Genome Biology

RESEARCH

Open Access

Telomere-to-telomere genome of *Stylosanthes guianensis* uncovers symbiotic adaptation to phosphorus-deficient soils

Pandao Liu[1,2,3†], Chun Liu[1,4†], Wenhui Pu[1†], Jianyu Zhang[1,4], Ranran Xu[1,4], Liting Liu[1,4], Jiajia Luo[1], Rui Huang[1], Lingyan Jiang[4], Hengfu Huan[1], Lijuan Luo[4], Guodao Liu[1*], Rongshu Dong[1*] and Zhijian Chen[1,2,3*]

†Pandao Liu, Chun Liu, and Wenhui Pu contributed equally to this work.

*Correspondence: Guodao_Liu@163.com; dongrongshu@126.com; zjchen@catas.cn

[1] Tropical Crops Genetic Resources Institute & State Key Laboratory of Tropical Crop Breeding, Chinese Academy of Tropical Agricultural Sciences, Haikou/Sanya 571101/572024, China
[2] Key Laboratory of Crop Gene Resources and Germplasm Enhancement in Southern China, Ministry of Agriculture and Rual Affairs, Haikou 571101, China
[3] Key Laboratory of Tropical Crops Germplasm Resources Genetic Improvement and Innovation of Hainan Province, Haikou 571101, China
[4] School of Breeding and Multiplication, Sanya Institute of Breeding and Multiplication, Hainan University, Sanya 572025, China

Abstract

Background: *Stylosanthes guianensis*, a representative tropical legume, exhibits remarkable adaptation to low-phosphorus acidic soils. As a symbiotic species, it forms root nodule associations with rhizobia to fix atmospheric nitrogen, potentially enhancing phosphate use efficiency. This study aims to decipher the mechanisms linking root nodule symbiosis to low-phosphate adaptation in *Stylosanthes guianensis*.

Results: We present the first gap-free, telomere-to-telomere genome of *Stylosanthes guianensis* (1.20 Gb), containing 82.28% repetitive sequences and 34,728 genes, with 99.30% BUSCO completeness and a 29.05 LTR Assembly Index score. Integrated genomic data and multi-omics analyses reveal a coordinated symbiotic strategy. Specifically, roots enhance flavonoid biosynthesis, likely driven by tandem duplication of chalcone reductase genes, to facilitate robust symbiont recruitment, while nodule development was regulated by a conserved network centered on the transcription factor NIN. In nodules, multiple phosphate starvation response pathways are activated, including enhanced phosphate transport and recycling, membrane lipid remodeling, and phosphate-conserving metabolic bypasses to support nitrogen fixation. Furthermore, co-upregulation of vitamin B6 and nitrogen assimilation pathways suggests a role in mitigating oxidative stress and sustaining metabolic balance.

Conclusions: This study reveals that root nodule symbiosis in *Stylosanthes guianensis* underpins a multifaceted adaptation to low-phosphate stress, integrating enhanced symbiotic signaling, conserved nodule development, reprogrammed phosphate metabolism, and improved antioxidant protection. These findings provide insights into stress-resilient symbiosis and a genomic foundation for improving nutrient efficiency in legumes.

Keywords: *Stylosanthes*, Root nodule symbiosis, Nitrogen fixation, Tropical legume, Genome assembly, Phosphorus deficiency, Acidic soils

Background

The symbiotic relationship between legumes and rhizobia converts atmospheric nitrogen (N_2) into plant-usable ammonia. This process, called symbiotic nitrogen fixation (SNF), occurs within specialized root-derived structures termed nodules [1, 2]. Legume-rhizobia symbiosis is pivotal for global agriculture and ecosystem sustainability. It contributes 40–60 million metric tons of biologically fixed N annually, equivalent to 60% of global synthetic fertilizer use [3, 4]. This symbiosis also enhances biodiversity in nutrient-poor soils.

The symbiosis initiates with a molecular cascade: legume roots secrete flavonoids that activate bacterial *nod* genes. This triggers the synthesis of lipochitooligosaccharide signaling molecules called Nod factors. These factors induce root hair curling and the formation of infection thread, a plant-derived tubular structure. The infection thread grows inward, guiding the enclosed rhizobia through the root hair and into the underlying cortical tissues. Simultaneously, Nod factor signaling induces reactivation of cell division in the root cortex, leading to the formation of a nodule primordium. Coordinated development of infection thread and nodule primordium ultimately results in nodule organogenesis, giving rise to a mature root nodule [2, 5, 6]. Within mature nodules, the infection threads release rhizobia into host cells. The released bacteria are enclosed by a plant-derived membrane, forming organelle-like structures called symbiosomes. Inside symbiosomes, rhizobia differentiate into bacteroids. These bacteroids express the nitrogenase enzyme complex, which catalyzes the reduction of N_2 to ammonia [2, 5]. This process requires substantial energy consumption [1].

SNF is highly sensitive to environmental stresses like soil acidification, salinity, and high temperature [7, 8]. Notably, over 50% of global arable land faces soil pH below 5.5. In China, approximately 40% of cultivated land is acidic (pH 4.0–5.5), predominantly located in southern regions [9–11]. Plants cultivated in acidic soils generally suffer from multiple nutrient-related constraints. These include low availability of phosphorus (P), calcium (Ca), and magnesium (Mg). Plant also face phytotoxicity caused by proton (H^+), aluminum (Al^{3+}), and manganese (Mn^{2+}) [11–14]. These factors collectively inhibit crop productivity.

Phosphorus plays fundamental roles in cellular energy metabolism, macromolecule biosynthesis, and membrane integrity maintenance. P deficiency severely impairs nodulation and nitrogenase activity. This disruption of root nodule symbiosis has been documented in soybean (*Glycine max*), common bean (*Phaseolus vulgaris*), and white lupin (*Lupinus albus*) [15–18]. Despite being affected by phosphate (Pi) starvation, legumes have developed multifaceted strategies to maintain symbiosis activity under low P stress. These include forming symbioses with effective rhizobial strains, establishing preferential P sinks within nodules, enhancing P remobilization, and regulating Pi starvation response (PSR) genes [11, 19–22].

For example, inoculation with effective rhizobial strains enhances soybean yield and elevates N and P contents in low-P acidic soils [19]. Furthermore, rhizobial inoculation can enhance P acquisition from insoluble P sources through increasing H^+ and organic anion exudation in white lupin and soybean [17, 23]. To date, many symbiotic genes have been characterized as fundamental for nodule development and functioning [2, 24]. Furthermore, numerous PSR genes are also implicated in maintaining Pi homeostasis and

nodulation efficiency, such as genes associated with Pi transporters, acid phosphatases, expansins, and signal transduction pathways [25–30]. Thus, precise regulation of root nodule symbiosis is key for improving both N and P efficiency in legumes.

The genus *Stylosanthes* Sw. (Fabaceae) comprises ~ 50 tropical legume species distributed across the Americas, Africa, and Asia, where acidic, low-fertility soils are widespread [31–33]. We previously assembled a high-quality genome of the drought-tolerant *S. angustifolia* [34], which provides a valuable genomic resource for studies in this genus. Among *Stylosanthes* species, *S. guianensis* (2n = 2x = 20) is the most widely cultivated and economically important, serving as livestock forage and green manure while also providing soil cover [33, 35]. It demonstrates remarkable adaptation to multiple growth-limiting factors in acidic soils, particularly P deficiency [36–40]. These traits make *S. guianensis* an ideal model for investigating the role of root nodule symbiosis in low-P adaptation.

Despite its importance, *S. guianensis* still lacks a high-quality reference genome, particularly a telomere-to-telomere (T2T) assembly. By contrast, several model legumes, including *G. max* [41], *P. vulgaris* [42], and *Medicago truncatula* [43], have already been assembled as T2T genomes. This gap has limited genetic and functional studies and slowed progress in understanding the molecular mechanisms underlying root nodule symbiosis and low-P adaptation. Comprehensive genome sequencing and high-resolution functional annotation are therefore urgently needed to advance both basic research and practical breeding in *S. guianensis*.

In this study, we report the first gap-free, T2T genome assembly of *S. guianensis*. The assembly was generated using an integrated sequencing strategy leveraging short-read DNA nanoball sequencing (DNBSEQ), long-read PacBio high-fidelity (HiFi) and Oxford Nanopore Technologies (ONT) sequencing, and high-throughput chromosome conformation capture (Hi-C). Using a multi-omics framework, we decipher the molecular mechanisms underlying root nodule symbiosis-mediated adaptation of *S. guianensis* to low-P acidic soils.

Results

Sequencing, assembly, and quality assessment of the *S. guianensis* genome

Whole-genome sequencing of the *S. guianensis* accession 'S1979' using short-read DNBSEQ technology yielded 149.48 Gb of high-quality data (Additional file 2: Table S1). A k-mer-based genome survey estimated the genome size at ~ 1.11 Gb, with a heterozygosity rate of ~ 1.51% (Additional file 1: Fig. S1). Subsequently, the genome was sequenced using long-read platforms, yielding 196.86 Gb of ONT reads (~ 177 × coverage) and 17.08 Gb of ultra-long ONT reads (~ 15 × coverage) (Additional file 2: Table S2), as well as 121.39 Gb of PacBio HiFi reads (~ 109 × coverage) (Additional file 2: Table S3). Hi-C sequencing was then performed to construct a chromosome-level reference genome, yielding 164.53 Gb of clean data (~ 147 × coverage) (Additional file 2: Table S4). Thus, multi-platform sequencing generated sufficient coverage for a high-quality genome assembly.

De novo genome assembly was performed by integrating PacBio HiFi, ONT, and Hi-C data, yielding 134 contigs, including 10 that spanned the full length of all 10 chromosomes of *S. guianensis*. This resulted in a gap-free assembly, as confirmed by Hi-C

contact maps (Fig. 1). We aligned the remaining 124 contigs to *Stylosanthes* organellar genomes retrieved from the National Center for Biotechnology Information (NCBI) database and the chromosomal genome of the *S. guianensis* assembly, identifying 47 chloroplast-derived and 8 mitochondrial-derived fragments (Additional file 2: Table S5). The other 69 contigs likely represent redundant sequences in the assembly, as at least half of each contig aligned to the chromosomal genome with an average sequence identity of 85% (Additional file 2: Table S6). Consequently, a gap-free genome assembly was successfully achieved.

