New Phytologist

Research

Genomic and transcriptomic insights into legume–rhizobia symbiosis in the nitrogen-fixing tree *Robinia pseudoacacia*

Bin Hu[1]*, Maxim Messerer[2]*, Georg Haberer[2], Thomas Lux[2], Vanda Marosi[2], Klaus F. X. Mayer[2,3], Kevin D. Oliphant[4], David Kaufholdt[4], Jutta Schulze[4], Lana-Sophie Kreth[4], Jens Jurgeleit[4], Robert Geffers[5], Robert Hänsch[1,4] and Heinz Rennenberg[1]

[1]Center of Molecular Ecophysiology (CMEP), College of Resources and Environment, Southwest University, No. 2, Tiansheng Road, Beibei District, 400715, Chongqing, China; [2]Plant Genome and Systems Biology, Helmholtz Center Munich, German Research Center for Environmental Health, D-85764, Munich-Neuherberg, Germany; [3]School of Life Sciences, Technical University Munich, 85354, Freising, Germany; [4]Institute for Plant Biology, Technische Universität Braunschweig, Humboldtstraße 1, D-38106, Braunschweig, Germany; [5]Genome Analytics, Helmholtz-Center for Infection Research (HZI), Inhoffenstraße 7, D-38124, Braunschweig, Germany

Authors for correspondence:
Bin Hu
Email: hubjoe@126.com

Robert Hänsch
Email: r.haensch@tu-bs.de

Received: *29 December 2024*
Accepted: *4 March 2025*

New Phytologist (2025)
doi: 10.1111/nph.70101

Key words: biological N_2 fixation, genome, legume, rhizobia inoculation, *Robinia pseudoacacia* L., transcriptional analysis.

Summary

- *Robinia pseudoacacia* L. (black locust) is a nitrogen (N)-fixing legume tree with significant ecological and agricultural importance. Unlike well-studied herbaceous legumes, *R. pseudoacacia* is a perennial woody species, representing an understudied group of legume trees that establish symbiosis with Mesorhizobium. Understanding its genomic and transcriptional responses to nodulation provides key insights into N fixation in long-lived plants and their role in ecosystem N cycling.
- We assembled a high-quality 699.6-Mb reference genome and performed transcriptomic analyses comparing inoculated and noninoculated plants. Differential expression and co-expression network analyses revealed organ-specific regulatory pathways, identifying key genes associated with symbiosis, nutrient transport, and stress adaptation.
- Unlike *Medicago truncatula*, which predominantly responds to nodulation in roots, *R. pseudoacacia* exhibited stem-centered transcriptional reprogramming, with the majority of differentially expressed genes located in stems rather than in roots. Co-expression network analysis identified gene modules associated with "leghemoglobins", metal detoxification, and systemic nutrient allocation, highlighting a coordinated long-distance response to N fixation.
- This study establishes *R. pseudoacacia* as a genomic model for nodulating trees, providing essential resources for evolutionary, ecological, and applied research. These findings have significant implications for reforestation, phytoremediation, forestry, and sustainable N management, particularly in depleted, degraded, and contaminated soil ecosystems.

Introduction

Nitrogen (N) is an essential macronutrient that limits plant growth and primary productivity across ecosystems (Elser *et al.*, 2007; LeBauer & Treseder, 2008). Although dinitrogen (N_2) comprises *c.* 78% of the Earth's atmosphere, most plants are unable to utilize it directly. Biological N fixation (BNF) enables certain plant species to overcome this limitation by forming symbiotic relationships with N-fixing bacteria. In legumes, this process is mediated by rhizobia, which colonize root nodules and differentiate into bacteroids, the symbiotic form of the bacteria responsible for N fixation (Oldroyd & Downie, 2006). Within these nodules, bacteroids convert atmospheric N into ammonia, which is subsequently assimilated by the plant. BNF plays a crucial role in global nutrient cycling and food security, accounting for > 25% of global primary crop production (Vitousek *et al.*, 2002; Galloway *et al.*, 2004; Ferguson *et al.*, 2010) and supplying over 50% of N accumulated in tropical forests (Batterman *et al.*, 2013). Global estimates suggest that BNF contributes *c.* 139 to 175 million tons of N per year to terrestrial ecosystems (Tate, 2020), with legumes alone fixing *c.* 50 million tons annually, nearly half the amount provided by synthetic fertilizers (EL Sabagh *et al.*, 2020).

In N-fixing nodules, root cells and the rhizobia symbiont engage in a tightly regulated exchange of metabolites and signals to establish the biochemical conditions necessary for N reduction and assimilation. This process is controlled by coordinated gene expression in both the legume host and rhizobia, regulating numerous biochemical and molecular pathways (Colebatch *et al.*, 2004; Küster *et al.*, 2004; Benedito *et al.*, 2008). Extensive research on these regulatory mechanisms has primarily focused on herbaceous model legumes such as *Lotus japonicus* and

*These authors contributed equally to this work.

Medicago truncatula, as well as crop species such as *Glycine max*, to elucidate the fundamental processes underlying the establishment and maintenance of legume–rhizobia symbioses (Martin *et al.*, 2017; Ferguson *et al.*, 2019; Quilbé *et al.*, 2021; Yang *et al.*, 2022). These studies have identified key regulators of nodulation, including nodule inception, which controls early nodule organogenesis (Schauser *et al.*, 1999); early nodulin, which regulates infection thread formation and cell differentiation (Fang & Hirsch, 1998); nuclear factor-Y, which integrates N signaling to control nodule development (Soyano *et al.*, 2013); and nodule cytein-rich peptides, which modulate bacteroid differentiation (Van de Velde *et al.*, 2010). However, the extent to which these transcriptional programs are conserved or modified in woody legumes remains unknown.

Biological N fixation is accompanied by systemic metabolic shifts that influence whole-plant physiology, including the reallocation of N and sulfur (S) assimilation from leaves to roots (Kalloniati *et al.*, 2015), increased CO_2 fixation and assimilation in leaves, enhanced respiration in roots, partial refixation of CO_2 produced during root energy metabolism, and its subsequent reassimilation in leaves (Fotelli *et al.*, 2011). Despite these advances, little is known about how N fixation is regulated in woody legumes. Establishing an effective symbiosis in trees likely involves not only metabolic reprogramming but also significant adjustments in membrane transport and systemic nutrient signaling, which remain largely unexplored in this plant group (Savage *et al.*, 2016; Liesche *et al.*, 2017; Gani *et al.*, 2021). Unlike herbaceous species, woody legumes tend to localize N and sulfur assimilation in root tissues, reducing the need for nutrient reallocation from leaves following nodule establishment (Rennenberg *et al.*, 2007; Kalloniati *et al.*, 2015; Liu *et al.*, 2022).

Robinia pseudoacacia (black locust) is a N-fixing woody legume with a global distribution and significant ecological impact (Richardson & Rejmanek, 2011; Li *et al.*, 2014; Nentwig *et al.*, 2018). It is both an invasive species and an economically valuable tree, known for rapid growth, high drought tolerance, and the ability to thrive in nutrient-poor soils. These traits make it a promising pioneer species for the restoration of degraded, marginal, and contaminated land (Vítková *et al.*, 2017; Liu *et al.*, 2020) and a suitable model for studying woody legume–rhizobia symbioses (Liu *et al.*, 2020). Although a *Robinia* genome was recently published (Wang *et al.*, 2023), it lacks key comparative analyses with both herbaceous legumes and non-N-fixing trees, limiting its utility in understanding the genomic basis of nodulation and N fixation in a broader evolutionary context. Understanding these differences is critical for establishing a molecular framework for further studies on *Robinia* cultivars, particularly in the context of nodulation and N fixation.

In this study, we present a high-quality genome assembly of *R. pseudoacacia* and analyze its transcriptomic response to rhizobia inoculation. Our aim was to establish *Robinia* as a model species for N-fixing woody legumes by (1) comparing its genome to those of other temperate plant species, including a non-N-fixing tree and two N-fixing herbaceous legumes; (2) elucidating the general molecular features of *Robinia*'s symbiotic N fixation through spatial RNA-seq analysis of multiple plant tissues, including roots, stems, and leaves; and (3) identifying key transport processes that are modified in response to symbiosis. These findings provide new insights into the molecular mechanisms of N fixation in a woody legume and may serve as a reference for broader studies on symbiotic N fixation across both woody and herbaceous legumes.

Materials and Methods

Plant materials

Robinia pseudoacacia L. seeds used in this study were purchased in October 2013 from a commercial tree nursery in Fushun City (41°51′N, 123°56′E), Liaoning Province in Northeast China, and subsequently transferred to collaborating laboratories in P.R. China, Germany, and Greece. As is the case for many *Robinia* plantations established decades ago, the exact genetic background of the provenance is unknown. However, recent studies have shown that *Robinia* seed provenance from the Dongbei (DB) region exhibits superior physiological and biochemical characteristics under nutritional stress and heavy metal exposure conditions, making it a promising source for the phytoremediation of degraded, nutrient-depleted, and contaminated soils (Liu *et al.*, 2024a, 2024b). The plantation sites near Fushun City are characterized by a temperate humid to semihumid continental monsoon climate (Lin *et al.*, 2017), with an annual precipitation of *c.* 800 mm. The soil at these sites is fertile, with a field water capacity of 27.1%, a mean total N content of 1.8 g kg^{-1}, total phosphorus of 0.65 g kg^{-1}, and organic carbon of 41.3 g kg^{-1}, all measured at a 0–10 cm soil depth (Duan *et al.*, 2010).

Robinia cultivation and *rhizobia* inoculation in a glasshouse

Seed germination followed the protocol described by H. Sun *et al.* (2023). Briefly, healthy seeds of *Robinia* DB provenance, genotype DB214, with a similar size were selected and soaked in concentrated sulfuric acid for 10 min. They were then washed with sterile water and subjected to surface sterilization using sodium hypochlorite, followed by further rinsing with sterile water. The seeds were placed on Petri dishes with moistened filter paper and incubated in a climate chamber (RLD-1000E-4; Ledian Ltd, Ningbo, China) at 25°C with a 12-h photoperiod at 400 μE m^{-2} s^{-1} photosynthetically active radiation (PAR). After germination, 24 seedlings were selected and transplanted into 20-cm-diameter pots containing 2.5 kg of a mixed soil substrate (95% sand, 5% vermiculite; Jialing River, Chongqing, China). The substrate had a water-holding capacity of 15.33% (Bi *et al.*, 2014). To prevent fungal and bacterial contamination, a carbendazim and thiodiazole-copper (95%) mixture was applied at 0.016 kg l^{-1} (w/v) (Luyiyuan Technology Development Co., Beijing, China). Previous studies have confirmed that these treatments effectively eliminate microbial contaminants while allowing successful nodulation in *Robinia* under controlled conditions (Heju, 2005; Yuan *et al.*, 2022; Ramangouda *et al.*, 2023; Liu *et al.*, 2024a, 2024b). Given that nodulation has been observed in multiple studies following similar sterilization treatments, we

consider it unlikely that the applied fungicide mixture significantly affected rhizobia colonization.

