GmNIN-associated TFs, hormone receptors, clock genes) with genomic prediction using SNF-associated markers offers a promising route to develop cultivars with enhanced nitrogen fixation, optimized plant architecture and reduced fertilizer dependency under variable environments (Zhou et al., 2025).

## 7 Prospects for the Application of Comparative Genomics in Improving Nitrogen Fixation in Legume Crops

### 7.1 Development of molecular markers for nitrogen fixation efficiency

Comparative genomics, coupled with high-throughput phenotyping, is rapidly transforming efforts to dissect and improve nitrogen fixation efficiency in legumes. Genome-wide association studies (GWAS) and diversity panels now routinely reveal dozens to hundreds of loci controlling nodulation traits such as nodule number, nodule biomass, nitrogenase activity, and plant nitrogen content across chickpea, cowpea, common bean and soybean (Herridge and Rose, 2000). In chickpea, GWAS for nodulation and biological nitrogen fixation (BNF) across multiple environments identified numerous stable marker–trait associations (MTAs) for traits like number of nodules and nodule fresh weight, with several SNPs validated in independent populations, illustrating the robustness of these associations. Many MTAs lie in or near genes encoding receptors, kinases and transporters previously implicated in symbiosis, making them strong candidates for functional markers. Similarly, in cowpea, a mini-core collection genotyped at high density revealed significant associations for nodule number, nodule efficiency and nodule dry weight on multiple chromosomes, with positional candidate genes near peak markers encoding proteins associated with BNF.

These discoveries have clear downstream breeding utility. Several studies emphasize that significant SNPs linked to nitrogen fixation traits can be converted into Kompetitive Allele-Specific PCR (KASP) markers to enable rapid and low-cost genotyping in breeding pipelines. Marker-assisted selection (MAS) can then be applied to stack favourable alleles for nodulation and BNF efficiency into elite backgrounds, particularly when trait heritability is moderate and environmental effects are strong, as observed for nodulation traits in cowpea and other legumes (Herridge and Rose, 2000). Reviews of legume genomics and breeding highlight that sequencing, resequencing, SNP discovery and the construction of high-density maps across major legumes provide a framework to develop marker sets tailored to nitrogen fixation traits, moving from anonymous markers to functionally annotated, trait-linked loci (Afzal et al., 2019). Integrating comparative genomics-via identification of conserved and lineage-specific symbiosis genes across species-with GWAS results further refines candidate lists and supports development of cross-species or crop-specific marker panels for nitrogen fixation efficiency (Dwivedi et al., 2014).

### 7.2 Molecular design breeding for nitrogen fixation-related traits

The growing body of genomic resources in legumes enables “molecular design breeding,” in which alleles at multiple loci affecting nitrogen fixation are rationally combined using markers, genomic selection and genome editing. Emerging tools such as genome-wide SNP arrays, genotyping-by-sequencing, and high-density reference genomes have greatly increased the resolution of QTL mapping and GWAS for complex traits, allowing breeders to model the polygenic architecture of nodulation and BNF in silico before constructing crosses (Afzal et al., 2019). Genomic selection (GS) models, trained on multi-environment phenotypes and genome-wide markers, can capture the small-to-moderate additive effects typical of nitrogen fixation traits, thereby outperforming MAS alone when many loci contribute (Pandey et al., 2016). Comparative genomic information from model legumes and multiple crop species helps define core symbiosis networks and orthologous candidate genes, which can be prioritized in GS models or targeted for editing (Dwivedi et al., 2014).

Molecular design breeding also increasingly incorporates functional validation and genome editing to create ideal ideotypes for nitrogen-efficient, climate-resilient legume varieties. Reviews of grain legume genomics stress that advances in host and rhizobial genome sequencing, combined with transcriptomics, are yielding candidate genes controlling symbiotic efficiency that can be introgressed or edited to enhance BNF under stress (Dwivedi et al., 2014; Mahto et al., 2025). For example, association studies in common bean and soybean have identified candidate genes encoding receptor-like kinases and other signaling components underlying SNF-related QTL; functional analysis using CRISPR and transcriptomics is being used to confirm causality and quantify pleiotropic effects. In chickpea, identification of key nodulation regulators such as CaNFP and its strong upregulation under favourable microbial treatments point to clear targets for genome editing to improve nodule formation and nitrogen fixation. Conceptually, molecular design breeding for nitrogen fixation will rely on integrating: (i) comparative maps of nodulation and BNF genes across legumes, (ii) trait-linked markers and genomic prediction, and (iii) precise editing or allele mining of regulatory and structural genes, with the explicit goal of optimizing both plant–microbe compatibility and whole-plant resource allocation to nodules (Dwivedi et al., 2014; Ferguson et al., 2018).