To assess telomere completeness in the assembled genome, TIDK software was used to identify the plant-typical telomeric sequence (TTTAGGG). All telomeres were detected except for a missing telomere at the start of chromosome 10. We then aligned ONT and PacBio HiFi reads to the genome and extracted those mapping to the first 1 Mb of chromosome 10, comprising 6,729 ultra-long ONT reads (147.1 Mb) and 38,378 PacBio HiFi reads (552.9 Mb). These reads were locally assembled using HiFiasm software. Alignment of the resulting contig to chromosome 10 revealed a 1.25 Mb sequence, of which the first 174 kb corresponded to the missing region and was subsequently incorporated. As a result, a gap-free, T2T genome assembly of *S. guianensis* was obtained, spanning 1.20 Gb across 10 chromosomes, with a contig N50 of 126.5 Mb and a GC content of 37.57% (Table 1; Fig. 1). In addition, 20 telomere and 10 centromere sequences were identified (Fig. 1; Additional file 2: Table S7). Therefore, a T2T genome assembly with all telomeres and centromeres was constructed.

We evaluated the *S. guianensis* genome assembly using multiple complementary metrics. Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis indicated 99.30% completeness, with 7.3% of BUSCOs complete and multi-copy (Additional file 2: Table S8). Long Terminal Repeat (LTR) integrity, assessed via the LTR Assembly Index (LAI), scored 29.05, corresponding to "Gold" quality (Table 1). Assembly continuity was

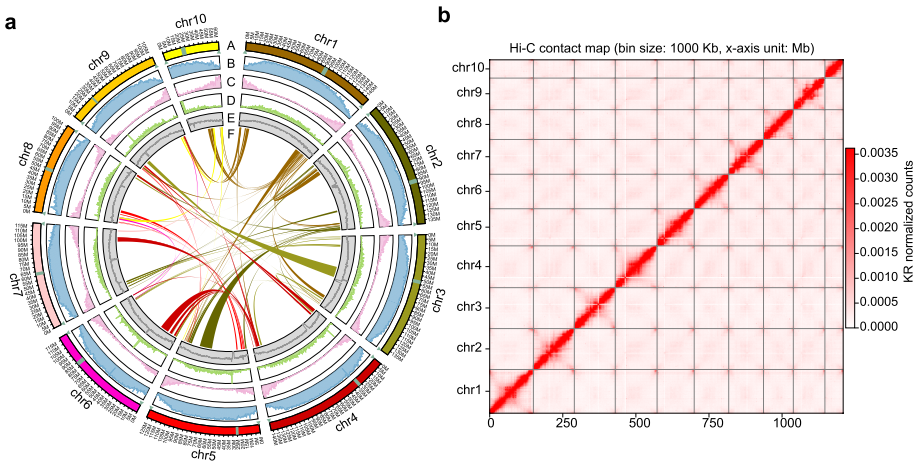

Fig. 1 Telomere-to-telomere (T2T) genome landscape and high-throughput chromosome conformation capture (Hi-C) contact map of *S. guianensis* accession 'S1979'. **a** Genome landscape. The tracks represent (A) the chromosomes, with triangles representing the telomere positions and rectangles within the chromosomes indicating the centromere regions; (B) transposable element density; (C) protein-coding gene density; (D) non-coding RNA density; (E) GC content; and (F) syntenic blocks within the *S. guianensis* genome. Blue triangles at the termini of each chromosome indicate the positions of telomeres, whereas the blue rectangles located in the central regions of the chromosomes denote the predicted centromere positions. Tracks B–E were calculated using 500-kb non-overlapping windows. **b** Hi-C contact map with a bin size of 1,000-kb

Table 1 Summary statistics of the genome assembly and annotation of *S. guianensis* accession 'S1979'

Elements	Value
Estimated genome size (Mb)	1,111.80
Assembly genome size (Mb)	1,204.87
Contig N50 (Mb)	126.5
Chromosome number	10
GC (%)	37.57
BUSCO (%)	99.30
LAI	29.05
GCI	100
Repetitive sequence (Mb)	991.38
Repetitive sequence (%)	82.28
No. protein coding genes	34,728
No. functional annotated genes	33,079

confirmed by aligning ONT and PacBio HiFi reads and analyzing them with Genome Continuity Inspector (GCI), which assigned a perfect score of 100 to the genome and each chromosome (Additional file 1: Figs. S2–S11; Additional file 2: Table S9). Collectively, these metrics confirm that the T2T assembly constitutes a high-quality, gap-free reference genome.

Genome annotation of the *S. guianensis* genome

We annotated repetitive sequences in the *S. guianensis* genome using a combination of de novo prediction and homology-based approaches. A total of 991.38 Mb of repetitive sequences were identified, accounting for 82.28% of the genome (Table 1; Fig. 1; Additional file 2: Table S10). LTR retrotransposons were the dominant class, accounting for 58.72% of the genome, whereas terminal inverted repeat (TIR) DNA transposons represented only 8.91%. Among LTR elements, Gypsy retrotransposons predominated (41.12%), substantially exceeding Copia elements (2.77%) (Additional file 2: Table S11). Within the TIR category, Mutator elements were the most prevalent (4.74%), followed by CACTA (2.85%) and hAT (1.13%) (Additional file 2: Table S11). Collectively, the *S. guianensis* genome is highly enriched in repetitive sequences, with LTR retrotransposons constituting a substantial proportion.

To support protein-coding gene prediction in *S. guianensis*, RNA sequencing (RNA-seq) was performed on 14 tissues, generating 136.89 Gb of clean data (Additional file 2: Table S12). Reads were analyzed through both de novo assembly and reference genome-guided transcript reconstruction, providing a foundation for subsequent gene prediction analyses. Full-length transcriptome sequencing was performed on the same tissues using PacBio Single Molecule Real-Time (SMRT), yielding 52.7 Gb of data (Additional file 2: Table S13) and identifying 182,741 transcripts. Open reading frames (ORFs) predicted from RNA-seq and Iso-Seq transcripts totaled 25,446 and 26,124, respectively. Additionally, de novo gene prediction applying GALBA and SNAP software generated 40,238 and 32,444 protein-coding genes. Homology-based prediction with amino acid sequences from *G. max*, *M. truncatula*, pigeonpea (*Cajanus cajan*), *Arabidopsis thaliana*, and *S. angustifolia* generated an average of 46,768 protein-coding genes. Integration of all

evidence with EvidenceModeler (EVM) produced a non-redundant set of 34,728 protein-coding genes in *S. guianensis* (Table 1; Additional file 2: Table S14). Gene number and length distributions were comparable to *C. cajan* and *S. angustifolia*, but lower than in *G. max* and *M. truncatula*. BUSCO analysis indicated 98.6% completeness (Additional file 2: Table S15). Coding sequence (CDS) length, exon number, and exon length were similar to those of *G. max*, *M. truncatula*, *C. cajan*, and *S. angustifolia* (Additional file 2: Table S14). Collectively, these results indicate that a reliable and highly complete set of 34,728 protein-coding genes was annotated in the *S. guianensis* genome.

For functional annotation, the *S. guianensis* gene set was queried against multiple reference databases, resulting in 33,079 genes (95.25% of the total) being assigned putative functions. NCBI Non-Redundant Protein Sequence Database (NR) and EggNOG yielded the highest annotation rates, at 95.16% and 91.40%, respectively (Additional file 2: Table S16). Based on functional annotations against the NR and EggNOG databases, Gene Ontology (GO) terms were assigned to 58.94% of the predicted protein-coding genes (Additional file 2: Table S16). The most abundant GO terms were cellular anatomical entity, binding, catalytic activity, cellular process, and metabolic process (Additional file 1: Fig. S12). Genome-wide analysis further identified 134 microRNAs (miRNAs), 16,442 small nuclear RNAs (snRNAs), 668 transfer RNAs (tRNAs), and 4,709 ribosomal RNAs (rRNAs) (Additional file 2: Table S17). Overall, the majority of predicted genes were successfully annotated, and a comprehensive set of non-coding RNAs was identified.

Comparative genomic analysis of leguminous plants

To investigate the genome evolution of *S. guianensis* and other legumes, we conducted an intra-species whole-genome synteny analysis using McScan software. We identified 256 syntenic blocks in *S. guianensis*, comparable to 290 in *M. truncatula* and 292 in *S. angustifolia*, but substantially fewer than the 1,072 blocks observed in *G. max*. Synonymous substitution rates (Ks) of genes within these blocks revealed a shared whole-genome duplication (WGD) among legumes, with a peak at 0.827, corresponding to ~ 50.92 million years ago (MYA) (Fig. 2a). No evidence of recent WGD was detected in *S. guianensis* (Fig. 2a). The WGD Ks peak of *S. guianensis* coincides precisely with that of *S. angustifolia* (Ks = 0.827), whereas *G. max* (Ks = 0.521) and *M. truncatula* (Ks = 0.725) display distinct peaks (Fig. 2a). Despite these differences, the overall distribution of Ks peaks aligns with the estimated timing of the Legume WGD (~ 59 MYA), suggesting that post-speciation evolutionary pressures have shaped the divergence of Ks values among these species. Thus, *S. guianensis* shares the ancient legume WGD event without any subsequent species-specific whole-genome duplication.

To explore the evolution of gene families in *S. guianensis* and other legumes, we identified and analyzed gene families across 11 legume species and Arabidopsis. In total, 36,252 orthologous groups (OGs) were detected, of which 8,914 were shared across all 12 species, 913 were legume-specific, and 394 were unique to *S. guianensis* (Additional file 1: Fig. S13). A maximum likelihood phylogeny based on 254 single-copy OGs showed that *S. guianensis* and *S. angustifolia* form sister branches, diverging ~ 9.4 MYA. The *Stylosanthes* lineage diverged from the wild diploid peanut relatives, *Arachis duranensis* and *Arachis ipaensis*, ~ 16.6 MYA (Fig. 2b). Analysis of gene family expansion and