Robinia seedlings were cultivated in a glasshouse from July to December 2020, with an average temperature of 25–35°C and a 12-h photoperiod at 500 μE m^{-2} s^{-1} PAR. From 1 July to 1 October, seedlings received 250 ml of a full nutrient solution (0.05 mM K_2SO_4, 0.5 mM KH_2PO_4, 0.25 mM $MgSO_4{\cdot}7H_2O$, 1 mM $CaCl_2$, 0.01 mM ferric citrate, 0.3 mM KNO_3, 0.1 mM NH_4NO_3, 2 μM H_3BO_3, 1 μM $MnSO_4{\cdot}H_2O$, 0.2 μM $CuSO_4{\cdot}5H_2O$, 0.1 μM $CoSO_4{\cdot}7H_2O$, 0.5 μM $ZnSO_4{\cdot}7H_2O$, and 0.1 μM $Na_2MoO_4{\cdot}2H_2O$) per pot every 14 d and 250 ml of deionized water per pot every 3–5 d without seepage to avoid leaching of nutrients from the bottom of the pots (modified from Flemetakis *et al.*, 2006; Mariangela *et al.*, 2011). The N concentration (0.5 mM) was adapted from Wang *et al.* (2018) and reflects conditions found in the Loess Plateau of northwest China (Liu *et al.*, 2013; Cao & Chen, 2017). In this study, nodules successfully formed under this N concentration, as shown in Supporting Information Figs S1, S2, indicating that this concentration is sufficient to engage symbiosis.

Seedlings were grown for 21 wk before rhizobia inoculation to ensure robust root development suitable for nodulation, as *Robinia* exhibits a slower early growth phase than herbaceous legumes. For inoculation, *Mesorhizobium huakuii* QD9 (Fig. S1; Liu *et al.*, 2019) was used. The strain was initially cultured on a Yeast Mannitol Agar solid medium (mannitol 10 g l^{-1}, yeast extract 0.12 g l^{-1}, NaCl 1 g l^{-1}, $MgSO_4$ 0.2 g l^{-1}, K_2HPO_4 0.5 g l^{-1}, and agar 20 g l^{-1}; pH = 6.8) (Vincent, 1970) and subsequently transferred to a Tryptone Yeast extract (TY) liquid medium (tryptone 6 g l^{-1}, yeast extract 3 g l^{-1}, and $CaCl_2$ 0.38 g l^{-1}) (Franzini *et al.*, 2013). The culture was incubated at 28°C on a shaker for 48 h before being centrifuged at 2000 ***g*** for 10 min at 25°C. The resulting bacterial pellet was resuspended in sterile water to an OD_{600} of 1.0. Each seedling's root was inoculated with 10 ml of this bacterial suspension using a sterile syringe. The first inoculation was performed 21 wk after the start of cultivation, followed by three additional inoculations at 7- to 10-d intervals. To ensure that any effects were specific to bacterial inoculation rather than to the TY medium itself, the sterilized TY medium was added to noninoculated plants in equal volumes (Fig. S2).

A total of two experimental treatments (inoculated vs noninoculated) were conducted, each with four replicates (n = 4), totaling eight pots. All four replicates per treatment were cultivated; however, only three were used for downstream analysis. To confirm that noninoculated plants did not establish symbiosis, we verified the complete absence of nodules in control seedlings (Fig. S2). However, while nodule formation is a strong indicator of successful rhizobia colonization, we acknowledge that without molecular validation, the possibility of undetected low-level colonization cannot be entirely ruled out.

Plant tissue sampling was conducted 5.5 months after seedling cultivation, corresponding to 21 d post inoculation (dpi). Whole seedlings were carefully removed from pots to minimize root damage and washed thoroughly with distilled water. Afterwards, root nodules from each seedling were counted and the total fresh weight of the entire plant was determined. Leaves, stems, roots, and nodules were dissected, transferred into 10-ml centrifuge tubes, immediately frozen in liquid N, transported to the laboratory, ground into a fine powder under liquid N using a freezer mill (Spex, Metuchen, NJ, USA), and stored at −80°C until further total RNA isolation and physiological analysis. For dry weight determination and further total N analysis, aliquots were dried at 65°C for 72 h to weight constancy. The tissue water contents and total N contents were determined according to Liu *et al.* (2024a), and the basic plant physiological parameters and rhizobia nodulation are shown in Table S1.

In vitro cultivation of *Robinia*

Seed germination was initiated by shaking seeds in concentrated H_2SO_4 for 20 min, followed by six washes with distilled H_2O. Surface sterilization was performed using a 6% NaClO solution (containing 0.025% Tween-20) for 15 min, followed by three additional rinses in sterile water. Plants were grown at 23°C under a 16-h photoperiod (120 μE m^{-2} s^{-1}) on a solid Murashige & Skoog (MS) medium (Murashige & Skoog, 1962). After seven days of germination, seedlings were transferred to culture tubes for 2 wk before shoot tips were transferred to culture jars for an additional 4 wk. The fast-growing DB genotype DB214 was used for further *in vitro* propagation. Cut shoot sections containing one leaf were transferred individually to a fresh MS medium for 4 wk, resulting in the generation of 636 individuals after 4 months. Chlorophyll levels were reduced by maintaining plants in darkness for 2 d before harvesting, followed by immediate freezing and grinding 45 g of leaves in liquid N for DNA isolation and sequencing.

DNA sequencing and assembly

High molecular weight (HMW) DNA was extracted from leaf tissue of a single *Robinia* isolate using a two-step process: First, nuclei were isolated, followed by genomic DNA purification using a cetyltrimethylammonium bromide-based extraction method. The quality of the extracted HMW-DNA was verified by pulsed-field gel electrophoresis, ensuring fragment sizes greater than 50 kbp. DNA sequencing was performed using the PacBio HiFi platform by the subcontractor NRGene (https://www.nrgene.com), generating 2 107 111 HiFi reads totaling 28.37 Gb of genomic sequence, corresponding to *c.* 40× coverage of the *Robinia* genome, assuming a genome size of *c.* 700 Mb. The mean and median read sizes were 13 436 bp and 12 935 bp, respectively.

Reads with a minimum length of 1.5 kb were subsequently assembled using *Hifiasm* with default parameters (Cheng *et al.*, 2021). To minimize redundancy and remove potential phase duplicates, we applied *purge_dups* (https://github.com/dfguan/purge_dups), following a standard genome assembly pipeline. Briefly, HiFi reads were remapped to the primary contigs using *minimap2* (Li, 2018) to generate read depth distributions, and a self-alignment of contigs was performed to identify and remove putative haplotigs and overlaps. The final purged assembly was obtained using scripts provided by the *purge_dups*

pipeline. To assess the completeness and quality of the genome assembly, we performed a Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis using *busco v5* with the *eudicot_odb10* dataset in both genome and protein modes for the primary assembly and protein-coding annotation, respectively (Simão *et al.*, 2015). To estimate genome size and heterozygosity, we performed 21-mer analysis using JELLYFISH 2.0 (Marçais & Kingsford, 2011) across the entire HiFi read dataset and retrieved a frequency histogram using standard protocols (https://github.com/gmarcais/Jellyfish). Subsequently, we used the GENOMSCOPE2 pipeline (https://github.com/tbenavi1/genomescope2.0; Ranallo-Benavidez *et al.*, 2020) to estimate genome-wide k-mer-based statistics.

RNA extraction and library generation

To investigate gene expression changes, spatial transcriptome analysis was conducted on nodules, leaves, stems, and roots from both rhizobia-inoculated and noninoculated treatment groups, with three biological replicates per condition ($n = 3$). Root-associated nodules and plant tissue samples were collected from 5.5-month-old seedlings, 21 days post inoculation. Total RNA was extracted from nodule and plant tissue samples using RNAiso PLUS reagent (No. 9109; Takara Bio Inc., Dalian, China) according to the manufacturer's protocol. RNA purity and concentration were determined using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA) (X. Sun *et al.*, 2023), and RNA integrity was assessed using a Bioanalyzer 2100 (Agilent Technologies, Waldbronn, Germany) to determine the RNA integrity number (Sun *et al.*, 2023b).

RNA sequencing libraries were prepared from 1 μg total RNA using the Ultra II Directional-RNA Seq Library Kit (NEB, Ipswich, MA, USA) following the manufacturer's protocol. Sequencing was performed on an Illumina NovaSeq 6000 using the NovaSeq 6000 S1 Reagent Kit (100 cycles, paired end run) with an average sequencing depth of *c.* 50 million reads per RNA sample (Modi *et al.*, 2021).

Gene annotation

Structural gene annotation was performed using a combination of *de novo* gene prediction and homology-based approaches, integrating RNA-seq, Oxford Nanopore Technologies (ONT) DirectRNA reads (dRNA), and protein datasets. RNA-seq data were initially mapped using STAR v.2.7.8a (Dobin *et al.*, 2013), and assembled into transcripts using STRINGTIE (v.2.1.5, parameters -m 150-t -f 0.3; Kovaka *et al.*, 2019). Fabaceae protein sequences from publicly available datasets (UniProt, https://www.uniprot.org, downloaded on 7 March 2022) were aligned to the genome sequence using GENOMETHREADER (v.1.7.1, arguments -startcodon -finalstopcodon -species rice -gcmincoverage 70 -prseedlength 7 -prhdist 4; Gremme *et al.*, 2005).

Oxford Nanopore Technologies dRNA datasets were assembled into transcripts following the recommended workflow (https://github.com/epi2me-labs/wf-transcriptomes) and aligned to the genome using GMAP (v.2018-07-04; Wu & Watanabe, 2005). Transcripts from RNA-seq, dRNA, and aligned protein sequences were merged using CUFFCOMPARE (v.2.2.1; Ghosh & Chan, 2016) and further refined with STRINGTIE (v.2.1.5, parameters --merge -m150). TRANSDECODER (v.5.5.0; http://transdecoder.github.io) was used to identify potential open reading frames and predict protein sequences.

Ab initio gene annotation was performed using AUGUSTUS (v.3.3.3; Hoff & Stanke, 2019), with additional predictions from GENEMARK (v.4.35; Ter-Hovhannisyan *et al.*, 2008). To minimize overprediction, hints were generated using RNA-seq, protein, and dRNA datasets (Hoff & Stanke, 2019). A species-specific AUGUSTUS training model was generated from a manually curated gene set. Gene predictions were integrated using EVIDENCEMODELLER (v.1.1.1; Haas *et al.*, 2008), with assigned weightings as follows: *ab initio* (Augustus: 5, GeneMark: 2), homology based (10), and comparative *ab initio* (7). Two rounds of PASA (v.2.4.1; Haas *et al.*, 2008) were conducted to refine untranslated regions (UTRs) and alternative splicing isoforms, using transcripts generated via genome-guided TRINITY (v.2.13.1; Grabherr *et al.*, 2011) and dRNA datasets.

Functional annotation was performed using BLASTP (ncbi-blast-2.3.0+, parameters -max_target_seqs 1 -evalue 1e–05; Altschul *et al.*, 1990) to compare predicted protein sequences to reference datasets (Uniprot Magnoliophyta, reviewed/Swissprot, downloaded 3 August 2016; and Fabaceae, downloaded 7 March 2022; https://www.uniprot.org). Candidate genes were classified as complete and valid protein-coding genes, noncoding transcripts, pseudogenes, or transposable elements (TEs). Transposable element identification was further refined using PTREP (release 19; http://botserv2.uzh.ch/kelldata/trep-db/index.html), with only hits displaying an *e*-value below 10^{-10} considered significant. Additionally, functional annotation was performed using the AHRD pipeline (https://github.com/groupschoof/AHRD) and MERCATOR (Schwacke *et al.*, 2019).