### 7.3 Nitrogen fixation research and sustainable agricultural development

Enhanced biological nitrogen fixation in legumes is central to sustainable agricultural systems that reduce dependence on synthetic fertilizers, lower greenhouse gas emissions, and maintain soil fertility. Global meta-analyses and quantitative reviews show that nitrogen derived from the atmosphere (Ndfa) in legumes can range from 5–99%, with averages near 68% across crops, and around 0.84 kg fixed N per kg total shoot N when belowground contributions are included, underscoring the large potential contribution of legumes to system-level N budgets. Other reviews of legume-driven BNF highlight multiple ecosystem services: improved soil structure, residual nitrogen benefits for subsequent crops, break-crop effects on pests and diseases, and enhanced biodiversity in rotation, intercropping, green manuring and alley-cropping systems (Lindström and Mousavi, 2019). Under organic and low-input systems, BNF is often the principal N source; systematic reviews indicate that optimizing species choice, cropping practice and biomass production is crucial to maximize Ndfa and stabilize yields under variable climate and sanitary conditions.

However, realizing this potential at scale requires aligning genetic gains in nitrogen fixation with agronomic, environmental and socioeconomic realities. Meta-analyses and reviews emphasize that BNF is highly sensitive to ecological context-soil fertility, climate, stand composition and rhizobial compatibility-so that local testing and adaptive management are essential. Climate change adds further constraints, as high temperatures and drought can drastically reduce nodulation and N fixation, calling for breeding of stress-tolerant, high-BNF varieties and the development of climate-smart management strategies. Comparative genomics and breeding for symbiotic efficiency, when coupled with improved inoculant technologies and microbiome management, can help produce cultivars that fix more nitrogen reliably across environments, thereby reducing fertilizer needs and associated emissions (Mahmud et al., 2020). Ultimately, sustained investment in nitrogen fixation research-from gene discovery and marker development to breeding and systems-level evaluation-will be crucial for designing diversified, legume-rich cropping systems that contribute meaningfully to global food security while staying within planetary nitrogen boundaries (Jhu and Oldroyd, 2023).

## 8 Summary and Outlook

Comparative genomics has established that key nitrogen fixation and nodulation genes are widely conserved across major legume crops, despite extensive genome duplication, polyploidy and divergence. Cross-species analyses of core nodulation and symbiotic signaling components, coupled with synteny and orthology mapping, show that a common set of regulatory and structural genes underpins nodule organogenesis, rhizobial infection, bacteroid differentiation and nodule metabolism in both model and crop legumes. Sequence-level comparisons of nitrogen fixation (NF) genes in multiple legumes highlight predominantly purifying selection and retention of duplicated copies, underscoring their essential role in legume biology. At the same time, pan-genomic and transcriptomic studies reveal substantial diversification in gene copy number, expression patterns and regulatory networks, particularly among transcription factors and receptors modulating nodulation and symbiotic nitrogen fixation (SNF). Parallel comparative genomics of rhizobial symbionts demonstrates large, open pan-genomes, extensive accessory gene content and multiple configurations of nodulation and nif clusters, with no simple core “symbiome,” but instead a spectrum of genetic strategies for establishing effective symbioses.