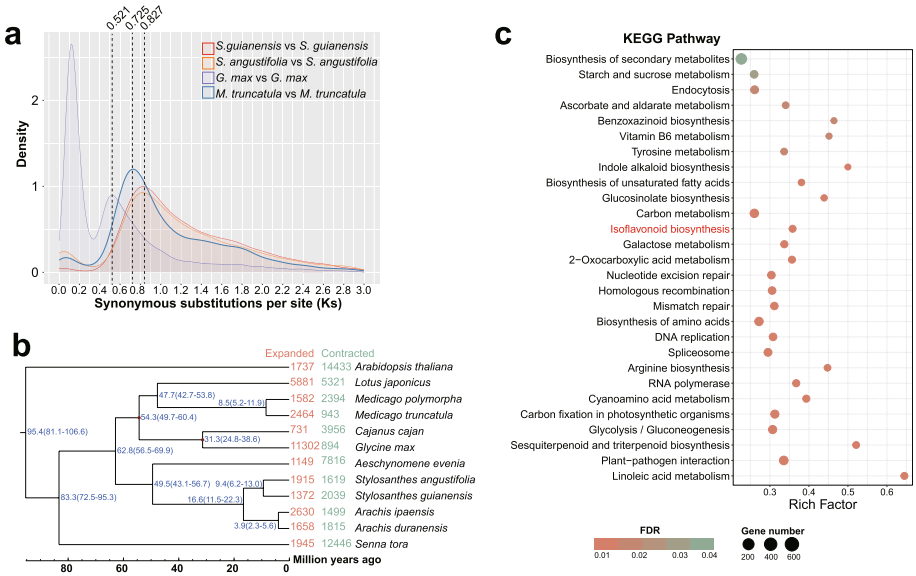

Fig. 2 Comparative genomic analysis of *S. guianensis* and other legume species. **a** Distribution of Ks values for syntenic genes in *S. guianensis*, *G. max*, *M. truncatula*, and *S. angustifolia*. The dashed lines and values represent the peak Ks values for the corresponding species. **b** Phylogenetic tree with estimated divergence times and gene family expansion/contraction analysis for 11 leguminous species and *A. thaliana*. The red circles indicate the divergence times derived from TimeTree. **c** KEGG enrichment analysis of significantly expanded pathways in the *S. guianensis* genome. All displayed pathways are significantly enriched at a false discovery rate (FDR) < 0.05

contraction revealed 1,372 expanded and 2,039 contracted OGs in *S. guianensis*, with the expanded OGs comprising 7,229 genes (Fig. 2b; Additional file 2: Tables S18–S20). Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis revealed that expanded OGs were significantly enriched in carbon metabolism, starch and sucrose metabolism, galactose metabolism, glycolysis/gluconeogenesis, linoleic acid metabolism, and the biosynthesis of secondary metabolites (Fig. 2c). Notably, isoflavonoid biosynthesis was also observed to be significantly enriched (Fig. 2c), suggesting that genes involved in flavonoid metabolism may have undergone adaptive evolution in *S. guianensis*. These results indicate that gene family expansion, particularly in metabolism-related pathways, may contribute to the environmental adaptation of *S. guianensis.*

Rhizobia isolation and inoculation effects on *S. guianensis* in acidic soils

A total of 46 rhizobial strains were isolated from the nodule of *S. guianensis*. These nodules were collected from the primary distribution regions of *S. guianensis* in southern China (Additional file 2: Table S21). Phylogenetic analysis based on 16S ribosomal RNA (rRNA) gene sequences revealed that all strains belonged to the genus *Bradyrhizobium* (Additional file 1: Fig. S14). Thus, *Bradyrhizobium* is the dominant microsymbiont associated with *S. guianensis* in acidic soils of southern China.

We performed inoculation analysis to assess the role of rhizobia symbiosis in the adaptation of *S. guianensis* to acidic soils. The rhizobial strain LZ3-2 isolated from acidic soils (Liuzhou, Guangxi; pH 4.5) was used. *S. guianensis* seedlings were subjected to two treatments: inoculation with *Bradyrhizobium* strain LZ3-2 (+Br) and non-inoculation (−Br). Plants were subsequently transplanted to a field in Danzhou, Hainan (pH

4.2; available N = 11.86 mg·kg^{-1}; available P = 2.18 mg·kg^{-1}) with no prior history of *S. guianensis* cultivation. Plants were harvested 60 days after inoculation. The results demonstrated that the +Br treatment significantly increased dry weight, N content, and P content in both shoots and roots compared with the −Br treatment (Fig. 3a–e). Notably, under the +Br treatment, N and P contents in the nodules were significantly higher than those in the roots (Fig. 3d, e). These results indicate that nodules are a strong sink for Pi and root nodule symbiosis enhances whole-plant Pi uptake and translocation.

Comparative analysis of SNF genes in nodules and roots

Based on previous literature [2, 6, 44, 45], we systematically curated a comprehensive dataset of SNF genes from *M. truncatula*, *Lotus japonicus*, *G. max*, and *P. vulgaris* (Additional file 2: Table S22). Using this reference dataset, we identified 742 putative SNF gene homologs in *S. guianensis*, which were classified into 11 symbiotic processes (Additional file 2: Table S23). Representative processes included early signaling, rhizobial infection, bacterial maturation, symbiosome formation, nodule organogenesis, nodule metabolism, autoregulation of nodule number, and senescence (Fig. 3f). Therefore, the *S. guianensis* genome harbors a comprehensive set of SNF gene homologs covering key symbiotic processes.

To comparative analysis of SNF genes in response to *Bradyrhizobium* infection in *S. guianensis*, RNA-seq was performed on nodules, non-inoculated roots (root−Br), and inoculated roots (root+Br). A total of 8,376 differentially expressed genes (DEGs) were identified in nodules relative to root+Br (4,058 upregulated and 4,318 downregulated). Compared with root−Br, 11,619 DEGs (5,773 upregulated and 5,846 downregulated) were detected in nodules (Additional file 1: Fig. S15; Additional file 2: Table S24). A venn diagram showed that 118 of the 742 SNF genes were upregulated in nodules relative to both root+Br and root−Br (Fig. 3g). These nodule-upregulated SNF genes were also assigned to the same 11 symbiotic processes described above (Additional file 2: Table S23). Thus, a core set of 118 SNF genes is significantly induced in nodules during symbiosis.

Among the SNF genes involved in nodule organogenesis, one *NODULE INCEPTION* (*NIN*) gene (*SgNIN*; SGM08g00013.1) exhibited the highest expression level in nodules (Fig. 3h). NIN is an RWP-RK DNA-binding domain containing family transcription factor that plays a central role in nodule organogenesis in *G. max*, *M. truncatula*, and *L. japonicus* [46–49]. Phylogenetic analysis showed that SgNIN clusters closely with its orthologs from *G. max* (GmNIN1a), *M. truncatula* (MtNIN), and *L. japonicus* (LjNIN) (Fig. 3i). In addition, the single-copy homolog of *GmNAC181* in *S. guianensis* (SGM04g01800.1) was also identified as a nodule-upregulated SNF gene (Additional file 1: Fig. S16; Additional file 2: Table S23). Interestingly, *GmNAC181* has been reported to act as a positive regulator of *GmNIN1a* in *G. max* [50]. These results suggest that an SgNIN-centered regulatory network may orchestrate root nodule symbiosis in *S. guianensis* under acidic soil conditions.

Comparative analysis of SNF genes in inoculated and non-inoculated roots

RNA-seq analysis identified 8,683 DEGs in root+Br compared with root−Br (4,532 upregulated and 4,151 downregulated) (Additional file 1: Fig. S15; Additional file 2:

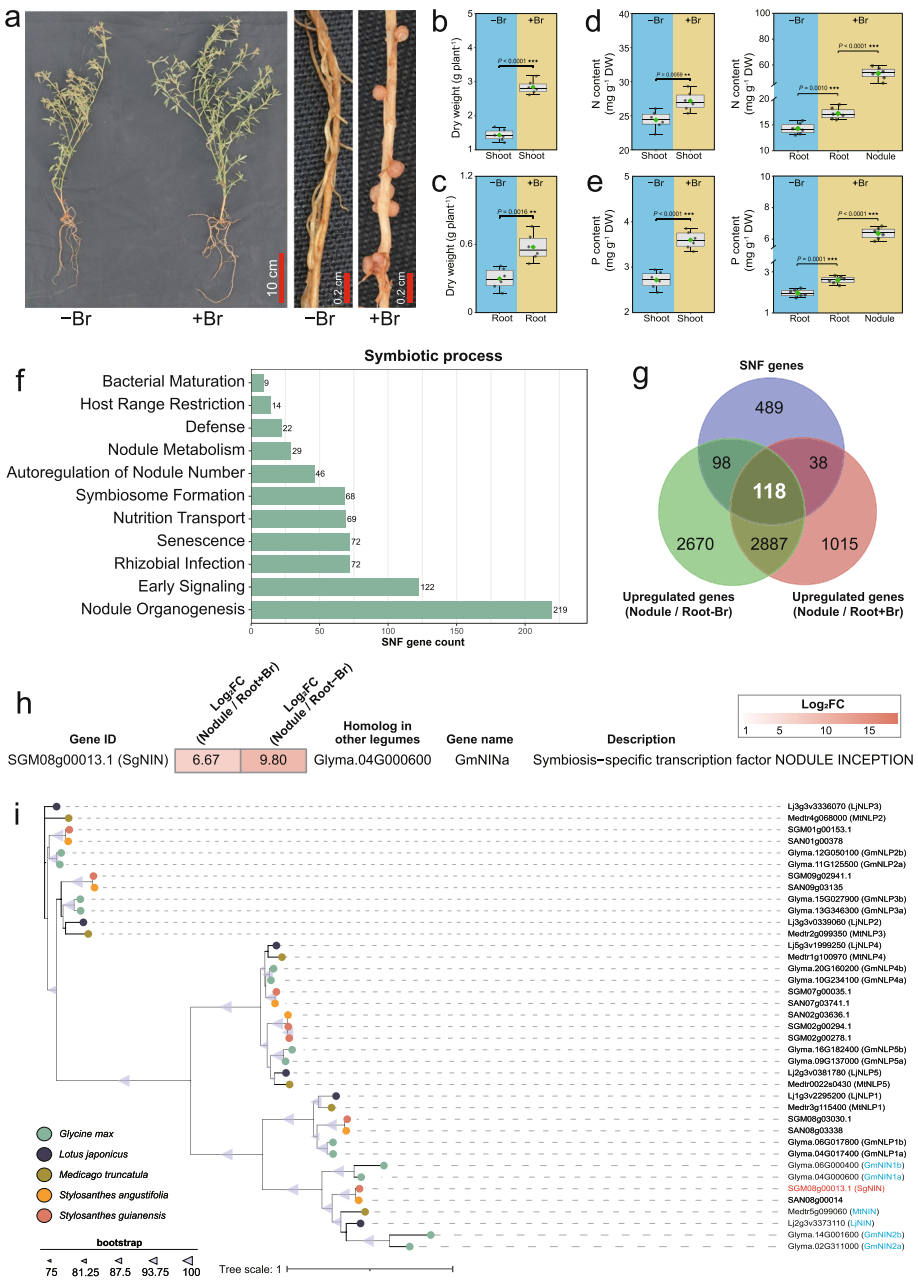

Fig. 3 Effects of rhizobial inoculation on *S. guianensis* growth and symbiotic nitrogen fixation (SNF)-related gene expression. **a**–**e** Phenotype (**a**), shoot dry weight (**b**), root dry weight (**c**), N content (**d**), and P content (**e**) of *S. guianensis* inoculated with *Bradyrhizobium* (+Br) or non-inoculated (−Br) under field conditions. Thirty-day-old seedlings were inoculated with *Bradyrhizobium* strain LZ3-2 for 1 h. Subsequently, seedlings were transplanted into a prepared field with no history of *S. guianensis* cultivation. Non-inoculated plants served as controls. Sixty days post-inoculation, shoots, roots, and nodules were separately harvested for analysis. In **b**–**e**, box plots display the median (center line), interquartile range (box), and minimum to maximum values (whiskers). Individual data points represent six independent biological replicates, and green diamonds indicate the mean values. Statistical significance between the two groups was determined using a two-sided Student's *t*-test (***P* < 0.01; ****P* < 0.001). **f** Distribution of SNF gene counts across symbiotic processes in *S. guianensis*. **g** Venn diagram showing the overlap between SNF genes and genes upregulated in nodules relative to non-inoculated (root−Br) and inoculated (root+Br) roots in *S. guianensis*. **h** Comparison of *SgNIN* expression in nodules relative to root−Br and root+Br groups. **i** Phylogenetic tree of NODULE INCEPTION (NIN) and NIN-like proteins (NLPs) in *S. guianensis*, *S. angustifolia*, *G. max*, *M. truncatula*, and *L. japonicus*