Predicted proteins were classified into high-confidence (HC) and low-confidence (LC) categories. High-confidence proteins were those with subject or query coverage above 80% in the UniMag database, or those without BLAST hits in UniMag but present in UniFaba and absent from PTREP. Low-confidence sequences include incomplete protein sequences with BLAST hits in UniMag or UniFaba but not in PTREP. A secondary refinement step promoted LC proteins with an AHRD score of three stars to high confidence. This annotation approach ensured comprehensive identification of protein-coding genes, UTRs, TEs, and alternative isoforms in *R. pseudoacacia*.

Identification of orthologs

Orthologous genes between *R. pseudoacacia* L., *M. truncatula* Gaertn., *G. max* (L.) Merr., and *Populus trichocarpa* Torr. & Gray were identified using GENESPACE (v.1.2.3; Lovell *et al.*, 2022). Genome and annotation data were obtained from Ensembl Plants (https://plants.ensembl.org). GENESPACE internally employs ORTHOFINDER (v.2.5.5; Mohammad *et al.*, 2022) to identify orthologous relationships among the input datasets. BEGINNING with v.2.4.0, ORTHOFINDER includes phylogenetic

hierarchical orthogroups, which provide increased accuracy over previous orthogroup classification methods.

Processing of RNA-seq data

Preprocessed ONT long reads were analyzed using the Nanopore Transcriptome DE pipeline (https://github.com/nanoporetech/pipeline-transcriptome-de; Love *et al.*, 2018) with default parameters. For differential gene expression, the pipeline uses EDGER (Robinson *et al.*, 2010) with the design '~condition'.

To functionally categorize gene expression patterns, we conducted Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis using the function camera() of the R package LIMMA (Ritchie *et al.*, 2015), mapping genes to metabolic and signaling pathways. Gene ontology (GO) enrichment analysis was conducted using the R package TOPGO (Alexa *et al.*, 2006) with the elim method, which accounts for hierarchical GO structures to minimize false positives. Heatmaps were generated using the PHEATMAP package in R (https://www.rdocumentation.org/packages/pheatmap/versions/1.0.12/topics/pheatmap), in which color intensities represent $-\log_{10}$ (*P*-value) for significant enrichment terms.

Weighted correlation network analysis (WGCNA) was performed using the R-package WGCNA (Langfelder & Horvath, 2008) to identify co-expressed gene modules and detect gene regulatory networks underlying symbiotic N fixation. The analysis included the top 5% of genes with the highest variance across samples, resulting in six distinct co-expression modules, categorized into two major clusters. The first cluster contained the blue (nodules and roots), green (nodules), and yellow (roots and stems of inoculated plants) modules, while the second cluster included the brown (leaves of inoculated plants), turquoise (leaves in general), and pink (stem under control conditions) modules.

Comparison between *Robinia* and *Medicago*

NCBI SRA datasets (PRJNA213402, PRJNA524899, and PRJNA554677) were selected for a comparative transcriptomics analysis, representing previously published nodulation-related RNA-seq datasets from *M. truncatula* grown under conditions comparable to those in this study. These datasets were chosen based on experimental similarity and sequencing depth to ensure compatibility. FASTQ files were downloaded, checked for quality control, and trimmed using FASTP (v.0.23.2; Chen, 2023). Reads were then mapped to the *M. truncatula* reference GENOME (v.4.0.56; Ensembl Plants) using SALMON (v.0.14.2; Patro *et al.*, 2017), and transcript abundances were quantified as transcripts per million (TPM).

From a total of 47 899 genes in *Medicago* and 72 352 for *Robinia*, 28 972 genes for *Medicago* and 27 406 for *Robinia* were identified as expressed (TPM > 0.5 in minimum two biological replicates), ensuring sufficient coverage of active transcriptional modules. Lowly expressed genes were filtered, to reduce noise, and variance-stabilizing transformation was performed using the R package DESEQ2 (v.1.34.0; Love *et al.*, 2014) with the design *'~tissue + treatment'* and the argument *'blind = FALSE'*. The 21 *Robinia* RNA-seq samples were processed under identical criteria to allow a direct comparison.

Co-expression networks were built separately for the two species using the WGCNA package (v.1.69; Langfelder & Horvath, 2012). To achieve scale-free topology, soft power threshold was determined as *Robinia* = 17; *Medicago* = 5. Unsigned networks were built using Pearson's correlation, capturing both positive and negative co-expression relationships. Modules were defined using *dynamicTreeCut* and *TOMtype* methods, with a minimum module size of 30, 'unsigned' network type and otherwise default settings. The analysis yielded 54 modules in *Robinia* and 61 modules in *Medicago*, with a gray module (containing 5484 genes in *Robinia*) representing unclassified genes. In *Medicago*, all genes were assigned to modules, indicating a stronger modular structure. Modules were merged using a cut height of 0.4, reducing the number to 24 in *Robinia* and 38 in *Medicago*. Modules were named according to species-specific conventions, such as 'Rob1' for *Robinia* and 'Med1' for *Medicago*, with the gray module excluded from further analysis.

Module eigengenes were calculated using the moduleEigengenes function of WGCNA, summarizing module expression as the first principal component of each module. Pearson's correlation and Fisher's exact test were applied to determine relationships between module eigengenes and binarized tissue categories.

Functional annotations for the longest, most representative isoforms of *M. truncatula* and *R. pseudoacacia* were assigned using MERCATOR4 (v.5.0; Schwacke *et al.*, 2019). Overrepresentation analysis for selected gene sets and modules was performed using the CLUSTERPROFILER R package (v.4.6; Wu *et al.*, 2021), applying the Benjamini–Hochberg false discovery rate (FDR) correction with a *P*-value cutoff of 0.05.

To compare conserved transcriptional networks, an overlap table was generated using a one-sided Fisher's exact test, identifying statistically significant co-expression modules shared between *Medicago* and *Robinia*. These overlapping modules provide insights into conserved regulatory mechanisms of nodulation in woody and herbaceous legumes, highlighting core symbiosis-related gene clusters while also revealing species-specific adaptations.

Results

Characteristics of *Robinia pseudoacacia* genome

To establish a genomic resource for *Robinia*, we assembled *c.* 2×10^6 PacBio HiFi reads totaling 28 Gb of genomic sequence into a high-quality genome. Our assembly complements a recently published assembly of a Northwest China variety (Wang *et al.*, 2023) by representing the genotype of the Northeast China (DB) region. The final primary assembly counted 680 contigs, totaling 699.6 Mb, with a contig N50 of 58.4 Mb. The 12 largest contigs range from 23.3 to 96.1 Mb. K-mer frequency statistics estimated a heterozygosity rate of 1.34% and a genome size of 645 Mb (Fig. S3). The completeness of our genome assembly

Table 1 Statistics of the assembly and gene annotation for the eudicot BUSCO dataset and annotation of *Robinia pseudoacacia*.

BUSCO class	Genome	Annotation HC	Annotation HC + LC
Complete	2297	2246	2264
Complete single copy	2193	2158	2171
Complete duplicated	104	88	93
Fragmented	5	21	23
Missing	24	59	39
Total	2326	2326	2326
Annotation	**Total**	**High**	**Low**
Genes	72 352	35 670	36 682
mRNAs	99 146	61 503	37 643
CDS	443 478	361 996	81 482
Exons	470 150	386 604	83 546
3′UTR	54 757	50 872	3885
5′UTR	60 727	57 218	3509

HC and LC refer to high- and low-confidence gene predictions.

is supported by a high proportion of complete BUSCO genes (98.8%) for the eudicot gene set.

Six of the largest contigs contained telomeric repeat clusters at both the 3′- and 5′-ends, suggesting near chromosome-level assemblies. Two additional large contigs exhibit telomeric sequences at only one end, with the opposite end enriched in ribosomal internal transcribed spacer sequences. These contigs (scaffold 000005 and scaffold 000008) also contain a high proportion of leucine-rich repeat genes at these termini, indicating either genuine structural features or potential mis-assemblies due to highly similar large-scale repeats (Fig. S4). Cytological studies have previously reported a *Robinia* karyotype of $2n = 2x = 22$ chromosomes, with nucleolar organizing regions (NORs) at the terminal sites of two homologous chromosome pairs (Liu *et al.*, 2006). In our assembly, we detected 23 terminal arrays of telomeric repeats, with the majority (20 out of 23) located at the ends of the 20 largest contigs, supporting the high contiguity of the assembly (Fig. S5). While these features may reflect true chromosomal structures, further cytogenetic validation is required to confirm the precise organization of nucleolar organizer regions-associated regions.

Structural gene annotation identified 72 352 genes, comprising 35 670 HC and 36 682 LC genes (Table 1). The BUSCO completeness assessment yielded 96.6% for HC-only genes and 97.3% for the combined HC + LC dataset (Table 1). These results, together with the recovery of distal telomeric sequences and overall genome statistics, support the generation of a near-chromosome-level high-quality *Robinia* genome with extensive completeness. Furthermore, our analysis revealed that *c.* 55% of the genome consists of TEs, predominantly long terminal repeat (LTR) retrotransposons (Table S2). Notably, many LTR elements were enriched in intergenic regions, which may play a role in gene regulation and genome structure.

A recently published *Robinia* genome (Wang *et al.*, 2023) is highly comparable to ours in terms of genome statistics and sequence-level organization (Fig. S6A,B). The reported genome size was *c.* 693.1 Mb (this study: 645 Mb), with an N50 of 59.9 Mb (58.4 Mb) and a heterozygosity rate of 1.13% (1.34%). The completeness assessment identified 98.2% complete BUSCO genes (this study: 98.8%), and 33 187 protein-coding genes were predicted (this study: 35670) with BUSCO completeness of 97% (this study: 96.6%). In addition to overall similar genome statistics, we observed strong contiguity and synteny between the two assemblies (Fig. S6A,B). However, the Wang *et al.* (2023). genome originates from a Northwest China variety, whereas our assembly represents a Northeast China (DB) genotype, providing a complementary resource for future comparative and evolutionary analyses.

To compare *Robinia* with other temperate plant species, we selected the N_2-fixing legumes *M. truncatula* and *G. max*, along with the non-N_2-fixing tree *P. trichocarpa*. Comparative genome analysis (Fig. 1a) revealed a high degree of chromosomal synteny between *Medicago* and *Robinia*. *Medicago* chromosomes 6 and 7 exhibit synteny with *Robinia* scaffolds 00005 and 00006, respectively. Additionally, *Robinia* scaffolds 00002 and 00003 are syntenic to *Medicago* chromosome 1, whereas *Robinia* scaffolds 00009 and 00011 correspond to *Medicago* chromosome 2. Orthologous gene framework analysis confirmed that most orthologous genes are shared across all investigated species, indicating high conservation of core gene content. Given its strong synteny with *Robinia*, *Medicago* was chosen for further comparative transcriptomic analyses. This close relationship supports the use of *R. pseudoacacia* as a model for studying N-fixing woody legumes in the context of symbiotic interactions and genomic evolution. Furthermore, *G. max* was not selected due to its polyploid genome and distinct domestication history, which may introduce confounding factors in comparative transcriptional analyses.