These genomic foundations have been translated into practical tools for dissecting natural variation in nitrogen fixation traits. High-density reference genomes, resequencing of diversity panels and genome-wide association studies (GWAS) have identified numerous loci associated with nodulation, nodule biomass, nitrogenase activity and plant nitrogen status in cowpea, soybean, common bean, chickpea and other legumes, frequently implicating receptors, kinases, transporters and regulatory genes. Many trait-linked SNPs can be converted into efficient markers (e.g., KASP) to support marker-assisted selection, and are increasingly integrated with transcriptome-wide association, eQTL mapping and chromatin accessibility data to pinpoint causal genes and cis-regulatory variants. Advances in functional genomics-including CRISPR/Cas-mediated editing, detailed nodule transcriptomics, and cross-species comparative analyses of nodulation genes-are clarifying the roles of key regulators such as NIN, CaNFP, FEN1 and multiple transcription factor families in controlling symbiotic efficiency, stress responses and host–rhizobium compatibility. Collectively, these efforts position comparative genomics as a central engine for both fundamental insight into the evolution of nitrogen-fixing symbioses and the rational improvement of nitrogen fixation in major legume crops.

Despite substantial progress, current comparative genomic research on nitrogen fixation in legumes faces several important limitations. First, genetic and genomic characterization is heavily biased toward a few species-particularly soybean, common bean, chickpea, pea and model legumes-while many regionally important or underutilized legumes remain genomically under-resourced. Even in well-studied crops, most analyses have focused on aboveground nodulation and SNF traits under controlled or limited field environments, whereas belowground nitrogen contributions, long-term soil fertility effects and performance under diverse management systems are far less quantified. On the microbial side, although comparative genomics of rhizobia and Bradyrhizobium has revealed extensive diversity and highlighted key symbiotic islands, the functional significance of many accessory genes, secretion systems and strain-specific modules for host specificity and field-level nitrogen fixation is still poorly resolved.

Second, nitrogen fixation is strongly context dependent, yet integration of genomic findings with environmental variability, microbiome complexity and agronomic management remains incomplete. Many GWAS and QTL studies report strong genotype-by-environment interactions and moderate heritability for SNF traits, complicating marker deployment and prediction of performance across sites and seasons. Comparative and functional studies often use single elite rhizobial strains and simplified conditions, which may not capture the competitive dynamics within native microbiomes that ultimately determine nodulation and fixation efficiency in farmers' fields. Furthermore, most comparative genomic work has centered on host plants or rhizobia in isolation; truly integrative, dual-genome or community-level approaches that connect host variants, microbial genotypes and soil ecological processes are still rare. Finally, while gene editing and advanced breeding strategies are emerging, regulatory, biosafety and adoption barriers can slow the translation of genomic discoveries into widely grown varieties, particularly in low-income regions that stand to benefit most from improved biological nitrogen fixation.

Future research on comparative genomics of nitrogen fixation in legumes will benefit from more comprehensive, multi-layered approaches spanning species, genomes and environments. Expansion of high-quality reference genomes, pan-genomes and diversity panels across a broader set of major and minor legumes, together with wild relatives, will help uncover hidden variation in nodulation and SNF genes and illuminate how domestication and improvement have shaped symbiotic capacities. Comparative analyses integrating structural variants, presence/absence variants and regulatory sequence diversity with functional genomics (transcriptomics, epigenomics, single-cell and spatial profiling of nodules) will be essential to dissect how conserved symbiotic modules are rewired across lineages and stress conditions. On the microbial side, large-scale comparative genomics of Bradyrhizobium and other rhizobia should be coupled with functional studies of candidate symbiosis and secretion genes to refine host-specific inoculant design and identify microbial traits that enhance competitiveness, persistence and nitrogen fixation under realistic field conditions.

Equally important will be the integration of comparative genomics with systems-level modeling and breeding for sustainability. Process-based and statistical models that link genomic variation in both host and rhizobial partners to BNF outputs, crop productivity and soil nitrogen dynamics under changing climates can guide ideotype design and management strategies. Genomic selection pipelines explicitly incorporating nitrogen fixation traits, stress resilience and interactions with rhizobial communities should be developed and tested in multi-environment trials, especially in low-input systems where legumes can most reduce fertilizer dependency. Future work should also explore translational opportunities beyond classical legumes, building on comparative studies with nodulating non-legumes and attempts to transfer nitrogen-fixing capabilities or associative fixation to cereals. Finally, bridging disciplinary boundaries-linking genomics, agronomy, soil ecology, climate science and socioeconomics-will be crucial to ensure that advances in comparative genomics of nitrogen fixation translate into deployable cultivars, adapted inoculants and cropping systems that contribute tangibly to global food security and environmental sustainability.