Table S24). Among these, 312 DEGs were SNF genes (Additional file 1: Fig. S17; Additional file 2: Table S23). Notably, SNF genes involved in flavonoid biosynthesis were markedly upregulated in root+Br relative to root−Br. These included 7 of 18 *chalcone reductase* (*CHR*) genes, 7 of 12 *chalcone synthase* (*CHS*) genes, 3 of 5 *2-hydroxyisoflavanone dehydratase* (*HID*) genes, all three *chalcone isomerase* (*CHI*) genes, and the single *isoflavone synthase* (*IFS*) gene (Fig. 4a). Consistent with gene expression, widely-targeted flavonoid metabolomic analysis revealed higher relative levels of specific metabolites in the root+Br than in the root−Br. These metabolites included isoliquiritigenin, naringenin chalcone, liquiritigenin, naringenin, daidzein, and genistein (Fig. 4b; Additional file 2: Table S25). Accordingly, rhizobial inoculation strongly activates the flavonoid biosynthesis pathway at both the transcriptional and metabolic levels.

Notably, *CHR* shows pronounced copy number expansion within the gene families of *S. guianensis* (OG0001439; Additional file 2: Tables S19 and S20). To characterize this expansion, we conducted a genome-wide identification of *CHR* genes in *S. guianensis*, *G. max*, and *M. truncatula*. The analysis revealed that *S. guianensis* harbors 18 *CHR* copies, substantially more than the 8 copies found in each of *G. max* and *M. truncatula* (Additional file 2: Table S26). Phylogenetic analysis and micro-synteny comparisons revealed species-specific tandem duplications on chromosome 8, resulting in ten *CHR* genes arranged in tandem in *S. guianensis* (Fig. 4c; Additional file 1: Fig. S18). To further validate this expansion, we performed reference-annotation-independent *CHR* identification in *G. max* and *M. truncatula*. Remarkably, the 18 *CHR* copies in *S. guianensis* were mapped to only five genomic loci in *G. max* and three loci in *M. truncatula*, strongly supporting a lineage-specific expansion of the *CHR* gene family in *S. guianensis* (Additional file 2: Table S27). This lineage-specific expansion of the *CHR* gene family may enhance root flavonoid biosynthesis, promoting rhizobial recruitment and efficient nodule formation in acidic soils.

Consistent with the enhanced flavonoid biosynthesis in inoculated roots, KEGG enrichment analysis revealed that upregulated genes in root+Br were also significantly enriched in the upstream phenylpropanoid biosynthesis pathway (Additional file 1: Fig. S19). These genes included two *phenylalanine ammonia-lyase* (*PAL*) genes, one *cinnamate 4-hydroxylase* (*C4H*) gene, and five *4-coumarate-CoA ligase* (*4CL*) genes, all of which showed higher expressed in root+Br than in root−Br (Additional file 1: Fig. S20). Thus, the phenylpropanoid-flavonoid pathway is transcriptionally activated in response to rhizobial inoculation.

Identification of PSR genes in nodules and inoculated roots

We performed weighted gene co-expression network analysis (WGCNA) to identify gene modules specifically associated with inoculated roots or nodules. WGCNA was conducted using expression profiles of DEGs identified in at least one pairwise comparison among the nodule, root+Br, and root−Br groups. To improve the robustness of network construction, these identified DEGs were analyzed alongside expression profiles from 11 additional tissues (Additional file 1: Fig. S21). WGCNA grouped the DEGs into 11 co-expression modules. Six modules exhibited close tissue-specific correlations ($r \geq 0.75$, $P < 0.01$; Fig. 5a). Among these, the brown module was most associated with nodules (2,339 genes, defined as the nodule-related module)

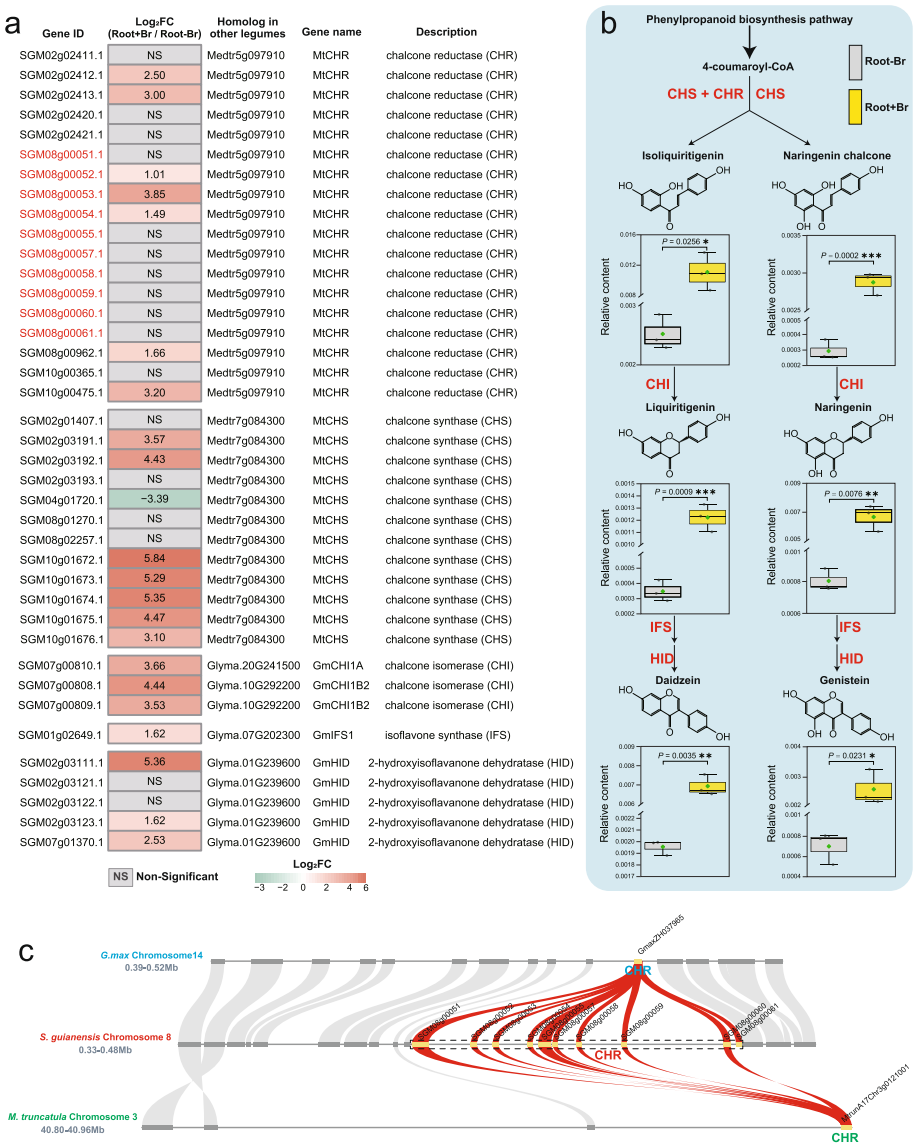

Fig. 4 Comparative analysis of flavonoid biosynthesis-related SNF gene expression and flavonoid accumulation in non-inoculated (root−Br) and inoculated (root+Br) roots of *S. guianensis*. **a** Heatmap showing differential expression of flavonoid biosynthesis-related SNF genes between root+Br and root−Br groups. **b** Differential accumulation of metabolites in the flavonoid biosynthesis pathway between root+Br and root−Br groups. The relative flavonoid content was derived from widely-targeted flavonoid metabolomic analysis. Box plots display the median (center line), interquartile range (box), and minimum to maximum values (whiskers). Individual data points represent three independent biological replicates, and green diamonds indicate the mean values. Statistical significance between the two groups was determined using a two-sided Student's *t*-test (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). **c** Microcollinearity of *CHR* genes in *S. guianensis* compared with those in *G. max* and *M. truncatula*. CHI, chalcone isomerase; CHR, chalcone reductase; CHS, chalcone synthase; IFS, isoflavone synthase; HID, 2-hydroxyisoflavanone dehydratase

(Fig. 5a; Additional file 2: Table S28). The red module was most associated with inoculated roots (1,226 genes, defined as the inoculated-root-related module) (Fig. 5a; Additional file 2: Table S28). Thus, WGCNA identified distinct gene modules specifically enriched in nodules versus inoculated roots.