Organ-specific transcription in *Robinia* upon rhizobia inoculation

To analyze organ-specific expression changes in response to rhizobia inoculation, we generated RNA-seq data from *Robinia* leaves, stems, and roots. Of the 72 352 genes annotated, 27 550 (16 682 HC and 10 868 LC) were expressed (median TPM > 0.5 of three replicates) in at least one tissue. Differential expression analysis between untreated and inoculated plants identified 4705 significant (FDR < 0.05) differentially expressed genes (DEGs): 308 (134 inoculated, 174 control) in roots, 4096 (1125 inoculated, 2971 control) in stems, and 638 (313 inoculated, 325 control) in leaves. The majority of DEGs were found in stems, indicating a strong transcriptional response in this organ, with a pronounced downregulation of stem-specific gene expression under inoculation. This pattern contrasts with *Medicago*, in which inoculation induces 6339 DEGs, predominantly in roots (89.4%), with 8.8% unique to shoots and only 1.8% shared between tissues (Gao *et al.*, 2022). The divergence in DEG distribution suggests fundamental differences between trees and herbaceous legumes in symbiotic N fixation, possibly due to the longer distances required for nutrient exchange in

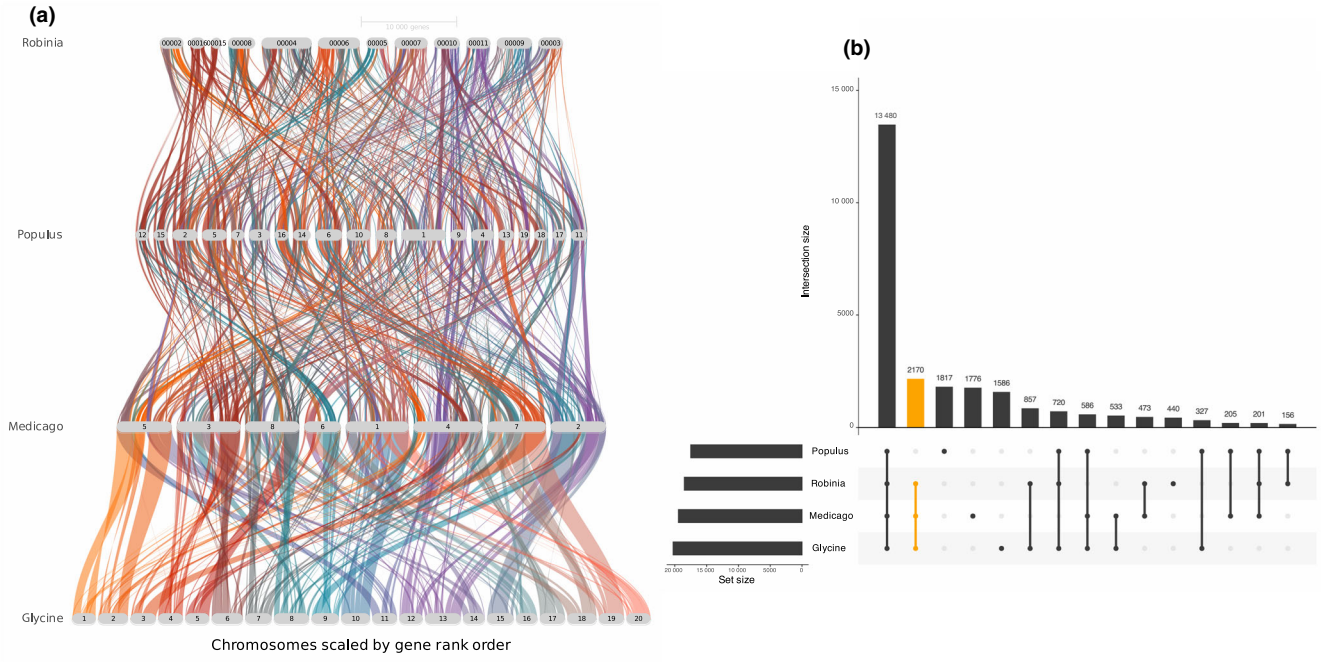

Fig. 1 Comparison between N_2-fixing (*Robinia pseudoacacia*, *Medicago truncatula*, and *Glycine max*) and non-N_2-fixing (*Populus trichocarpa*) species. (a) Chromosome syntenic network across genomes. (b) Upset plot showing group-specific orthologs.

woody species. Notably, 5.2% (213/4096) of stem DEGs in *Robinia* encode transporters, highlighting the importance of solute redistribution between photosynthetic tissues and the root system.

In woody plants and nodulating herbaceous species, key metabolic pathways, such as S assimilation (Kalloniati *et al.*, 2015), N assimilation (Herschbach *et al.*, 2012), and carbon (C) metabolism (Fotelli *et al.*, 2011), are often reprogrammed upon symbiosis, requiring enhanced metabolite transport between roots and shoots. To determine whether the observed shifts in *Robinia* are functionally linked to stem-specific transcriptional responses, we examined the expression patterns of genes involved in S, N, phosphate (P), and C metabolism (Fig. 2a) and compared them with those of *Medicago*. Only significantly regulated DEGs (FDR < 0.05) were considered.

S assimilation involves glutathione (GSH) synthesis and transport, essential for N fixation and oxidative stress protection (Bela *et al.*, 2015). In *Robinia*, five S-transporters were significantly downregulated in stems, with one transporter upregulated in leaves. By contrast, *Medicago* roots exhibited two upregulated and one downregulated S-transporter. Notably, a GSH transporter was upregulated in *Robinia* leaves but was not differentially expressed in *Medicago*. GSH peroxidases, which mitigate oxidative stress during rhizobia colonization (Harrison *et al.*, 2005; Temme & Tudzynski, 2009), were differentially expressed: In *Robinia*, one GSH peroxidase was downregulated in roots and three were regulated in stems (two up, one down), whereas in *Medicago*, three were downregulated in roots. Serine acetyltransferase (SAT), a key enzyme in S assimilation, showed no significant regulation in *Medicago*, whereas one SAT gene was downregulated in *Robinia* stems, suggesting potential differences in S metabolic reallocation.

In many trees, N assimilation is partially shifted from leaves to roots, resembling the metabolic adjustments of herbaceous legumes under BNF (Herschbach *et al.*, 2012). Typically, amino compounds are transported from roots to leaves via stems, with ureides being predominant in N_2-fixing plants and amides in trees. In *Robinia*, genes encoding ureide synthesis enzymes were expressed not only in roots but also in stems, with significantly lower expression in leaves. This suggests a shift from amide-based to ureide-based assimilation upon inoculation. Glutamine synthetase (GS), a key enzyme for ammonium assimilation (Bernard & Habash, 2009), was downregulated in *Robinia* stems (two genes), while in *Medicago*, one GS gene was downregulated in roots. A general trend of transporter gene downregulation was observed in *Robinia*, with amino acid transporters down in leaves, amino acid and molybdate transporters down in roots, and ammonium and molybdate transporters down in stems. By contrast, *Medicago* exhibited downregulation of approximately two-thirds of amino acid transporters in roots, with two transporters upregulated in shoots. Only two molybdate transporters were downregulated in *Medicago* roots, whereas ammonium transporters were upregulated in both roots (one gene) and shoots (two genes). These differences highlight distinct N transport strategies between *Robinia* and *Medicago*.

BNF enhances P availability in the soil, which can increase phosphatase expression in roots. In *Robinia*, phosphatases were significantly upregulated in leaves but downregulated in both stems and roots. Similarly, *Medicago* roots exhibited phosphatase downregulation, while in shoots, expression was balanced (three

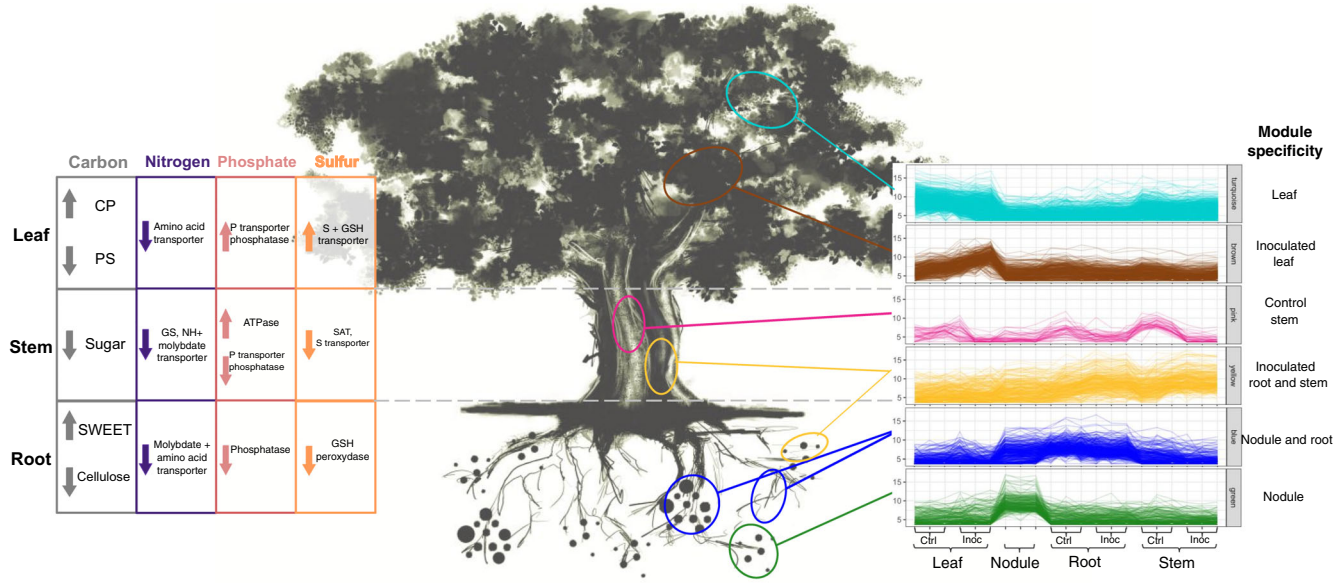

Fig. 2 Transcriptional overview of *Robinia pseudoacacia*. Left scheme: Overview of metabolic changes due to inoculation. Arrow up means inoculation; arrow down means control (R-limma, $P < 0.05$). CP, carboxypeptidase; GS, glutamine synthetase; GSH, glutathione; P, phosphate; PS, photosynthesis; S, sulfate; SAT, serine acetyltransferase; SWEET, bidirectional sugar transporter SWEET. Right scheme: Weighted correlation network analysis modules showing tissue-specific gene expression. The following modules have been identified: blue (specific for nodules and roots), brown (specific for leaves of inoculated plants), green (specific for nodules), pink (specific for stems of control plants), turquoise (specific for leaves), and yellow (specific for roots and stems of inoculated plants). Ctrl, control noninoculation; Inoc, inoculation.

up, three down). In *Robinia*, P-transporters were upregulated in leaves but downregulated in stems, whereas in *Medicago*, they were predominantly upregulated in roots. Additionally, two V-type proton ATPase genes were significantly enriched in *Robinia* stems, while all ATPase genes were downregulated in *Medicago* roots. These findings suggest contrasting phosphate uptake and redistribution strategies in trees vs herbaceous legumes.