## Acknowledgments

Thanks to the reviewers for providing detailed comments and guidance on the manuscript of this study. The reviewers' keen insights into the issues and attention to detail have greatly benefited the authors.

## Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

### References

Adepoju S., 2025, Advancements in legume breeding for nitrogen fixation, Bulletin of University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, Horticulture, 82(2): 21.
https://doi.org/10.15835/buasvmcn-hort:2025.0026

Afzal M., Alghamdi S.S., Migdadi H.M., Khan M.A., Mirza S.B., and El-Harty E.H., 2019, Legume genomics and transcriptomics: from classic breeding to modern technologies, Saudi Journal of Biological Sciences, 27: 543-555.
https://doi.org/10.1016/j.sjbs.2019.11.018

Ayra L., Jiménez-Nopala G., De La Rosa C., Fuentes S., Ramírez M., Leija A., and Hernández G., 2025, The common bean miR172c microRNA, a relevant regulator of the N-fixing symbiosis, is activated by SPL and AGL/MADS-domain transcription factors, Physiologia Plantarum, 3: 177.
https://doi.org/10.1111/ppl.70566

Cardoso-Silva C.B., Quintanilha-Peixoto G.H., Turquetti-Moraes D.C., Almeida-Silva F., and Venancio T.M., 2025, Gene family expansions and nodule-specific expression patterns reveal the recruitment of beta-glucosidases and cytochrome P450 genes to nodulation in soybean, bioRxiv, 7(3): 662952.
https://doi.org/10.1101/2025.07.03.662952

Chaulagain D., and Frugoli J., 2021, The regulation of nodule number in legumes is a balance of three signal transduction pathways, International Journal of Molecular Sciences, 11: 22.
https://doi.org/10.3390/ijms22031117

De Lima F.B., Boldt A.B.W., Kava V., Galli-Terasawa L.V., and Adamoski D., 2022, Epigenetics' role in the common bean (*Phaseolus vulgaris* L.) and soybean (Glycine max (L.) Merr.) nodulation: a review, Plant Molecular Biology Reporter, 21: 1-11.
https://doi.org/10.1007/s11105-021-01333-3

De Lima J.S., De Souza A.R.R., Nunes A.A.A., Rivadavea W.T., Zaro G.C., and Da Silva G.M., 2024, Expanding agricultural potential through biological nitrogen fixation: recent advances and diversity of diazotrophic bacteria, Australian Journal of Crop Science, 18(6): 324-333.
https://doi.org/10.21475/ajcs.24.18.06.p4104

Dinkins R.D., Hancock J.M., Bickhart D.M., Sullivan M.L., and Zhu H., 2022, Expression and variation of the genes involved in rhizobium nodulation in red clover, Plants, 2: 11.
https://doi.org/10.3390/plants11212888

Djordjevic M.A., Mohd-Radzman N.A., and Imin N., 2015, Small-peptide signals that control root nodule number, development, and symbiosis, Journal of Experimental Botany, 66(17): 5171-5181.
https://doi.org/10.1093/jxb/erv357

Dwivedi S.L., Sahrawat K.L., Upadhyaya H.D., Mengoni A., Galardini M., Bazzicalupo M., Biondi E.G., Hungria M., Kaschuk G., Blair M.W., and Ortiz R., 2014, Advances in host plant and rhizobium genomics to enhance symbiotic nitrogen fixation in grain legumes, Advances in Agronomy, 129: 1-116.
https://doi.org/10.1016/bs.agron.2014.09.001

Feng J., Lee T., Schiessl K., and Oldroyd G.E.D., 2021, Processing of NODULE INCEPTION controls the transition to nitrogen fixation in root nodules, Science, 374: 629-632.
https://doi.org/10.1126/science.abg2804

Ferguson B.J., Indrasumunar A., Hayashi S., Lin M.-H., Lin Y.-H., Reid D.E., and Gresshoff P.M., 2010, Molecular analysis of legume nodule development and autoregulation, Journal of Integrative Plant Biology, 52(1): 61-76.
https://doi.org/10.1111/j.1744-7909.2010.00899.x