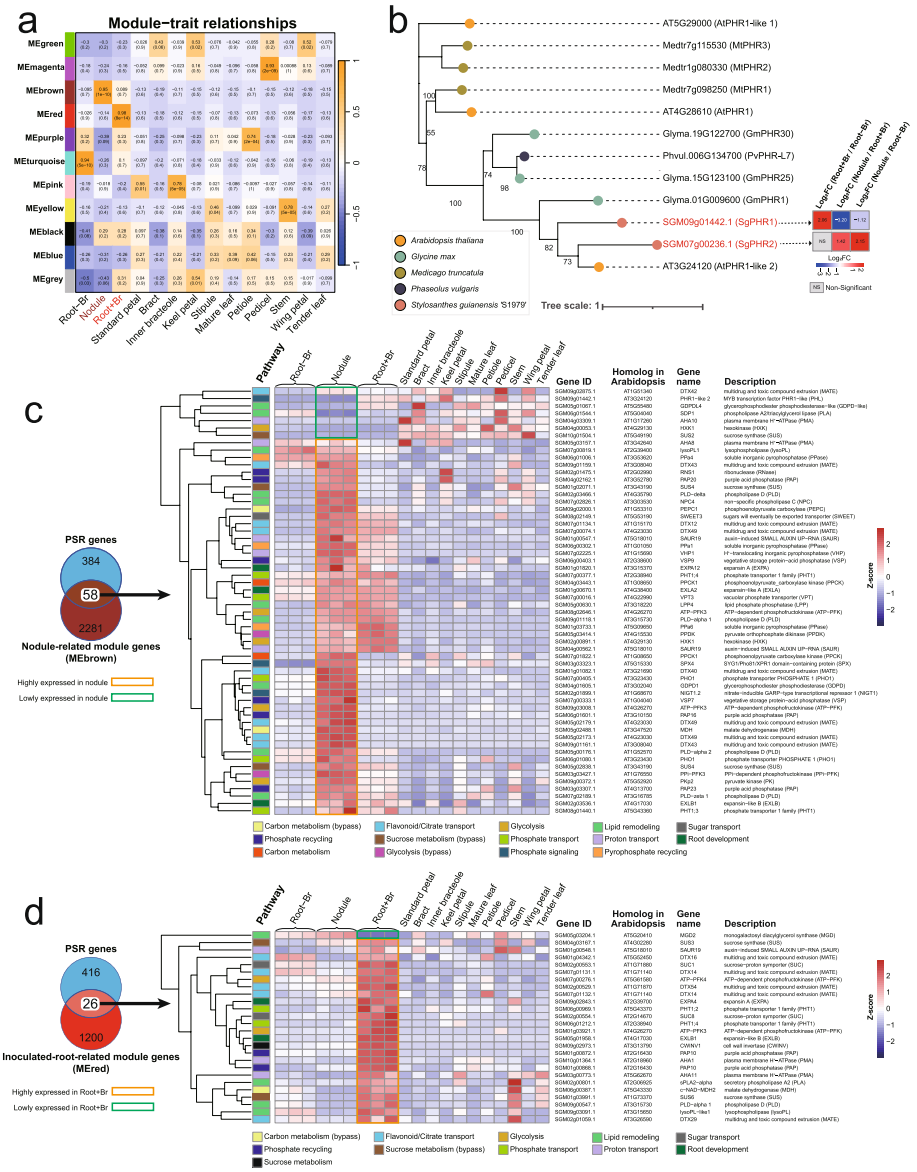

Fig. 5 Pi starvation response (PSR) genes in co-expression modules associated with nodules and inoculated roots of *S. guianensis*. **a** Identification of modules related to nodules and inoculated roots via weighted gene co-expression network analysis (WGCNA). WGCNA was performed using DEGs identified from pairwise comparisons among nodules, non-inoculated (root−Br), and inoculated (root+Br) roots, along with expression profiles from 11 additional tissues. The color key ranges from blue (negative) to white (neutral) to orange (positive), representing Pearson correlation coefficients from − 1 to 1. **b** Phylogenetic tree of PHOSPHATE STARVATION RESPONSE 1 (PHR1) and PHR-like (PHL) proteins from *S. guianensis*, *A. thaliana*, *G. max*, *M. truncatula*, and *P. vulgaris*. The heatmap depicts the differential expression of the *SgPHR1* and *SgPHR2* genes from *S. guianensis* across the nodule, root+Br, and root−Br groups. **c** Expression heatmap showing the overlap between PSR genes and nodule-related module genes. **d** Expression heatmap showing the overlap between PSR genes and inoculated-root-related module genes. Gene expression values were normalized to Z-scores

Based on previous studies [51, 52], we systematically compiled a dataset of 414 PSR genes in the model plant *A. thaliana* (Additional file 2: Table S29). Using this dataset, we identified 442 putative PSR genes in *S. guianensis* (Additional file 2: Table S30). In *A. thaliana*, PHR1 and its homologs PHR1-like (PHL1/PHL2) are central positive

regulators of PSR [53]. Here, we identified two homologous genes in *S. guianensis*, designated *SgPHR1* (SGM09g01442.1) and *SgPHR2* (SGM07g00236.1) (Fig. 5b; Additional file 2: Table S30). Phylogenetic analysis revealed that SgPHR1 and SgPHR2 cluster with GmPHR1 from *G. max* and AtPHR1-like 2 from *A. thaliana* (Fig. 5b). Additionally, *SgPHR1* was highly expressed in inoculated roots (root+Br), while *SgPHR2* was predominantly expressed in nodules (Fig. 5b). These results suggest their respective roles in regulating PSR in symbiotic roots and nodules.

Intersection analysis identified shared genes between PSR genes and the tissue-related modules. Specifically, 58 PSR genes overlapped with the nodule-related module (51 highly and 7 lowly expressed in nodules), and 26 PSR genes overlapped with the inoculated-root-related module (25 highly and 1 lowly expressed in inoculated roots) (Fig. 5c, d). Both sets contained genes involved in carbon metabolism, glycolysis, lipid remodeling, sugar transport, sucrose metabolism, phosphate recycling, phosphate transport, proton transport, and root development (Fig. 5c, d). Thus, distinct subsets of PSR genes are preferentially associated with the nodule-related and inoculated-root-related modules, implicating divergent strategies for symbiosis-related lowP adaptation in each tissue.

Notably, a greater number of PSR genes were highly or specifically expressed in nodules across multiple pathways. For example, in the phosphate transport pathway, two *phosphate transporter 1* (*PHT1*) genes (SGM08g01440.1 and SGM07g00377.1) were preferentially expressed in nodules. Additionally, two *PHOSPHATE 1* (*PHO1*) genes (SGM06g01080.1 and SGM07g00405.1) and one *vacuolar phosphate transporter* (*VPT*) gene (SGM07g00016.1) also showed high expression in nodules (Fig. 5c). In contrast, only two *PHT1* genes (SGM06g00969.1 and SGM06g01212.1) were highly expressed in inoculated roots (Fig. 5d). In the phosphate recycling pathway, three *purple acid phosphatase* (*PAP*) genes (SGM06g01601.1, SGM04g02162.1, and SGM03g03307.1), one *ribonuclease* (*RNase*) gene (SGM02g01475.1), and two *vegetative storage protein–acid phosphatase* (*VSP*) genes (SGM07g00333.1 and SGM06g00403.1) were highly expressed in nodules (Fig. 5c). By comparison, only two *PAPs* (SGM01g00868.1 and SGM01g00872.1) exhibited high expression levels in inoculated roots (Fig. 5d). Thus, nodules deploy a more extensive array of PSR genes for phosphate transport and recycling compared with inoculated roots.

In the lipid remodeling pathway, four *phospholipase D* (*PLD*) genes (SGM09g01118.1, SGM05g00176.1, SGM02g03466.1, and SGM07g02189.1) were highly expressed in nodules (Fig. 5c). In contrast, only one *PLD* gene (SGM09g00547.1) was predominantly expressed in roots (Fig. 5d). Furthermore, one *non-specific phospholipase C* (*NPC*) gene (SGM07g02826.1), one *glycerophosphodiester phosphodiesterase* (*GDPD*) gene (SGM04g01605.1), and one *lipid phosphate phosphatase* (*LPP*) gene (SGM05g00630.1) exhibited exclusively high expression in nodules (Fig. 5c). Lipid remodeling genes are preferentially activated in nodules, suggesting enhanced membrane phospholipid turnover during symbiosis.

In the flavonoid/citrate transport pathway, seven *multidrug and toxic compound extrusion* (*MATE*) genes (SGM01g01082.1, SGM05g02173.1, SGM05g02179.1, SGM07g00074.1, SGM07g01134.1, SGM09g01159.1, and SGM09g01161.1) showed high expression in nodules (Fig. 5c). Meanwhile, five *MATE* genes (SGM01g04342.1,

SGM02g00529.1, SGM02g01059.1, SGM07g01131.1, and SGM07g01132.1) exhibited high expression in inoculated roots (Fig. 5d). In the sugar transport pathway, one *sugars will eventually be exported transporter* (*SWEET*) gene (SGM08g02149.1) was highly expressed in nodules (Fig. 5c). By comparison, two *sucrose–proton symporter* (*SUC*) genes (SGM02g00553.1 and SGM02g00554.1) were highly expressed in inoculated roots (Fig. 5d). Thus, distinct transporter families involved in flavonoid/citrate and sugar transport are differentially regulated between nodules and inoculated roots, reflecting tissue-specific transport and secretory processes for these metabolites.

Pyrophosphate (PPi) recycling and energy-efficient metabolism in nodules

PPi is generated as a byproduct during the biosynthesis of nucleic acids, proteins, polysaccharides, and membrane lipids [54]. Pyrophosphatase catalyzes the hydrolysis of one molecule of PPi into two molecules of Pi. In nodules, three *soluble inorganic pyrophosphatase* (*PPase*) genes (SGM06g01006.1, SGM06g00302.1, and SGM01g03733.1) exhibited preferential expression (Fig. 5c). Consistently, PPase activity was higher in nodules than in inoculated roots (root+Br), but the relative PPi content was lower in nodules (Fig. 6a, b). Enhanced PPase activity in nodules facilitates PPi recycling and Pi generation, helping to meet the high P demand.

Another key finding was the high expression in nodules of genes encoding Pi- and adenylate-independent bypass enzymes, which function in glycolysis, sucrose metabolism,

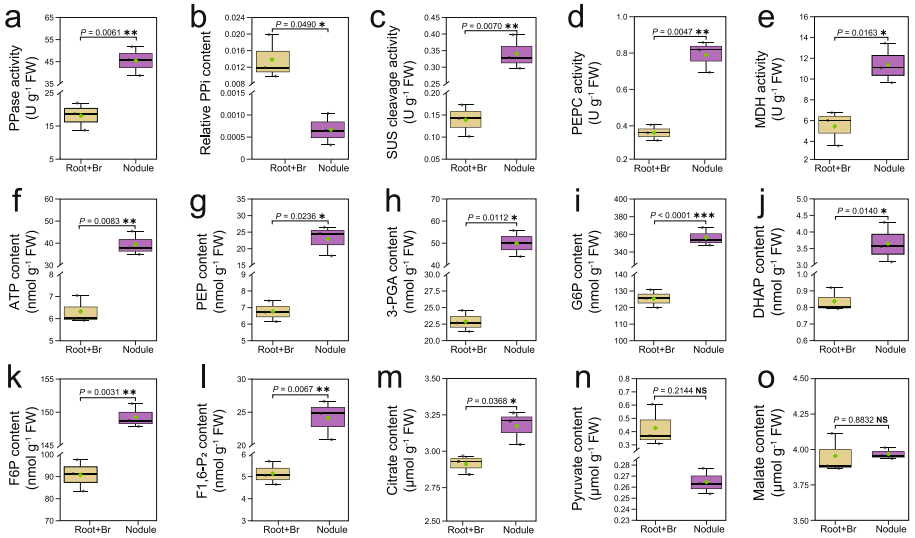

Fig. 6 Comparative analysis of central carbon metabolism-associated enzyme activities and metabolite accumulation in nodules and inoculated roots (root+Br) of *S. guianensis*. **a** Pyrophosphatase (PPase) activity. **b** Relative pyrophosphate (PPi) content, as determined by untargeted metabolomic profiling. **c** Sucrose synthase (SUS) cleavage activity. **d** Phosphoenolpyruvate carboxylase (PEPC) activity. **e** Malate dehydrogenase (MDH) activity. **f** Adenosine 5′-triphosphate (ATP) content. **g** Phosphoenolpyruvate (PEP) content. **h** 3-phosphoglycerate (3-PGA) content. **i** Glucose-6-phosphate (G6P) content. **j** Dihydroxyacetone phosphate (DHAP) content. **k** Fructose-6-phosphate (F6P) content. **l** Fructose-1,6-bisphosphate (F-1,6-P_2) content. **m** Citrate content. **n** Pyruvate content. **o** Malate content. Box plots display the median (center line), interquartile range (box), and minimum to maximum values (whiskers). Individual data points represent three independent biological replicates, and green diamonds indicate the mean values. Differences between the two groups were assessed for statistical significance using a two-sided Student's *t*-test (**P* < 0.05; ***P* < 0.01; ****P* < 0.001), with "NS" indicating no significant difference