To support nitrogenase activity in nodules, additional sugars are produced in leaves via photosynthesis and transported to roots to fuel respiration. In *Robinia*, nearly all DEGs related to photosynthesis were downregulated in leaves, a finding consistent with previous reports (Liu *et al.*, 2024b). A similar pattern was observed in *Medicago*, confirming a general downregulation of photosynthetic activity upon inoculation. Sugar transporters, particularly SWEET proteins, play a critical role in carbon allocation under N fixation (Zhu *et al.*, 2024). In *Robinia*, we identified 27 SWEET transporters, of which eight were significantly differentially expressed. Among these, one was upregulated and one downregulated in leaves, one was up-regulated in roots, and five were regulated in stems (one up, four down). By contrast, in *Medicago*, SWEET transporters were only differentially expressed in roots, where four out of six were downregulated. The isolated upregulation of a single SWEET transporter in *Robinia* roots may reflect a localized adjustment in C allocation to nodules, rather than a broad enhancement of sugar transport.

Together, these findings highlight major metabolic reprogramming in *Robinia* during symbiosis, with an unexpectedly strong transcriptional response in stems. This suggests that trees exhibit a distinct mode of metabolic integration compared with herbaceous legumes, potentially involving stem-mediated coordination of resources. However, rather than an increased reliance on active solute transport through the stem, our data indicate a predominant downregulation of key transporter families, including P, S, N, and sugar transporters. This contrasts with *Medicago*, in which transporter regulation is primarily observed in roots. These findings reflect the structural and physiological constraints of woody N-fixing plants.

Co-expression modules are correlated with tissues and inoculation

Using WGCNA of significant DEGs, we identified two major clusters of highly co-expressed genes, each divided into three modules significantly correlated with specific tissues and inoculation status (Figs 2b, S7). The first cluster contained the blue (nodules and roots), green (nodules), and yellow (roots and stems of inoculated plants) modules, while the second cluster included the brown (leaves of inoculated plants), turquoise (leaves in general), and pink (stem under control conditions) modules. These modules represent distinct transcriptional programs associated with symbiotic N fixation, nutrient transport, and stress adaptation in *Robinia*.

Nodule-specific module (green, 226 genes) The green module contained genes predominantly expressed in nodules, including several copies of leghemoglobin, a key protein that facilitates BNF by maintaining a low-oxygen environment for nitrogenase activity (Ott *et al.*, 2005). Additionally, this module was enriched

in plant cadmium resistance (PCR) genes, which enhance tolerance to cadmium (Cd) accumulation, a major pollutant in agricultural soils of China (Lin *et al.*, 2020). The upregulation of Cd resistance genes in nodules suggests that *Robinia* may employ a protective strategy to mitigate heavy metal toxicity during symbiosis. Other highly abundant genes in this module include transporters for sugar, phosphate, sulfate, and copper, supporting the idea that nodules function as active hubs for metabolite exchange with the plant host.

Nodule and root-specific module (blue, 354 genes) The blue module, specific to both nodules and roots, was enriched for multiple copies of the Bowman–Birk type trypsin inhibitors, which are known to regulate endogenous proteases and provide protection against microbial and insect proteases (Gitlin-Domagalska *et al.*, 2020). These inhibitors are commonly found in legumes and cereals, playing a crucial role in plant defense mechanisms. Their enrichment in *Robinia* roots and nodules suggests a dual role in protecting against both symbiotic and pathogenic microbes (Gitlin-Domagalska *et al.*, 2020). The presence of root-specific transporters in this module indicates active regulation of nutrient acquisition and exchange between roots and nodules.

Roots and stems of inoculated plants (yellow, 250 genes) The yellow module was associated with roots and stems in rhizobia-inoculated plants, containing several heat shock proteins, lipid transfer proteins, defensins, glucose phosphatases, glutathione peroxidases and S-transferases, homeobox proteins, lectins, and pathogenesis-related proteins. Many of these genes play roles in stress adaptation and immune responses, suggesting that BNF induces systemic molecular changes beyond root tissues. The enrichment of glutathione peroxidases and S-transferases also indicates an increased demand for oxidative stress mitigation in stems and roots upon inoculation (Madhu Sharma *et al.*, 2022).

Leaves of inoculated plants (brown, 353 genes) The brown module was highly specific to the leaves of inoculated plants, showing strong expression of solute transporters, including several sugar and inositol transporters, likely reflecting increased energy demand during N fixation. Additionally, genes encoding amino acid permeases and proline transporters were highly expressed, supporting long-distance N transport from roots to leaves. The presence of magnesium, zinc, phosphate, sulfate, and calcium transporters suggests an extensive metabolic shift in leaf nutrient homeostasis under symbiosis. Although ureide degradation in leaves is typically rapid, we identified a highly expressed ureide permease-like protein, indicating possible differences in N remobilization strategies between *Robinia* and herbaceous legumes.

General leaf-specific module (turquoise, 765 genes) As expected, the turquoise module was dominated by photosynthesis-related genes, reinforcing the role of leaves as primary carbon sources. Additionally, multiple solute transporters were highly expressed, facilitating the movement of amino acids, sulfate, potassium, inositol, sugars, phosphate, sulfite, and zinc. This highlights the complex coordination between carbon and N metabolism under symbiosis, ensuring efficient resource allocation between leaves and belowground tissues.

Stems under control conditions (pink, 40 genes) The pink module, the smallest of all identified co-expression modules, was specific to stems of noninoculated plants. This module was highly enriched in cell wall-related genes, suggesting that uninoculated stems maintain active cell wall remodeling processes. Additionally, we identified the expression of plant Cd resistance proteins and detoxification transporter proteins, which may indicate a basal detoxification strategy independent of BNF. The small size of this module suggests that stems under control conditions exhibit lower metabolic activity than inoculated stems, in which symbiosis-related metabolic adjustments are necessary.

These WGCNA-derived co-expression modules provide insight into the organ-specific molecular responses of *Robinia* to rhizobia inoculation. The green and blue modules highlight nodule-specific gene expression patterns, including the activation of leghemoglobins, Cd resistance genes, and protease inhibitors. The yellow and brown modules emphasize metabolic and transport-related shifts in stems and leaves, reinforcing the systemic impact of symbiosis. Finally, the pink module suggests that uninoculated stems exhibit lower transcriptional activity, which may shift upon BNF activation. Together, these findings suggest that *Robinia* coordinates a highly specialized molecular network to balance nutrient exchange, stress adaptation, and nodule function, distinct from herbaceous legumes.

Nodulation-related differences between *Robinia* and *Medicago*

To better understand the regulatory mechanisms underlying nodulation in *R. pseudoacacia*, we conducted a comparative analysis with *M. truncatula*, an extensively studied model legume. Despite their divergence *c.* 41.31 million years ago (Ma) (Lebvre *et al.*, 2010; Zhao *et al.*, 2021), both species retain key features of N-fixing symbiosis, making *Medicago* a relevant comparative framework. Notably, *Robinia* has not undergone a whole genome duplication (WGD), facilitating a clearer one-to-one gene comparison.

To establish a direct comparison, we curated a *Medicago* RNA-seq dataset comprising 59 publicly available samples from NCBI SRA and integrated them with our *Robinia* RNA-seq dataset. Co-expression networks were constructed separately for each species, yielding 38 modules for *Medicago* and 23 for *Robinia*. Among the 72 352 genes in *Robinia* and 47 899 in *Medicago*, 27 406 and 28 972 genes, respectively, were identified as expressing TPM > 0.5. We observed strong positive correlations ($P \geq 0.99$) between network modules and specific tissues, confirming biologically meaningful gene clusters (Fig. 3a). Notably, the modules Med1 and Rob1 correlated with shoot and leaf tissues, Med2 and Rob4 with nodules, and Med4 and Rob5 with root tissues. Additionally, the Med1 and Rob1 modules were enriched in photosynthesis-related genes, reinforcing their

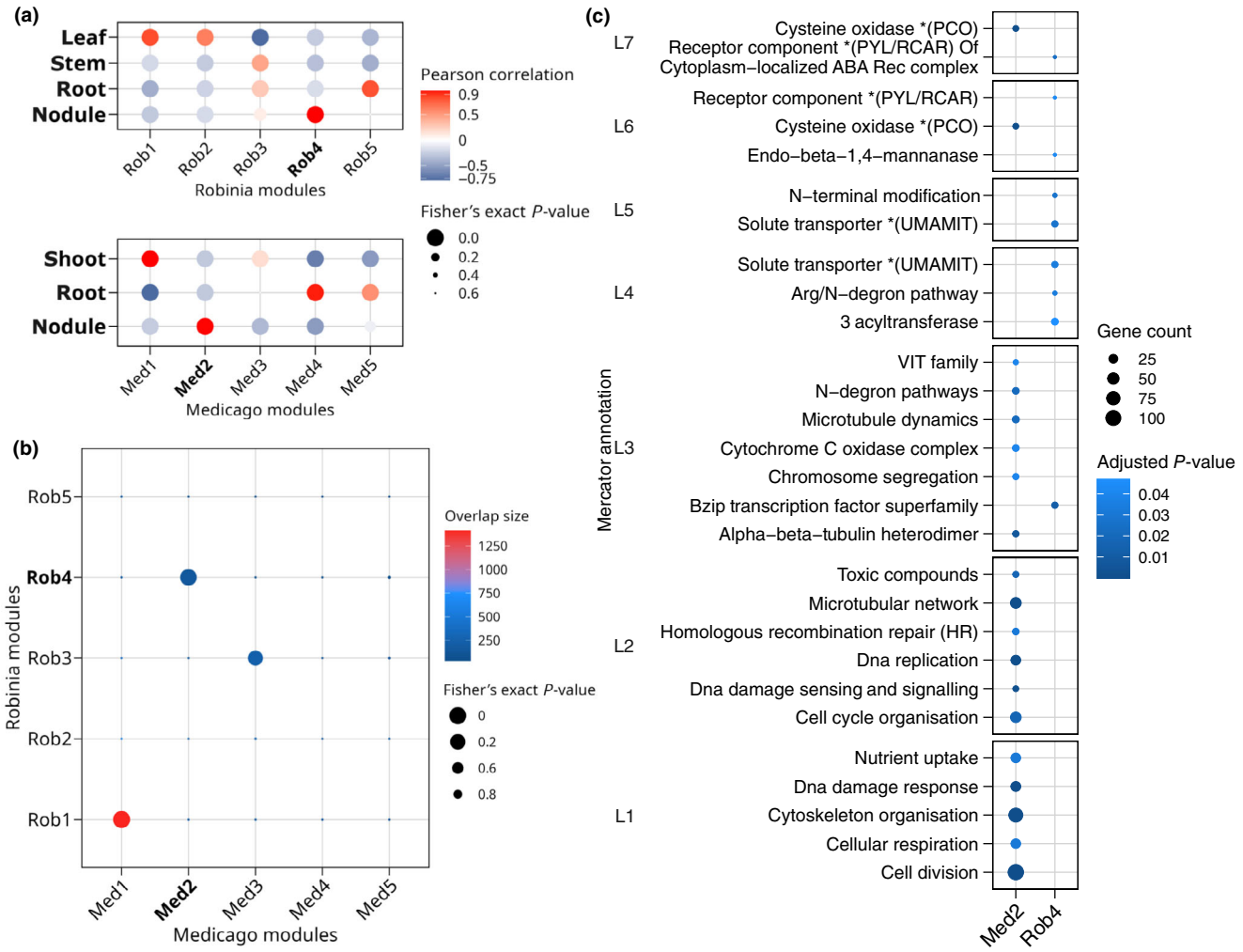

Fig. 3 Nodulation-related differences between *Robinia pseudoacacia* and *Medicago truncatula*. (a) Module–tissue correlation of *Robinia* and *Medicago* top five co-expression network modules. (b) Co-expression network module comparison by their shared number of orthologs for the five biggest modules of both species. (c) Enrichment of Mercator annotations for nodulation modules. Mercator annotation terms are sorted into Level 1–7 categories based on their BIN positions and labeled as L1–L7.

functional association with leaf metabolism (Fig. S8C). These cross-species correlations confirm the reliability of the module–trait associations and indicate conserved transcriptional networks.