Ferguson B.J., Mens C., Hastwell A.H., Zhang M., Su H., Jones C.H., Chu X., and Gresshoff P.M., 2018, Legume nodulation: the host controls the party, Plant, Cell and Environment, 42(1): 41-51.
https://doi.org/10.1111/pce.13348

Gou C., Zhang G., Deng Z., Lin C., Li H., Liu H., and Fang X., 2025, Genome-wide analysis of the DNA-binding with one finger gene family reveals soybean expression pattern and functional analysis, International Journal of Molecular Sciences, 3: 26.
https://doi.org/10.3390/ijms26136192

Goyal R.K., Mattoo A.K., and Schmidt M.A., 2021, Rhizobial-host interactions and symbiotic nitrogen fixation in legume crops toward agriculture sustainability, Frontiers in Microbiology, 3: 12.
https://doi.org/10.3389/fmicb.2021.669404

Guo K., Yang J., Yu N., Luo L., and Wang E., 2022, Biological nitrogen fixation in cereal crops: progress, strategies, and perspectives, Plant Communications, 11: 4.
https://doi.org/10.1016/j.xplc.2022.100499

Haskett T.L., Cooke L., Green P., and Poole P.S., 2025, Regulation of rhizobial nodulation genes by flavonoid-independent NodD supports nitrogen-fixing symbioses with legumes, Environmental Microbiology, 2: 27.
https://doi.org/10.1111/1462-2920.70014

Herridge D.F., and Rose I.A., 2000, Breeding for enhanced nitrogen fixation in crop legumes, Field Crops Research, 65: 229-248.
https://doi.org/10.1016/S0378-4290(99)00089-1

Jain D., Jones L.A., and Roy S., 2022, Gene editing to improve legume-rhizobia symbiosis in a changing climate, Current Opinion in Plant Biology, 71: 102324.
https://doi.org/10.1016/j.pbi.2022.102324

Jhu M.-Y., and Oldroyd G.E.D., 2023, Dancing to a different tune, can we switch from chemical to biological nitrogen fixation for sustainable food security?, PLOS Biology, 7: 21.
https://doi.org/10.1371/journal.pbio.3001982

Jithesh T.K., James E.K., Iannetta P.P.M., Howard B.F., Dickin E.T., and Monaghan J.M., 2024, Recent progress and potential future directions to enhance biological nitrogen fixation in faba bean (*Vicia faba* L.), Plant-Environment Interactions, 11: 5-8.
https://doi.org/10.1002/pei3.10145

Kebede E.K., 2021, Contribution, utilization, and improvement of legumes-driven biological nitrogen fixation in agricultural systems, Frontiers in Sustainable Food Systems, 5: 767998.
https://doi.org/10.3389/fsufs.2021.767998

Kim D., Parupalli S., Azam S., Lee S.-H., and Varshney R.K., 2013, Comparative sequence analysis of nitrogen fixation-related genes in six legumes, Frontiers in Plant Science, 2: 4-21.
https://doi.org/10.3389/fpls.2013.00300

Ladha J.K., Peoples M.B., Reddy P.M., Biswas J.C., Bennett A., Jat M.L., and Krupnik T.J., 2022, Biological nitrogen fixation and prospects for ecological intensification in cereal-based cropping systems, Field Crops Research, 283: 108541.
https://doi.org/10.1016/j.fcr.2022.108541

Lee T., Orvošová M., Batzenschlager M., Batista M.B., Bailey P.C., Mohd-Radzman N.A., Gurzadyan A., Stuer N., Mysore K.S., Wen J., Ott T., Oldroyd G.E.D., and Schiessl K., 2024, Light-sensitive short hypocotyl genes confer symbiotic nodule identity in the legume *Medicago truncatula*, Current Biology, 34: 825-840.e7.
https://doi.org/10.1016/j.cub.2024.01.018

Leng J., Xu R., Liu Y., Jiang T., Hu H., Ding Z., and Dai S., 2025, Genome-wide analysis of GmMYB S20 transcription factors reveals their critical role in soybean nodulation, Plants, 1: 14-11.
https://doi.org/10.3390/plants14142240