and carbon metabolism. These genes included one *PPi-dependent phosphofructokinase* (*PPi-PFK*) gene (SGM03g03427.1), one *pyruvate orthophosphate dikinase* (*PPDK*) gene (SGM05g03414.1), two *sucrose synthase* (*SUS*) genes (SGM01g02071.1 and SGM05g02838.1), one *phosphoenolpyruvate carboxylase* (*PEPC*) gene (SGM09g02000.1), and one *malate dehydrogenase* (*MDH*) gene (SGM05g02488.1) (Fig. 5c). In line with this, activities of SUS, PEPC, and MDH were significantly higher in nodules than in root+Br (Fig. 6c–e). Thus, nodules upregulate Pi- and adenylate-independent bypass pathways to sustain carbon metabolism under low-P conditions.

Moreover, the enhancement of these bypass pathways was associated with altered metabolite levels. The contents of adenosine 5′-triphosphate (ATP) and several P-containing metabolites were significantly higher in nodules than in root+Br (Fig. 6f–l). These metabolites included phosphoenolpyruvate (PEP), 3-phosphoglycerate (3-PGA), glucose-6-phosphate (G6P), dihydroxyacetone phosphate (DHAP), fructose-6-phosphate (F6P), and fructose-1,6-bisphosphate (F-1,6-P_2) (Fig. 6g–l). Interestingly, citrate content was also higher in nodules than in root+Br (Fig. 6m). However, the levels of pyruvate and malate did not differ significantly between nodules and root+Br (Fig. 6n, o). The accumulation of phosphorylated intermediates in nodules reflects a metabolic shift toward P-conserving, energy-efficient glycolysis, ensuring adequate P allocation to sustain nitrogen fixation under low-P conditions.

Enhancement of vitamin B6 and nitrogen metabolisms in nodules

KEGG enrichment analysis revealed that genes upregulated in nodules relative to inoculated roots (root+Br) were significantly enriched in multiple pathways, including vitamin B6 and nitrogen metabolism (Fig. 7a). In the vitamin B6 metabolism pathway, one *pyridoxal 5'-phosphate synthase pdxS subunit* (*PDX1*) gene and three *pyridoxal phosphate phosphatase* (*PLPP*) genes were upregulated in nodules (Fig. 7b). Consistent with this result, untargeted metabolomic analysis revealed higher accumulation of three vitamin B6 components (pyridoxal, pyridoxine, and pyridoxamine) in nodules than in root+Br (Fig. 7c; Additional file 2: Table S31). Vitamin B6 biosynthesis is enhanced in nodules, potentially contributing to oxidative stress tolerance during nitrogen fixation.

Additionally, genes involved in nitrogen metabolism were upregulated in nodules. These included four *glutamine synthetase* (*GS*) genes, one *glutamate synthase* (*GOGAT*) gene, two *asparagine synthetase* (*AS*) genes, two *nucleoside hydrolase* (*NSH*) genes, one *xanthine dehydrogenase* (*XDH*) gene, two *allantoin synthase* (*ALNS*) genes, and one *allantoinase* (*ALN*) gene (Fig. 7d). Correspondingly, the levels of four nitrogen fixation products, including xanthosine, uric acid, and ureides (allantoin and allantoate), were significantly higher in nodules than in root+Br (Fig. 7e; Additional file 2: Table S31). Thus, nodules exhibit elevated ureide biosynthesis and upregulated nitrogen assimilation gene expression, consistent with active symbiotic nitrogen fixation.

Discussion

Decoding the T2T genome of *S. guianensis* to explore how root nodule symbiosis mediates adaptation to low-P acidic soils

The genus *Stylosanthes* Sw. plays important agroecological roles in tropical ecosystems [33, 35]. However, genomic resources for this genus remain scarce, particularly regarding

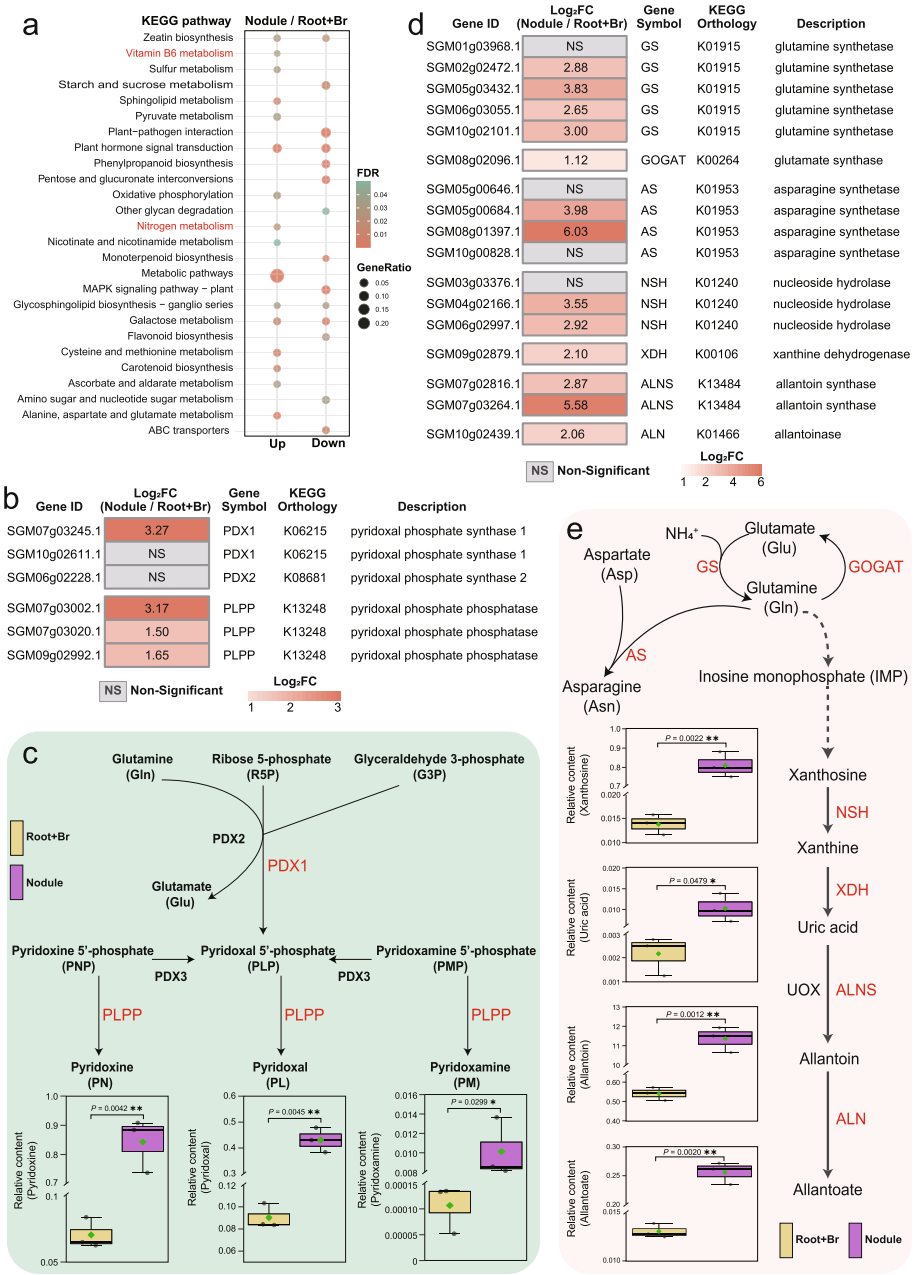

Fig. 7 Comparative analysis of vitamin B6 and nitrogen metabolism in nodules and inoculated roots (root+Br) of *S. guianensis*. **a** Enriched KEGG pathways of differentially expressed genes between the nodule and root+Br groups. All displayed pathways are significantly enriched with a false discovery rate (FDR) < 0.05. **b** Differential expression heatmap of *PDX* and *PLPP* genes involved in vitamin B6 metabolism between the nodule and root+Br groups. **c** Differential accumulation of metabolites in the vitamin B6 metabolism pathway between the nodule and root+Br groups. **d** Differential expression heatmap of genes involved in nitrogen metabolism between the nodule and root+Br groups. **e** Differential accumulation of metabolites in the nitrogen metabolism pathway between the nodule and root+Br groups. In **c** and **e**, The relative metabolite levels were derived from untargeted metabolomic analysis. Box plots display the median (center line), interquartile range (box), and minimum to maximum values (whiskers). Individual data points represent three independent biological replicates, and green diamonds indicate the mean values. Statistical significance between the two groups was determined using a two-sided Student's *t*-test (**P* < 0.05; ***P* < 0.01). ALN, allantoinase; ALNS, allantoin synthase; AS, asparagine synthetase; GOGAT, glutamate synthase; GS, glutamine synthetase; NSH, nucleoside hydrolase; PDX, pyridoxal phosphate synthase; PLPP, pyridoxal phosphate phosphatase; XDH, xanthine dehydrogenase

high-quality T2T assemblies. To address this gap, we present the T2T, gap-free genome of *S. guianensis*, the most widely cultivated and utilized species in the genus, known for its adaptability to low-P acidic soils. The analysis showed that *S. guianensis* shares a WGD event with other legumes, with no subsequent species-specific WGD events (Fig. 2a). This suggests its genetic diversity arises from alternative mechanisms like tandem gene duplication, gene loss, or expression changes. Comparative genomic analysis revealed significant gene expansion in *S. guianensis*. These expanded genes enrich in secondary metabolite biosynthesis, indole alkaloid biosynthesis, vitamin B6 metabolism, linoleic acid metabolism, carbon metabolism, and isoflavonoid biosynthesis (Fig. 2c).