Using OrthoFinder, we identified 9431 shared orthogroups expressed in both species, allowing a direct comparison of co-expression modules between *Robinia* and *Medicago*. A significant overlap of 143 shared orthogroups (out of 2296 and 524 orthogroups in Med2 and Rob4, respectively) suggests that key nodulation regulatory networks are conserved between both species (Fig. 3b). However, functional enrichment analysis revealed substantial differences (Figs 3c, S9–S11). Molybdate ion transport was exclusively enriched in *Medicago*, consistent with its role in enhancing nodulation efficiency (Zhou *et al.*, 2017; Sary *et al.*, 2020). The absence of this pathway in *Robinia* suggests either a lower dependency on molybdate for nitrogenase activity or an alternative regulatory mechanism. Amino acid transport and oxygen binding were commonly enriched in both species, supporting their shared metabolic adaptations for N fixation. By contrast, hormone-mediated signaling pathways and hormone binding were uniquely enriched in *Robinia*, implicating plant hormone regulation in its nodulation strategy. Specifically, genes encoding abscisic (ABA) receptors and basic (region) leucine zipper (BZIP) transcription factors were exclusive to *Robinia*, suggesting a distinct role for ABA in woody legume nodulation. ABA has been previously implicated in the modulation of root architecture and nodulation in legumes (Liu *et al.*, 2022), supporting its potential role in *Robinia*'s symbiotic adaptation. Given that ABA signaling influences root development and stress responses, its involvement in *Robinia* nodulation may indicate an alternative regulatory mechanism compared with herbaceous legumes.

These findings highlight both conserved and species-specific regulatory mechanisms governing nodulation in *Robinia* and *Medicago*. While major metabolic pathways related to amino acid transport and oxygen binding appear shared, the absence of molybdate enrichment and the presence of ABA-specific

regulation in *Robinia* suggest key divergences. This comparative analysis provides a molecular framework for understanding nodulation in woody legumes, paving the way for future research into the evolutionary diversification of N-fixing symbioses.

Discussion

In this study, we generated a high-quality reference genome for *R. pseudoacacia* and analyzed its transcriptional responses to nodulation. The assembled genome size of 699.6 Mb aligns well with previous *C*-value estimates ranging from 684 to 734 Mb (Olszewska & Osiecka, 1984; Bai *et al.*, 2012). Comparative genomic analysis with two herbaceous legumes (*M. truncatula* and *G. max*) and one non-N_2-fixing tree (*P. trichocarpa*) revealed a high degree of synteny between *Robinia* and *Medicago*, consistent with their divergence *c.* 41 Ma (Lebvre *et al.*, 2010; Zhao *et al.*, 2021). The conservation of large orthologous gene sets across all species suggests that key genetic components required for symbiosis have been maintained.

A key finding from our transcriptomic analysis is the unexpectedly high proportion of DEGs in stems upon rhizobia inoculation. Unlike *Medicago* (Gao *et al.*, 2022), in which 90% of DEGs are found in roots following inoculation, nearly 90% of DEGs in *Robinia* were located in stems. This suggests a fundamental difference in how nodulation signals are regulated in woody vs herbaceous legumes. The observed downregulation of transcription in stems could reflect early-stage systemic reprogramming, prioritizing root-associated N assimilation over shoot growth (Ruffel *et al.*, 2008). Alternatively, it may indicate a delayed or prolonged metabolic adjustment due to the longer transport distances in trees, requiring stems to coordinate sugar, N, and metabolite redistribution more actively than in herbaceous species (Gessler *et al.*, 2004; Luo *et al.*, 2020).

To further investigate these organ-specific transcriptional patterns, we conducted WGCNA, which identified six major co-expression modules strongly correlated with tissue and inoculation status. These modules provide insights into how different plant organs contribute to N fixation and metabolic regulation. The nodule-specific module contained multiple copies of leghemoglobin – a key oxygen-binding protein essential for nitrogenase function (Ott *et al.*, 2005)—as well as PCR genes, which may protect nodules from metal toxicity in contaminated soils (Lin *et al.*, 2020). Upregulation of leghemoglobin and Remorin genes was previously described upon rhizobia inoculation (Liang *et al.*, 2018; Welch *et al.*, 2020). These findings suggest that *Robinia* has evolved mechanisms to detoxify heavy metals in nodules while maintaining symbiotic efficiency (Fan *et al.*, 2018). The stem-specific modules were enriched in transporters involved in amino acid, sulfate, and sugar transport, supporting the hypothesis that stems act as metabolic relay stations, coordinating the long-distance exchange of N and carbon metabolites between leaves and roots (De Schepper & Steppe, 2010).

Comparison between *Robinia* and *Medicago* transcriptomes revealed both conserved and species-specific nodulation responses. Unlike *M. truncatula* (Young *et al.*, 2011) and *G. max* (Schmutz *et al.*, 2010), which underwent additional WGDs after the legume-wide WGD, *R. pseudoacacia* has not experienced a more recent lineage-specific WGD (Wang *et al.*, 2023). This distinction suggests that while *Robinia* shares ancestral duplication events with other legumes, its genome has remained more stable in terms of large-scale duplications. Co-expression module analysis identified significant correlations between nodulation-associated gene clusters, particularly in the Med2 and Rob4 modules, which were strongly associated with nodule formation (Fig. 3b). However, molybdate transport genes, which are known to enhance rhizobia symbiosis in *Medicago* (Zhou *et al.*, 2017; Sary *et al.*, 2020), were absent from *Robinia*'s nodule-specific module. This suggests that *Robinia* may rely on alternative mechanisms to optimize nitrogenase activity, possibly through distinct metal cofactor homeostasis pathways (González-Guerrero *et al.*, 2014). Further investigations using proteomics and ionomics profiling will be necessary to determine whether *Robinia* employs alternative molybdate uptake pathways or post-transcriptional regulatory mechanisms for nitrogenase activation. Another key difference is the strong enrichment of ABA-related signaling in *Robinia*, as evidenced by the presence of ABA receptor genes and BZIP transcription factors in nodule-associated modules. This suggests that woody legumes may integrate hormonal signaling pathways into their nodulation responses in a way that differs from herbaceous legumes (Guinel, 2015). Future studies should investigate whether ABA plays a direct regulatory role in *Robinia* nodulation, potentially linking it to stress responses and resource allocation in long-lived trees (Ding *et al.*, 2008; Tominaga *et al.*, 2009).

Finally, our findings have implications for ecosystem function and sustainable forestry. *R. pseudoacacia* is a globally distributed tree species that thrives in nutrient-poor and depleted, degraded, and contaminated soils, making it an ideal candidate for reforestation and soil restoration programs, as our previous studies already demonstrated (Hu *et al.*, 2017; Liu *et al.*, 2020, 2024a, 2024b). The observed stem-centered transcriptional reprogramming and the presence of metal detoxification genes in nodules suggest that *Robinia* has evolved unique adaptations to symbiotic N fixation under variable environmental conditions (Liu *et al.*, 2020, 2024b; Yao *et al.*, 2022).

In conclusion, this study provides the first comprehensive genomic and transcriptomic resource for a N-fixing tree, uncovering fundamental differences in nodulation regulation between trees and herbaceous legumes. The unexpected stem-specific transcriptional responses, the absence of molybdate-dependent nodulation regulation, and the strong involvement of ABA signaling highlight *Robinia*'s distinct symbiotic strategy. Future work should aim to functionally characterize the identified regulatory pathways and determine their role in optimizing N fixation efficiency in woody legumes, which could contribute to developing improved reforestation, phytoremediation, and sustainable forestry strategies.

Acknowledgements

The authors thank their colleagues Julia Vrebalov from Cornell CALS (www.cals.cornell.edu/), Nir Kfir, Yoni Roter, and Evan Staton from NRgene (www.nrgene.com) and Biomarker

Technologies Co., Beijing, China (http://www.biomarker.com.cn) for their supportive work on the DNA and RNA levels. The authors also thank Verena Messerer for editing Fig. 2. This work was supported by funding from the 'Double-First Class' Initiative Program for Foreign Talents of Southwest University and the 'Prominent Scientist Program' of Chongqing Talents (cstc2021ycjh-bgzxm0002 for BH & cstc2021ycjh-bgzxm0020 for HR). MM was funded by the German Research Foundation (DFG) – TRR 356.

Competing Interests

None declared.

Author contributions

BH, MM, KFXM, RH and HR were involved in conceptualization, methodology, formal analysis, visualization, and writing – original draft. KDO were involved in writing, editing, revising manuscript, RNA-seq analysis and genomic analysis. VM, GH, VM, MM and TL were involved in methodology, formal analysis and statistical analysis. DK, JS, L-SK, JJ and RG were involved in formal analysis. KDO, RH and HR were involved in writing – review and editing. BH, RH and HR were involved in resources, conceptualization, design of the experiment, validation, funding acquisition, and writing – review and editing. All authors approved the final version of the manuscript. BH and MM contributed equally to this work.

ORCID

Robert Hänsch https://orcid.org/0000-0003-1310-8419
Georg Haberer https://orcid.org/0000-0002-6612-6939
Bin Hu https://orcid.org/0000-0001-9814-5096
David Kaufholdt https://orcid.org/0000-0001-7509-2833
Thomas Lux https://orcid.org/0000-0002-5543-1911
Vanda Marosi https://orcid.org/0000-0002-1444-2972
Klaus F. X. Mayer https://orcid.org/0000-0001-6484-1077
Maxim Messerer https://orcid.org/0000-0003-0554-5045
Kevin D. Oliphant https://orcid.org/0000-0003-2757-8315
Heinz Rennenberg https://orcid.org/0000-0001-6224-2927

Data availability

RNA-seq data were submitted to EMBL-EBI-Annotare (https://www.ebi.ac.uk/fg/annotare) under the ArrayExpress accession no. E-MTAB-14236. Assembly and annotation were submitted to zenodo (DOI: 10.5281/zenodo.14887937).

References

Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**: 1600–1607.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Bai C, Alverson WS, Follansbee A, Waller DM. 2012. New reports of nuclear DNA content for 407 vascular plant taxa from the United States. *Annals of Botany* **110**: 1623–1629.

Batterman SA, Hedin LO, Breugel M, Ransijn J, Craven DJ, Hall JS. 2013. Key role of symbiotic dinitrogen fixation in tropical forest secondary succession. *Nature* **502**: 224–227.

Bela K, Horváth E, Gallé Á, Szabados L, Tari I, Csiszár J. 2015. Plant glutathione peroxidases: emerging role of the antioxidant enzymes in plant development and stress responses. *Journal of Plant Physiology* **176**: 192–201.

Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Udvardi MK. 2008. A gene expression atlas of the model legume *Medicago truncatula*. *The Plant Journal* **55**: 504–513.