Lepetit M., and Brouquisse R., 2023, Control of the rhizobium-legume symbiosis by the plant nitrogen demand is tightly integrated at the whole plant level and requires inter-organ systemic signaling, Frontiers in Plant Science, 3: 14-31.
https://doi.org/10.3389/fpls.2023.1114840

Li Q., Guo L., Wang H., Zhang Y., Fan C., and Shen Y., 2019, In silico genome-wide identification and comprehensive characterization of the BES1 gene family in soybean, Heliyon, 5(1): 21-31.
https://doi.org/10.1016/j.heliyon.2019.e01868

Li Y., Pei Y., Shen Y., Zhang R., Kang M., Li D., and Chen Y., 2022, Progress in the self-regulation system in legume nodule development-AON (autoregulation of nodulation), International Journal of Molecular Sciences, 3(23):31-53.
https://doi.org/10.3390/ijms23126676

Lindström K., and Mousavi S.A., 2020, Effectiveness of nitrogen fixation in rhizobia, Microbial Biotechnology, 13(5): 1314-1335.
https://doi.org/10.1111/1751-7915.13517

Liu Y., Wu L., Baddeley J.A., and Watson C.A., 2011, Models of biological nitrogen fixation of legumes: a review, Agronomy for Sustainable Development, 31: 155-172.

Mahmud K., Makaju S., Ibrahim R., and Missaoui A., 2020, Current progress in nitrogen fixing plants and microbiome research, Plants, 9(1): 97.
https://doi.org/10.3390/plants9010097

Mahto R.K., Chandana B.S., Singh R.K., Kumar A., Kumar S., Yadav R.K., Dey D., Hamwieh A., and Kumar R., 2025, Symbiotic nitrogen fixation for sustainable chickpea yield and prospects for genome editing in changing climatic situations, Frontiers in Plant Science, 16: 11-21.
https://doi.org/10.3389/fpls.2025.1621191

Mahto R.K., Chandana B.S., Singh R.K., Talukdar A., Swarnalakshmi K., Suman A., Dey D., and Kumar R., 2025, Uncovering potentials of an association panel subset for nitrogen fixation and sustainable chickpea productivity, BMC Plant Biology, 25: 693.
https://doi.org/10.1186/s12870-025-06244-z

Nelson M.S., and Sadowsky M.J., 2015, Secretion systems and signal exchange between nitrogen-fixing rhizobia and legumes, Frontiers in Plant Science, 6: 491.
https://doi.org/10.3389/fpls.2015.00491

Nzepang D., Cissoko M., Gully D., Hocher V., Rami J.-F., Fall S., Fonceka D., and Svistoonoff S., 2025, Transcriptomic analysis reveals genetic factors underlying impaired symbiotic nitrogen fixation in lines derived from crosses between cultivated peanut (*Arachis hypogaea* L.) and its wild ancestors, BMC Genomics, 26: 11739.
https://doi.org/10.1186/s12864-025-11739-y

Palmero F., Fernández J.A., García F.O., Haro R.J., Prasad P.V.V., Salvagiotti F., and Ciampitti I.A., 2022, A quantitative review into the contributions of biological nitrogen fixation to agricultural systems by grain legumes, European Journal of Agronomy, 136: 126514.
https://doi.org/10.1016/j.eja.2022.126514

Pandey M.K., Roorkiwal M., Singh V.K., Ramalingam A., Kudapa H., Thudi M., Chitikineni A., Rathore A., and Varshney R.K., 2016, Emerging genomic tools for legume breeding: current status and future prospects, Frontiers in Plant Science, 7: 455.
https://doi.org/10.3389/fpls.2016.00455

Pankievicz V.C.S., Irving T.B., Maia L.G.S., and Ané J.-M., 2019, Are we there yet? The long walk towards the development of efficient symbiotic associations between nitrogen-fixing bacteria and non-leguminous crops, BMC Biology, 17: 99.
https://doi.org/10.1186/s12915-019-0710-0

Qiao L., Lin J., Suzaki T., and Liang P., 2023, Staying hungry: a roadmap to harnessing central regulators of symbiotic nitrogen fixation under fluctuating nitrogen availability, aBIOTECH, 5: 107-113.
https://doi.org/10.1007/s42994-023-00123-7