This genomic foundation enables investigation of how root nodule symbiosis mediates adaptation of *S. guianensis* to low-P acidic soils. Previous studies reported acid-tolerant *Bradyrhizobium* strains efficiently fix nitrogen and enhance low-P tolerance in *S. guianensis* [55–58]. Our results confirmed rhizobial symbiosis improved *S. guianensis* growth under low-P conditions. Furthermore, inoculated plants had higher P content not only in nodules but also in roots and shoots than non-inoculated plants (Fig. 3a–e). Thus, nodules act as a strong Pi sink, and rhizobial symbiosis enhances whole-plant N and P nutrients in *S. guianensis*, consistent with other legumes [17, 19, 23, 59].

Coordinated regulation of SNF in *S. guianensis* during symbiosis

This study delineated the expression landscape of SNF genes during symbiosis in *S. guianensis*. The results confirmed deep conservation of core symbiotic regulators and revealed species-specific genetic adaptations for efficient nodulation.

A key finding is the strong activation of flavonoid biosynthesis in roots upon *Bradyrhizobium* perception. Upregulation of genes related to the dedicated pathway (*CHS*, *CHR*, *CHI*, *HID*) probably elevated the levels of flavonoids like naringenin, daidzein, and genistein (Fig. 4a, b). These compounds are established nod-gene inducers, facilitating partner recognition and infection [6, 60–62]. Notably, the *CHR* gene family is remarkably expanded via tandem duplications in *S. guianensis* (Fig. 4c). CHR is critical for synthesizing 5-deoxyisoflavonoids, key symbiotic signals for many legumes [2, 63]. We propose this gene dosage effect, combined with strong post-inoculation induction, enhances flavonoid production, ensuring reliable rhizobial recruitment and infection.

Transcriptomic comparison of nodules versus roots revealed a dedicated symbiotic program, with 118 SNF genes upregulated in nodules (Fig. 3g). Central to this was the transcription factor SgNIN. NIN is a master regulator conserved across legumes for rhizobial infection and nodule organogenesis [64–66]. Preferential expression of *SgNIN* in nodules suggests its role in nodule organogenesis. Furthermore, a co-upregulated NAC-domain transcription factor (homologous to GmNAC181) was identified. GmNAC181 directly activates *GmNIN1a* expression in soybean [50]. This suggests preservation of a hierarchical regulatory network centered on *SgNIN* in *S. guianensis*. Precise modulation of this core module likely ensures successful nodule growth and development in low-P acidic soils.

Systematic modulation of PSR in root nodule symbiosis

Our WGCNA analysis revealed a sophisticated, compartmentalized regulation of PSR genes during symbiosis. This involves an activation of a core PSR program in

inoculated roots. This program is later amplified in nodules to meet the metabolic demands of SNF under low-P condition.

We identified two PHR homologs with distinct expression: *SgPHR1* in roots and *SgPHR2* in nodules (Fig. 5b). PHR is a master regulator in legume-rhizobia symbiosis [18, 67]. Under P deficiency, PHR can inhibit nodulation to conserve Pi for the host, as demonstrated in *P. vulgaris* (PvPHR-L7) and *M. truncatula* (MtPHR2) [18, 68]. This process may involve suppressing key SNF genes (e.g., *NIN*) or flavonoid pathways [18, 68]. Conversely, under P sufficiency, PHR proteins promote nodulation. They likely activate flavonoid biosynthesis (e.g., MtPHR2 in *M. truncatula*) or improve root growth and nutrient status (e.g., GmPHR1 in *G. max*) [68, 69]. Thus, precise spatiotemporal control of PHR activity is crucial for balancing the Pi costs and benefits of symbiosis.

S. guianensis employs a spatially partitioned strategy to manage P nutrition under Pi-limited symbiosis (Fig. 3e). In inoculated roots, high expression of two *PHT1* genes suggests their roles in Pi uptake in roots (Fig. 5d), as observed in *G. max* [25, 26]. In addition, upregulation of specific *MATE* genes was observed in roots with rhizobial inoculation (Fig. 5d), which may enhance citrate/flavonoid exudation, thereby mobilizing soil Pi and/or amplifying rhizobial signaling [70–72]. These root-level adaptations prime the plant for improved Pi acquisition and translocation, aligning with increased whole-plant P content.

Within the nodules, a specialized and intensified Pi management program is characterized by a comprehensive reinforcement of several pathways (Fig. 8). For example, in nodules, not only two *PHT1* homologs but also two *PHO1* homologs and one *VPT* homolog exhibited high expression levels (Fig. 5c), suggesting optimized Pi allocation from the host and sequestration of Pi in symbiotic tissues, as observed in other legumes [25, 26, 73–75]. Moreover, the pronounced induction of genes encoding enzymes involved in phospholipid catabolism, including *PLD*, *NPC*, *phospholipase A* (*PLA*), *lysophospholipase* (*lysoPL*), *GDPD*, and *LPP*, reflects a reprogramming of membrane lipid metabolism in nodules under low-P conditions (Fig. 8). When experiencing Pi starvation, plants can replace phospholipids with sulfolipids and galactolipids to conserve Pi through the action of these enzymes [76, 77]. This remodeling likely liberates Pi for essential metabolic processes while maintaining membrane integrity in infected cells, a critical adaptation for nitrogen fixation.

Another key finding is the activation of Pi-conserving metabolic bypasses in nodules. Genes encoding enzymes involved in central carbon metabolic bypass pathways (e.g., *SUS*, *PPi-PFK*, *PPDK*, *PEPC*) were markedly induced (Fig. 8). These enzymes reduce adenylate (ATP and ADP) and Pi consumption by utilizing PPi as a phosphoryl donor and are thus regarded as alternative "Pi-free" metabolic pathways [76, 78]. In addition, the concomitant high soluble inorganic PPase activity in nodules (Fig. 6), which likely hydrolyzes PPi into available Pi [54, 79, 80]. The elevated levels of phosphorylated glycolytic intermediates (e.g., PEP, G6P, F6P, F-1,6-P_2) and ATP in nodules provide direct metabolic evidence (Fig. 6). This reprogramming redirects carbon flux, minimizes Pi dependency, and secures Pi for nitrogen fixation. It supports the nodule's role as a Pi sink, regulating Pi homeostasis.

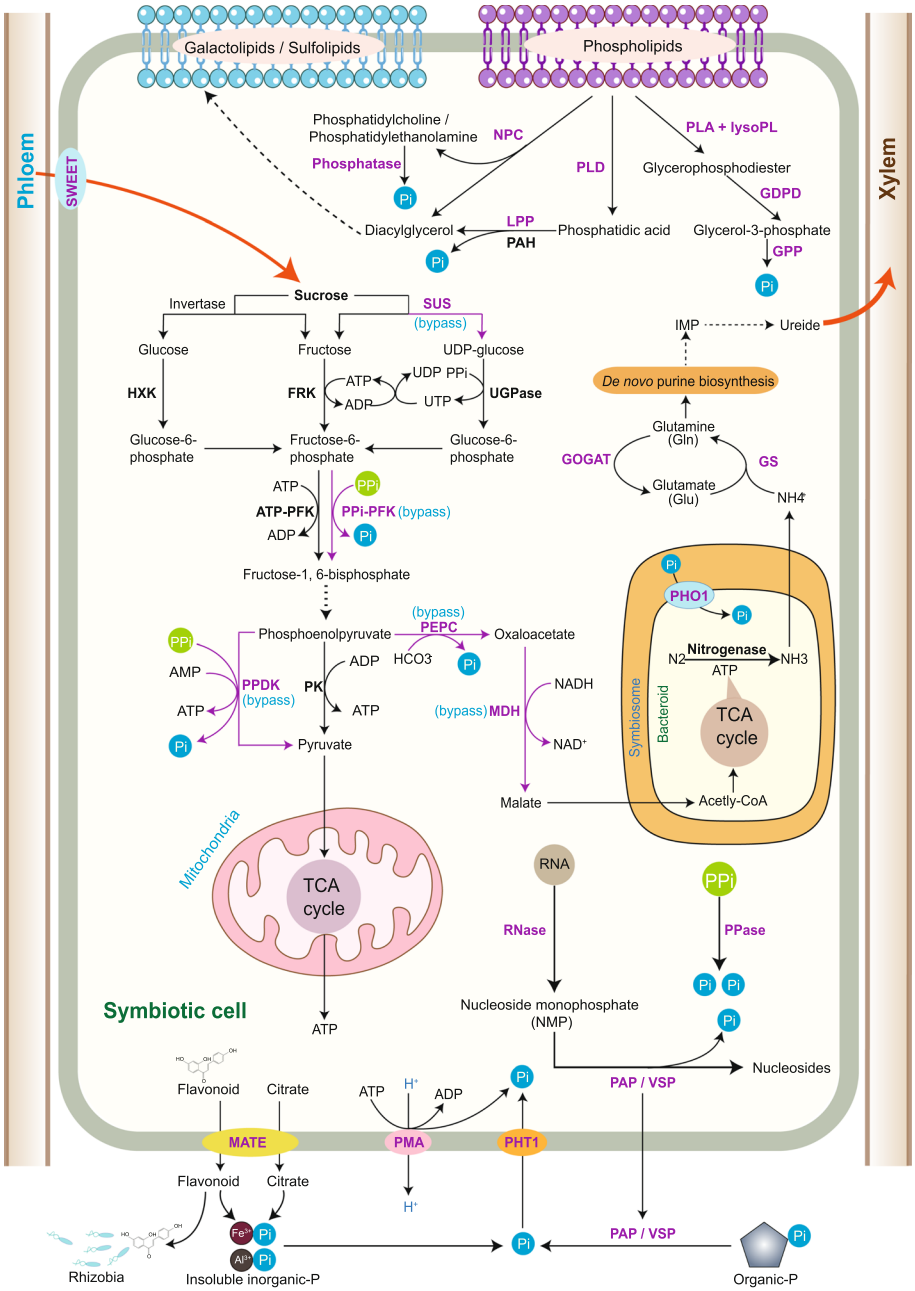

Fig. 8 Schematic model of major metabolic activities and transport processes in the nodules of *S. guianensis* under low-P acidic soil conditions. Genes with high expression levels in nodule-related metabolic and transport pathways are marked in purple. GDPD, glycerophosphodiester phosphodiesterase; GOGAT, glutamate synthase; GPP, glycerol-3-phosphatase; GS, glutamine synthetase; LPP, lipid phosphate phosphatase; lysoPL, lysophospholipase; MATE, multidrug and toxic compound extrusion; MDH, malate dehydrogenase; NPC, non-specific phospholipase C; PAP, purple acid phosphatase; PEPC, phosphoenolpyruvate carboxylase; PHO1, phosphate transporter PHOSPHATE 1; PHT1, phosphate transporter 1 family; PLA, phospholipase A; PLD, phospholipase D; PMA, plasma membrane H^+-ATPase; PPase, soluble inorganic pyrophosphatase; PPDK, pyruvate orthophosphate dikinase; PPi-PFK, PPi-dependent phosphofructokinase; RNase, ribonuclease; SUS, sucrose synthase; SWEET, sugars will eventually be exported transporter; VSP, vegetative storage protein-acid phosphatase

Integration of vitamin B6 and nitrogen metabolism for nodule function

Beyond the above outlined strategies for nodulation and PSR, the nodule exhibited coordinated enhancement of vitamin B6 and nitrogen metabolism. This integration functionally mitigates oxidative stress. The coordinated upregulation of *PDX1* and *PLPP* genes promoted the accumulation of vitamin B6 vitamers (pyridoxal, pyridoxine, and pyridoxamine) in nodules (Fig. 7b, c), indicating a reinforced antioxidant system. Beyond its role as an enzymatic cofactor, vitamin B6 is a potent non-enzymatic antioxidant that scavenges reactive oxygen species (ROS) [81–83]. Nodules are considered major sites of ROS generation due to high metabolic activity and oxygen-sensitive nitrogenase [84, 85]. The induced vitamin B6 likely protects cellular structures and enzymes from oxidative damage, maintaining nodule function.