Bernard SM, Habash DZ. 2009. The importance of cytosolic glutamine synthetase in nitrogen assimilation and recycling. *New Phytologist* **1823**: 608–620.

Bi Y, Zou H, Zhu C. 2014. Dynamic monitoring of soil bulk density and infiltration rate during coal mining in sandy land with different vegetation. *International Journal of Coal Science & Technology* **1**: 198–206.

Cao Y, Chen Y. 2017. Coupling of plant and soil C : N : P stoichiometry in black locust (*Robinia pseudoacacia*) plantations on the Loess Plateau, China. *Trees* **31**: 1559–1570.

Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**: e107.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**: 170–175.

Colebatch G, Desbrosses G, Ott T, Krusell L, Montanari O, Kloska S, Kopka J, Udvardi MK. 2004. Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. *The Plant Journal* **39**: 487–512.

De Schepper V, Steppe K. 2010. Development and verification of a water and sugar transport model using measured stem diameter variations. *Journal of Experimental Botany* **618**: 2083–2099.

Ding Y, Kaló P, Yendrek CR, Sun J, Liang Y, Marsh JF, Harris JM, Oldroyd GE. 2008. Abscisic acid coordinates Nod factor and cytokinin signaling during the regulation of nodulation in *Medicago truncatula*. *Plant Cell* **20**: 2681–2695.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Duan X, Xie Y, Liu G, Gao X, Lu H. 2010. Field capacity in black soil region, northeast China. *Chinese Geographical Science* **20**: 406–413.

El Sabagh A, Hossain A, Islam MS, Fahad S, Ratnasekera D, Meena RS, Hasanuzzaman M. 2020. Nitrogen fixation of legumes under the family Fabaceae: adverse effect of abiotic stresses and mitigation strategies. In: Hasanuzzaman M, Araújo S, Gill S, eds. *The plant family Fabaceae: biology and physiological responses to environmental stresses*. Singapore: Springer, 75–111.

Elser JJ, Bracken MES, Gruner DS, Harpole WS, Hillebrand H, Smith JE. 2007. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecology Letters* **10**: 1135–1142.

Fan M, Xiao X, Guo Y, Zhang J, Wang ET, Chen W, Lin Y, Wei G. 2018. Enhanced phytoremdiation of Robinia pseudoacacia in heavy metal-contaminated soils with rhizobia and the associated bacterial community structure and function. *Chemosphere* **197**: 729–740.

Fang Y, Hirsch AM. 1998. Studying early nodulin gene ENOD40 expression and induction by nodulation factor and cytokinin in transgenic alfalfa. *Plant Physiology* **116**: 53–68.

Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE, Gresshoff PM. 2010. Molecular analysis of legume nodule development and autoregulation. *Journal of Integrated Plant Biology* **52**: 61–76.

Ferguson BJ, Mens C, Hastwell AH, Zhang M, Su H, Jones CH, Chu X, Gresshoff PM. 2019. Legume nodulation: the host controls the party. *Plant, Cell & Environment* **42**: 41–51.

Flemetakis E, Efrose RC, Ott T, Stedel C, Aivalakis G, Udvardi MK, Katinakis P. 2006. Spatial and temporal organization of sucrose metabolism in *Lotus*

japonicus nitrogen-fixing nodules suggests a role for the elusive alkaline/neutral invertase. *Plant Molecular Biology* **62**: 53–69.

Fotelli MN, Tsikou D, Kolliopoulou A, Aivalakis G, Katinakis P, Udvardi MK, Rennenberg H, Flemetakis E. 2011. Nodulation enhances dark CO_2 fixation and recycling in the model legume *Lotus japonicus*. *Journal of Experimental Botany* **62**: 2959–2971.

Franzini VI, Azcón R, Méndes FL, Aroca R. 2013. Different interaction among Glomus and Rhizobium species on *Phaseolus vulgaris* and *Zea mays* plant growth, physiology and symbiotic development under moderate drought stress conditions. *Plant Growth Regulation* **70**: 265–273.

Galloway JN, Dentener FJ, Capone DG, Boyer EW, Howarth RW, Seitzinger SP, Vöosmarty CJ. 2004. Nitrogen cycles: past, present and future. *Biogeochemistry* **70**: 153–226.

Gani U, Vishwakarma RA, Misra P. 2021. Membrane transporters: the key drivers of transport of secondary metabolites in plants. *Plant Cell Reports* **40**: 1–18.

Gao Y, Selee B, Schnabel EL, Poehlman WL, Chavan SA, Frugoli JA, Feltus FA. 2022. Time series transcriptome analysis in *Medicago truncatula* shoot and root tissue during early nodulation. *Frontiers in Plant Science* **13**: 861639.

Gessler A, Kopriva S, Rennenberg H. 2004. Regulation of nitrate uptake at the whole-tree level: interaction between nitrogen compounds, cytokinins and carbon metabolism. *Tree Physiology* **24**: 1313–1321.

Ghosh S, Chan CKK. 2016. Analysis of RNA-Seq data using TopHat and Cufflinks. *Plant Bioinformatics: Methods and Protocols* **1374**: 339–361.

Gitlin-Domagalska A, Maciejewska A, Dębowski D. 2020. Bowman-Birk inhibitors: insights into family of multifunctional proteins and peptides with potential therapeutical applications. *Pharmaceuticals* **13**: 421.

González-Guerrero M, Matthiadis A, Sáezá LTA. 2014. Fixating on metals: new insights into the role of metals in nodulation and symbiotic nitrogen fixation. *Frontiers in Plant Science* **5**: 45.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Regev A. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* **29**: 644.

Gremme G, Brendel V, Sparks ME, Kurtz S. 2005. Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology* **47**: 965–978.

Guinel FC. 2015. Ethylene, a hormone at the center-stage of nodulation. *Frontiers in Plant Science* **6**: 1121.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, Wortman JR. 2008. Automated eukaryotic gene structure annotation using evidence modeler and the program to assemble spliced alignments. *Genome Biology* **9**: 1–22.

Harrison J, Jamet A, Muglia CI, Van de Sype G, Aguilar OM, Puppo A, Frendo P. 2005. Glutathione plays a fundamental role in growth and symbiotic capacity of *Sinorhizobium meliloti*. *Journal of Bacteriology* **187**: 168–174.

Heju H. 2005. Spring nursery soil disinfection. *Northern Horticulture* **45**: 35–43.

Herschbach C, Gessler A, Rennenberg H. 2012. Long-distance transport and plant internal cycling of N- and S-compounds. In: Lüttge U, Beyschlag W, Büdel B, Francis D, eds. *Progress in botany 73. Progress in botany*, vol. **73**. Berlin, Heidelberg, Germany: Springer.

Hoff KJ, Stanke M. 2019. Predicting genes in single genomes with Augustus. *Current Protocols in Bioinformatics* **65**: e57.

Hu B, Zhou M, Dannenmann M, Saiz G, Simon J, Bilela S, Liu X, Hou L, Chen H, Zhang S *et al.* 2017. Comparison of nitrogen nutrition and soil carbon status of afforested stands established in degraded soil of the Loess Plateau, China. *Forest Ecology and Management* **389**: 46–58.

Kalloniati C, Krompas P, Karalias G, Udvardi MK, Rennenberg H, Herschbach C, Flemetakis E. 2015. Nitrogen-fixing nodules are an important source of reduced sulfur, which triggers global changes in sulfur metabolism in *Lotus japonicus*. *Plant Cell* **27**: 2384–2400.

Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology* **20**: 1–13.

Küster H, Hohnjec N, Krajinski F, El Yahyaoui F, Manthey K, Gouzy J, Dondrup M, Meyer F, Kalinowski J, Brechenmacher L *et al.* 2004. Construction and validation of cDNA-based Mt6k-RIT macro- and microarrays to explore root endosymbioses in the model legume *Medicago truncatula*. *Journal of Biotechnology* **108**: 95–113.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**: 559.

Langfelder P, Horvath S. 2012. Fast R functions for robust correlations and hierarchical clustering. *Journal of Statistical Software* **46**: 1–17.

LeBauer DS, Treseder KK. 2008. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology* **89**: 371–379.

Lebvre B, Timmers T, Mbengue M, Moreau S, Hervé C, Tóth K, Bittencourt-Silvestre J, Klaus D, Deslandes L, Godiard L *et al.* 2010. A remorin protein interacts with symbiotic receptors and regulates bacterial infection. *Proceedings of Natural Academy of Science, USA* **107**: 2343–2348.

Li G, Xu G, Guo K, Du S. 2014. Mapping the global potential geographical distribution of black locust (*Robinia pseudoacacia* L.) using herbarium data and a maximum entropy model. *Forests* **5**: 2773–2792.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.

Liang P, Stratil TF, Popp C, Marín M, Folgmann J, Mysore KS, Wen J, Ott T. 2018. Symbiotic root infections in *Medicago truncatula* require remorin-mediated receptor stabilization in membrane nanodomains. *Proceedings of National Academy of Science, USA* **115**: 5289–5294.

Liesche J, Pace MR, Xu Q, Li Y, Chen S. 2017. Height-related scaling of phloem anatomy and the evolution of sieve element end wall types in woody plants. *New Phytologist* **214**: 245–256.

Lin J, Gao X, Zhao J, Zhang J, Chen S, Lu L. 2020. *Plant Cadmium Resistance 2* (*SaPCR2*) facilitates cadmium efflux in the roots of hyperaccumulator *Sedum alfredii* Hance. *Frontiers in Plant Science* **11**: 568887.

Lin Y, Feng Z, Wu W, Yang Y, Zhou Y, Xu C. 2017. Potential impacts of climate change and adaptation on maize in northeast China. *Agronomy Journal* **109**: 1476–1490.

Liu B, Chen C, Li X, Qi L, Han S. 2006. Karyotype analysis and physical mapping of 45S rDNA in eight species of Sophora, Robinia, and Amorpha. *Frontiers of Biology in China* **3**: 290–294.

Liu R, Hu B, Dannenmann M, Giesemann A, Geilfus CM, Li C, Rennenberg H. 2024b. Significance of phosphorus deficiency for the mitigation of mercury toxicity in the *Robinia pseudoacacia* L. – rhizobia symbiotic association. *Journal of Hazardous Materials* **467**: 133717.

Liu R, Hu B, Flemetakis E, Dannenmann M, Geilfus C, Haensch R, Wang D, Rennenberg H. 2024a. Antagonistic effect of mercury and excess nitrogen exposure reveals provenance-specific phytoremediation potential of black locust-rhizobia symbiosis. *Environmental Pollution* **342**: 123050.

Liu X, Fan Y, Long J, Wei R, Kjelgren R, Gong C, Zhao J. 2013. Effects of soil water and nitrogen availability on photosynthesis and water use efficiency of *Robinia pseudoacacia* seedlings. *Journal of Environmental Sciences* **25**: 585–595.

Liu X, Hu B, Chu C. 2022. Nitrogen assimilation in plants: current status and future prospects. *Journal of Genetics and Genomics* **49**: 394–404.

Liu ZS, Chen W, Jiao S, Wang X, Fan M, Wang E, Wei G. 2019. New insight into the evolution of symbiotic genes in black locust-associated rhizobia. *Genome Biology and Evolution* **11**: 1736–1750.