Qiao M., Sun R., Wang Z., Dumack K., Xie X., Dai C., Wang E., Zhou J., Sun B., Peng X., Bonkowski M., and Chen Y., 2024, Legume rhizodeposition promotes nitrogen fixation by soil microbiota under crop diversification, Nature Communications, 15: 47159.
https://doi.org/10.1038/s41467-024-47159-x

Qiao Z., Pingault L., Nourbakhsh-Rey M., and Libault M., 2016, Comprehensive comparative genomic and transcriptomic analyses of the legume genes controlling the nodulation process, Frontiers in Plant Science, 7: 34.
https://doi.org/10.3389/fpls.2016.00034

Ren B., Wang X., Duan J., and Ma J., 2019, Rhizobial tRNA-derived small RNAs are signal molecules regulating plant nodulation, Science, 365(6456): 919-922.
https://doi.org/10.1126/science.aav8907

Rizwan H.M., He J., Nawaz M.A., Lu K., and Wang M., 2025, The members of zinc finger-homeodomain (ZF-HD) transcription factors are associated with abiotic stresses in soybean: insights from genomics and expression analysis, BMC Plant Biology, 25: 6028.
https://doi.org/10.1186/s12870-024-06028-x

Roy S., Liu W., Nandety R.S., Crook A., Mysore K.S., Pislariu C.I., Frugoli J., Dickstein R., and Udvardi M.K., 2019, Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation, The Plant Cell, 32(1): 15-41.
https://doi.org/10.1105/tpc.19.00279

Schulte C.C.M., Borah K., Wheatley R.M., Terpolilli J.J., Saalbach G., Crang N., De Groot D.H., Ratcliffe R.G., Kruger N.J., Papachristodoulou A., and Poole P.S., 2021, Metabolic control of nitrogen fixation in rhizobium-legume symbioses, Science Advances, 7: eabh2433.
https://doi.org/10.1126/sciadv.abh2433

Sharma V., Bhattacharyya S., Kumar R., Kumar A., Ibañez F., Wang J., Guo B., Sudini H.K., Gopalakrishnan S., DasGupta M., Varshney R.K., and Pandey M.K., 2020, Molecular basis of root nodule symbiosis between Bradyrhizobium and "crack-entry" legume groundnut (*Arachis hypogaea* L.), Plants, 9: 276.
https://doi.org/10.3390/plants9020276

Shen Y., Li D., Kang M., Pei Y., Zhang R., Tao W., Huang S., Song W., Li Y., Huang W., Wang D., and Chen Y., 2023, Biological and genomic analysis of a symbiotic nitrogen fixation defective mutant in *Medicago truncatula*, Frontiers in Plant Science, 14: 1209664.
https://doi.org/10.3389/fpls.2023.1209664

Shumilina J., Soboleva A., Abakumov E., Shtark O., Zhukov V., and Frolov A., 2023, Signaling in legume-rhizobia symbiosis, International Journal of Molecular Sciences, 24: 17397.
https://doi.org/10.3390/ijms242417397

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(6468): 1021-1023.
https://doi.org/10.1126/science.aax2153

Tang J., Yang S., Li S., Yue X., Jin T., Yang X., Zhang K., Yang Q., Liu T., Zhao S., Gai J., and Li Y., 2025, Editing a gibberellin receptor gene improves yield and nitrogen fixation in soybean, Journal of Integrative Plant Biology, 68: 75-95.
https://doi.org/10.1111/jipb.70026

Tsoy O., Ravcheev D., Čuklina J., and Gelfand M.S., 2016, Nitrogen fixation and molecular oxygen: comparative genomic reconstruction of transcription regulation in Alphaproteobacteria, Frontiers in Microbiology, 7: 1343.
https://doi.org/10.3389/fmicb.2016.01343

Tsyganov V.E., and Tsyganova A.V., 2020, Symbiotic regulatory genes controlling nodule development in *Pisum sativum* L., Plants, 9: 1741.
https://doi.org/10.3390/plants9121741

Vlk D., Trnĕný O., and Řepková J., 2022, Genes associated with biological nitrogen fixation efficiency identified using RNA sequencing in red clover (*Trifolium pratense* L.), Life, 12: 1975.
https://doi.org/10.3390/life12121975