Liu ZS, Hu B, Bell TL, Flemetakis E, Rennenberg H. 2020. Significance of mycorrhizal associations for the performance of N_2-fixing Black Locust (*Robinia pseudoacacia* L.). *Soil Biology & Biochemistry* **145**: 107776.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550.

Love MI, Soneson C, Patro R. 2018. Swimming downstream: statistical analysis of differential transcript usage following Salmon quantification. *F1000Research* **7**: 952.

Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J. 2022. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife* **11**: e78526.

Luo Y, Liu D, Jiao S, Liu S, Wang X, Shen X, Wei G. 2020. Identification of *Robinia pseudoacacia* target proteins responsive to *Mesorhizobium amphore* CCNWGS0123 effector protein NopT. *Journal of Experimental Botany* **71**: 7347–7363.

Madhu Sharma A, Kaur A, Tyagi S, Upadhyay SK. 2022. Glutathione peroxidases in plants: innumerable role in abiotic stress tolerance and plant development. *Journal of Plant Growth Regulation* **42**: 598–613.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**: 764–770.

Mariangela NF, Daniela T, Anna K, Georgios A, Panagiotis K, Michael KU, Rennenberg H, Emmanouil F. 2011. Nodulation enhances dark CO_2 fixation and recycling in the model legume *Lotus japonicus*. *Journal of Experimental Botany* **62**: 2959–2971.

Martin FM, Uroz S, Barker DG. 2017. Ancestral alliances: plant mutualistic symbioses with fungi and bacteria. *Science* **356**: eaad4501.

Modi A, Vai S, Caramelli D, Lari M. 2021. The illumina sequencing protocol and the NovaSeq 6000 system. *Methods in Molecular Biology* **2242**: 15–42.

Mohammad NS, Nazli R, Zafar H, Fatima S. 2022. Effects of lipid based multiple micronutrients supplement on the birth outcome of underweight pre-eclamptic women: a randomized clinical trial. *Pakistan Journal of Medical Sciences* **38**: 219–226.

Murashige T, Skoog FK. 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum* **15**: 473–497.

Nentwig W, Bacher S, Kumschick S, Pyšek P, Vilà M. 2018. More than "100 worst" alien species in Europe. *Biological Invasions* **20**: 1611–1621.

Oldroyd GED, Downie JA. 2006. Nuclear calcium changes at the core of symbiosis signalling. *Current Opinion in Plant Biology* **9**: 351–357.

Olszewska MJ, Osiecka R. 1984. The relationship between 2 C DNA content, systematic position, and the level of nuclear DNA endoreplication during differentiation of root parenchyma in some dicotyledonous shrubs and trees. Comparison with herbaceous species. *Biochemie und Physiologie der Pflanzen* **179**: 641–657.

Ott T, van Dongen JT, Günther C, Krusell L, Desbrosses G, Vigeolas H, Bock V, Czechowski T, Geigenberger P, Udvardi MK. 2005. Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Current Biology* **15**: 531–535.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**: 417–419.

Quilbé J, Lamy L, Brottier L, Leleux P, Fardoux J, Rivallan R, Benichou T, Guyonnet R, Becana M, Villar I *et al.* 2021. Genetics of nodulation in *Aeschynomene evenia* uncovers mechanisms of the rhizobium-legume symbiosis. *Nature Communications* **12**: 829.

Ramangouda G, Naik MK, Nitnavare R, Yeshvekar R, Bhattacharya JK, Bhatnagar-Mathur PK, Sharma MK. 2023. Genetic enhancement of trichoderma asperellum biocontrol potentials and carbendazim tolerance for chickpea dry root rot disease management. *PLoS ONE* **18**: e280064.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* **11**: 1432.

Rennenberg H, Herschbach C, Haberer K, Kopriva S. 2007. Sulfur metabolism in plants: are trees different? *Plant Biology* **9**: 620–637.

Richardson DM, Rejmanek M. 2011. Trees and shrubs as invasive alien species – a global review. *Diversity and Distributions* **17**: 788–809.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**: e47.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Ruffel S, Freixes S, Balzergue S, Tillard P, Jeudy C, Martin-Magniette M, van der Merwe MJ, Kakar K, Gouzy J, Fernie AR *et al.* 2008. Systemic signaling of the plant nitrogen status triggers specific transcriptome responses depending on the nitrogen source in *Medicago truncatula*. *Plant Physiology* **146**: 2020–2035.

Sary D, Niel E, Abdelhameid N. 2020. Effect of inoculation with biofertilizers and molybdenum on the growth and productivity of Faba Bean (*Vicia faba* L.) grown in sandy soil under drip irrigation. *Alexandria Science Exchange Journal* **41**: 105–121.

Savage JA, Clearwater MJ, Haines DF, Klein T, Mencuccini M, Sevanto S, Turgeon R, Zhang C. 2016. Allocation, stress tolerance and carbon transport in plants: how does phloem physiology affect plant ecology. *Plant, Cell & Environment* **4**: 709–725.

Schauser L, Roussis A, Stiller J, Jens S. 1999. A plant regulator controlling development of symbiotic root nodules. *Nature* **402**: 191–195.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson WC, Hyten DL, Song Q, Thelen JJ, Cheng J *et al.* 2010. Genome sequence of the palaeopolyploid soybean. *Nature* **465**: 120.

Schwacke R, Ponce-Soto GY, Krause K, Bolger AM, Arsova B, Hallab A, Usadel B. 2019. MapMan4: a refined protein classification and annotation framework applicable to multi-omics data analysis. *Molecular Plant* **12**: 879–892.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.

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

Sun H, Liu R, Yuan H, Zhou M, Liu Z, Hu B, Rennenberg H. 2023. Interaction of nitrogen availability in the soil with leaf physiological traits and nodule formation of *Robinia pseudoacacia*-rhizobia symbiosis depends on provenance. *Plant and Soil* **490**: 239–259.

Sun X, Yu T, Bin M, Hu C, Bi F, Peng X, Yi G, Zhang X. 2023. Transcriptomic analysis reveals the defense mechanisms of citrus infested with *Diaphorina citri*. *Horticultural Plant Journal* **9**: 450–462.

Tate RL. 2020. Nitrogen fixation. In: Tate RL, ed. *Soil microbiology*. New York, NY, USA: John Wiley & Sons, Inc.

Temme N, Tudzynski P. 2009. Does *Botrytis cinerea* ignore H_2O_2-induced oxidative stress during infection? Characterization of botrytis activator protein 1. *Molecular Plant–Microbe Interactions* **228**: 987–998.

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Research* **18**: 1979–1990.

Tominaga A, Nagata M, Futsuki K, Abe H, Uchiumi T, Abe M, Kucho K, Hashiguchi M, Akashi R, Hirsch AM *et al.* 2009. Enhanced nodulation and nitrogen fixation in the abscisic acid low-sensitive mutant enhanced nitrogen fixation 1 of *Lotus japonicus*. *Plant Physiology* **151**: 1965–1976.

Van de Velde W, Zehirov G, Szatmári Á, Debreczeny M, Ishihara H, Kevei Z, Farkas AE, Mikuláss KR, Nagy A, Tiricz H *et al.* 2010. Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* **327**: 1122–1126.

Vincent JM. 1970. *A manual for the practical study of root-nodule bacteria*. Oxford, UK: Oxford Press.

Vítková M, Mullerova J, Sadlo J, Pergl J, Pysek P. 2017. Black locust (*Robinia pseudoacacia*) beloved and despised: a story of an invasive tree in central Europe. *Forest Ecology and Management* **384**: 287–302.

Vitousek PM, Cassman K, Cleveland C, Crews T, Field CB, Sprent JI. 2002. Towards an ecological understanding of biological nitrogen fixation. *Biogeochemistry* **58**: 1–45.

Wang X, Guo X, Yu Y, Cui H, Wang R, Guo W. 2018. Increased nitrogen supply promoted the growth of non-N_2-fixing woody legume species but not the growth of N_2-fixing *Robinia pseudoacacia*. *Scientific Reports* **8**: 17896.

Wang Z, Zhang X, Lei W, Zhu H, Wu S, Liu B, Ru D. 2023. Chromosome-level genome assembly and population genomics of *Robinia pseudoacacia* reveal the genetic basis for its wide cultivation. *Communications Biology* **6**: 797.

Welch CJ, Talaga ML, Kadav PD, Edwards JL, Bandyopadhyay P, Dam TK. 2020. A capture and release method based on noncovalent ligand cross-linking and facile filtration for purification of lectins and glycoproteins. *Journal of Biological Chemistry* **295**: 223–236.

Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L *et al.* 2021. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovations* **2**: 100141.

Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**: 1859–1875.

Yang J, Lan L, Jin Y, Yu N, Wang D, Wang E. 2022. Mechanisms underlying legume–rhizobium symbioses. *Journal of Integrated Plant Biology* **64**: 244–267.

Yao Y, Zhang X, Huang Z, Li H, Huang J, Corti G, Wu Z, Qin X, Zhang Y, Ye X *et al.* 2022. A field study on the composition, structure, and function of endophytic bacterial community of *Robinia pseudoacacia* at a composite heavy metals tailing. *The Science of the Total Environment* **850**: 157874.

Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Roe BA. 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**: 520–524.

Yuan H, Hu B, Zhsh L, Sun HG, Zhou M, Rennenberg H. 2022. Physiological responses of black locust-rhizobia symbiosis to water stress. *Physiologia Plantarium* **174**: e13641.

Zhao Y, Zhang R, Jiang KW, Qi J, Hu Y, Guo J, Ma H. 2021. Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. *Molecular Plant* **14**: 748–773.

Zhou J, Deng B, Zhang Y, Cobb AB, Zhang Z. 2017. Molybdate in rhizobial seed-coat formulations improves the production and nodulation of *Alfalfa*. *PLoS ONE* **12**: e0170179.

Zhu Y, Tian Y, Han S, Wang J, Liu Y, Yin J. 2024. Structure, evolution, and roles of SWEET proteins in growth and stress responses in plants. *International Journal of Biological Macromolecules* **263**: 130441.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Schematic graph of rhizobia strain inoculation and noninoculation control treatments on the roots of *Robinia pseudoacacia* seedlings.

Fig. S2 Phenotypic schemes of roots (a) and seedlings (b) of *Robinia pseudoacacia* with rhizobia inoculation and noninoculation treatments.

Fig. S3 Genome scope analysis of the *Robinia pseudoacacia* assembly.

Fig. S4 Gene and DNA transposon density plot of the largest contigs.

Fig. S5 Size and telomer locations of the 20 largest contigs of the *Robinia pseudoacacia* assembly.

Fig. S6 (A) Dot plot showing a whole genome comparison between *Robinia pseudoacacia* on the nucleotide level. (B) Dot plot showing a comparison between two *Robinia pseudoacacia* genomes based on gene synteny, showing only the 10 largest scaffolds.

Fig. S7 Results of weighted correlation network analysis.

Fig. S8 Co-expression network analysis of *Robinia pseudoacacia* and *Medicago truncatula*.

Fig. S9 Gene ontology term enrichment analysis – biological process.

Fig. S10 Gene ontology term enrichment analysis – cellular component.

Fig. S11 Gene ontology term enrichment analysis – molecular function.

Table S1 Plant biomass, water content, nitrogen contents, and nodulation of *Robinia pseudoacacia* seedlings of Dongbei provenance as affected by rhizobia inoculation.

Table S2 Statistics of the transposable element prediction.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.