Wang L., Li R., Liang J., Tian T., Ji J., Chen R., Zhou Y., Fan Q., Ning G., Larkin R.M., Becana M., and Duanmu D., 2022, Single cell-type transcriptome profiling reveals genes that promote nitrogen fixation in the infected and uninfected cells of legume nodules, Plant Biotechnology Journal, 20: 616-618.
https://doi.org/10.1111/pbi.13778

Wang L., Mai C., He S., Niu B., Jia G., Yang T., Xu Y., Ren M., Zhao X., Liu X., and Kong Z., 2025, Dynamic 3D chromatin organization and epigenetic regulation of gene expression in peanut nodules, Journal of Integrative Plant Biology, 67: 2624-2642.
https://doi.org/10.1111/jipb.70007

Wang L., Tian T., Liang J., Li R., Xin X., Qi Y., Zhou Y., Fan Q., Ning G., Becana M., and Duanmu D., 2023, A transcription factor of the NAC family regulates nitrate-induced legume nodule senescence, New Phytologist, 239: 18896.
https://doi.org/10.1111/nph.18896

Wang L., Wang L., Tan Q., Fan Q., Zhu H., Hong Z., Zhang Z., and Duanmu D., 2016, Efficient inactivation of symbiotic nitrogen fixation related genes in Lotus japonicus using CRISPR-Cas9, Frontiers in Plant Science, 7: 1333.
https://doi.org/10.3389/fpls.2016.01333

Wu Z., Sun Y., Yang J., Liu Z., Niu Y., and Zhang X., 2025, Gene expression and regulatory networks provide new insights into the similarity between nitrogen fixing and arbuscular mycorrhizal symbioses, Plant and Cell Physiology.
https://doi.org/10.1093/pcp/pcaf082

Xia M., Tang L., Zhai H., Liu Y., Zhang L., and Chen D., 2025, Genome-wide identification and evolutionary analysis of the GATA transcription factor family in nitrogen-fixing legumes, Plants, 14: 2456.
https://doi.org/10.3390/plants14162456

Zhang M., Liu Y., Shi H., Guo M., Chai M., He Q., Yan M., Cao D., Zhao L., Cai H., and Qin Y., 2018, Evolutionary and expression analyses of soybean basic leucine zipper transcription factor family, BMC Genomics, 19: 4511.
https://doi.org/10.1186/s12864-018-4511-6

Zheng J., Sun L., Wang D., He L., Du W., Guo S., and Wang L., 2023, Roles of a CCR4-NOT complex component GmNOT4-1 in regulating soybean nodulation, Frontiers in Plant Science, 14: 1172354.
https://doi.org/10.3389/fpls.2023.1172354

Zhong C., Hu G., Hu C., Xu C., Zhang Z., and Ning K., 2024, Comparative genomics analysis reveals genetic characteristics and nitrogen fixation profile of Bradyrhizobium, iScience, 27: 108948.
https://doi.org/10.1016/j.isci.2024.108948

Zhong X., Wang J., Shi X., Bai M., Yuan C., Cai C., Wang N., Zhu X., Kuang H., Wang X., Su J., He X., Liu X., Yang W., Yang C., Kong F., Wang E., and Guan Y., 2024, Genetically optimizing soybean nodulation improves yield and protein content, Nature Plants, 10: 736-742.
https://doi.org/10.1038/s41477-024-01696-x

Zhou C., Wang H., Zhu X., Li Y., Zhang B., Tadege M., Wu S., Qi Z., and Xia Z., 2025, Functional genomics: from soybean to legume, International Journal of Molecular Sciences, 26: 6323.
https://doi.org/10.3390/ijms26136323

Zorin E., Kliukova M., Afonin A., Gribchenko E., Gordon M., Sulima A., Zhernakov A., Kulaeva O., Romanyuk D., Kusakin P., Tsyganova A., Tsyganov V., Tikhonovich I., and Zhukov V., 2022, A variable gene family encoding nodule-specific cysteine-rich peptides in pea (*Pisum sativum* L.), Frontiers in Plant Science, 13: 884726.
https://doi.org/10.3389/fpls.2022.884726