Genes for nitrogen assimilation (*GS*, *GOGAT*, *AS*) and ureide biosynthesis (*XDH*, *ALNS*, *ALN*) were upregulated, corroborated by elevated ureide precursors in nodules (Fig. 7d, e). This indicates that *S. guianensis* nodules preferentially export fixed nitrogen as ureides (allantoin and allantoate), which are major nitrogen sources for the host (Fig. 8), as also observed in soybean, cowpea (*Vigna unguiculata*), and common bean [1, 86, 87]. Furthermore, the high-flux nitrogen fixation pathways are prone to generating metabolic by-products that can exacerbate oxidative stress [84, 88, 89]. The induced vitamin B6 system likely scavenges ROS, protecting the activity of enzymes in nitrogen metabolism. This ensures efficient nitrogen fixation and export of ureide, a critical strategy for symbiosis in low-P acidic soils.

Conclusions

This study revealed root nodule symbiosis not merely as an N fixation mechanism, but as a gateway that reprograms Pi homeostasis and metabolic pathways. This reprogramming enables plant survival in low-P soils. Key elements included expanded flavonoid biosynthesis genes enhanced rhizobial recruitment, a core symbiotic regulatory network centered on SgNIN, activation of Pi-conserving metabolic bypasses, and coupling of vitamin B6 metabolism with efficient nitrogen assimilation. These findings offer potential targets for engineering nutrient-efficient crops. The newly assembled T2T gap-free genome of *S. guianensis* holds substantial implications for plant genomics, comparative genomics, and legume cultivation improvement. Future work should explore functional validation of candidate genes and their epistatic interactions under field conditions.

Methods

Plant materials

The *Stylosanthes guianensis* accession 'S1979' used in this study was provided by the Tropical Crops Genetic Resources Institute (TCGRI), Chinese Academy of Tropical Agricultural Sciences (CATAS), Hainan, China.

Genome sequencing

Genomic DNA was extracted from tender leaves using a modified cetyltrimethylammonium bromide (CTAB) method optimized for plant tissue. After quality assessment, high-quality genomic DNA was used for DNBSEQ, PacBio HiFi, ONT, and Hi-C sequencing.

For short-read DNBSEQ sequencing, library construction was performed at BGI-Shenzhen (Shenzhen, China). Genomic DNA was randomly fragmented, followed by end repair, adapter ligation, and circularization. Rolling circle amplification was then applied to generate DNA nanoballs, which were sequenced on the DNBSEQ-T7 platform using paired-end (PE) sequencing with a read length of 150 bp. Raw reads were filtered and quality controlled using SOAPnuke software [90].

For ONT sequencing, both standard (15 kb) and ultra-long (150 kb) libraries were constructed. For the 15 kb library, DNA fragments of approximately 15 kb were size-selected using the BluePippin system automated nucleic acid recovery system (Sage Science, USA) and used for ONT library preparation. Sequencing was performed on the Nanopore PromethION platform (Oxford Nanopore Technologies, UK) at BGI-Shenzhen (Shenzhen, China) using single-end mode. For the ultra-long library, genomic DNA fragments longer than 150 kb were selected using the SageHLS HMW library system (Sage Science). Ultra-long ONT sequencing was carried out on the PromethION platform at the Genome Center of Grandomics (Wuhan, China) using single-end sequencing. Raw reads from both standard and ultra-long ONT sequencing were filtered using NanoFilt (v2.8.0; https://github.com/wdecoster/nanofilt) [91].

For PacBio HiFi sequencing, libraries were prepared using the SMRTbell® prep kit 3.0 (PacBio). High-quality libraries (20 kb) were sequenced on the PacBio Revio system at the Genome Center of Grandomics (Wuhan, China) using single-end sequencing. HiFi reads were generated using circular consensus sequencing (CCS) software (v8.0.0; https://github.com/PacificBiosciences/ccs) with a minimum of three passes and a minimum read quality (RQ) of 0.99.

For Hi-C sequencing, the libraries were constructed at the Genome Center of Grandomics (Wuhan, China) following a standard in situ Hi-C protocol with minor modifications for plant tissues. Approximately 2 g of fresh plant tissue was cut into small strips (1 to 2 mm), and crosslinked with formaldehyde under vacuum infiltration, after which the reaction was quenched with glycine. Crosslinked tissues were washed, frozen in liquid nitrogen, and ground to a fine powder for nuclei isolation. Isolated nuclei were lysed, and chromatin DNA was digested with the restriction enzyme DpnII (NEB). The resulting restriction fragment ends were filled in and labeled with biotinylated nucleotides, followed by proximity ligation under dilute conditions to favor intranuclear ligation events. After ligation, crosslinks were reversed by proteinase K treatment, and DNA was purified using a commercial DNA purification kit. Purified Hi-C DNA was sheared to an average fragment size of approximately 400 bp and used for library construction. The final Hi-C libraries were sequenced on the MGI-2000 platform using paired-end 150 bp reads. Raw Hi-C reads were filtered using SOAPnuke software (v2.1.7) to obtain high-quality sequencing data [90].

Genome assembly and assessment

The genome was assembled de novo using HiFiasm software (v0.20.0; https://github.com/chhylp123/hifiasm) with a hybrid approach [92], combining PacBio HiFi, ONT (standard and ultra-long reads), and Hi-C data [93]. Subsequently, Hi-C data alignment and chromosome grouping were carried out using Juicer (v1.6; https://github.com/aidenlab/juicer) and 3D de novo assembly (3D-DNA, v180922; https://github.com/aidenlab/

3d-dna) pipeline [94, 95]. The resulting Hi-C crosslinking data were visualized using Juicebox (v1.11.08; https://github.com/aidenlab/Juicebox), which confirmed that the 10 assembled contigs corresponded to 10 chromosomes, with the remaining contigs showing no Hi-C crosslinking signals.

To further investigate the unplaced contigs, chloroplast and mitochondrial sequences from the genus *Stylosanthes* were retrieved from NCBI (https://www.ncbi.nlm.nih.gov/nuccore/). Unplaced contigs were then aligned to the collected sequences using minimap2 (v2.28-r1209; https://github.com/lh3/minimap2) [96], followed by statistical analysis of alignment length, identity, and query coverage. Contigs with a query coverage > 50% were designated as chloroplast or mitochondrial fragments. Subsequently, the remaining contigs were aligned against the chromosomes of *Stylosanthes guianensis* using the same approach to further identify potential redundant sequences.

Telomeric sequences in plants (TTTAGGG) were searched using TIDK (https://github.com/tolkit/telomeric-identifier), which revealed the absence of the telomeric sequence at the start of chromosome 10. To resolve this, we aligned the ONT and PacBio HiFi reads to the assembled genome using winnowmap (v2.03; https://github.com/marbl/winnowmap) [97]. Based on the alignment results, we used SAMtools (v1.9; https://github.com/samtools/samtools) [98] to extract ONT and PacBio HiFi reads that aligned to positions from 1 to 1 Mb of chromosome 10. Moreover, these reads were locally assembled using both HiFiasm and NextDenovo (v2.5.2; https://github.com/Nextomics/NextDenovo) software [99]. The consistency of the assembly was confirmed by synteny analysis with MUMmer (v4.0.4; https://github.com/mummer4/mummer) [100], which demonstrated excellent agreement between the HiFiasm assembly and the first 1 Mb of chromosome 10. The assembly was extended by 174 kb, which was then added to the start of chromosome 10.

To visualize the Hi-C crosslinking interactions, the Hi-C data were mapped to the T2T genome using HapHiC (https://github.com/zengxiaofei/HapHiC) [101], and the Hi-C interaction heatmap was generated. Centromere regions were annotated using CentIER software [102].

The quality of the genome assembly was assessed using several approaches, including evaluations based on conserved genes, LTRs, and single-base resolution. Genome completeness was evaluated with BUSCO (v5.8.0; https://gitlab.com/ezlab/busco) [103] (with parameters set to -m geno) using the embryophyta_odb10 dataset. Additionally, LTR_retriever (v3.0.2; https://github.com/oushujun/LTR_retriever) [104] was performed to identify complete LTRs and calculate the LTR Assembly Index. Furthermore, genome assembly continuity was evaluated at single-base resolution using GCI (v1.0; https://github.com/yeeus/GCI), based on ONT and PacBio HiFi reads [105].

Rhizobial strain isolation and purification

Root nodules were collected from *Stylosanthes guianensis* at 15 long-term field sites (> 5 years of continuous cultivation) in southern China, including Sichuan, Yunnan, Guangdong, Guangxi, and Hainan provinces. Soil properties, including soil pH, and N, P, and K contents, from all sampling locations are listed in Additional file 2: Table S21. The collected nodules were used to isolate and purify rhizobia as previously described [19]. Fresh nodules were surface-sterillized with 95% (v/v) ethanol for 3 min, 0.2% (w/v)

$HgCl_2$ for 5 min, and rinsed with sterile deionized H_2O. Bisected nodules were aseptically streaked onto solid yeast mannitol agar (YMA) medium. The YMA medium contained 3 $g{\cdot}L^{-1}$ yeast extract, 10 $g{\cdot}L^{-1}$ mannitol, 0.2 $g{\cdot}L^{-1}$ $MgSO_4{\cdot}7H_2O$, 0.1 $g{\cdot}L^{-1}$ (0.0017 M) NaCl, 0.25 $g{\cdot}L^{-1}$ K_2HPO_4, 0.25 $g{\cdot}L^{-1}$ KH_2PO_4 and 15 $g{\cdot}L^{-1}$ agar (pH 6.8). The culture was incubated inverted at 28 °C for 6–10 days. Presumptive rhizobial colonies were subcultured three times on the same medium to obtain axenic cultures. Gram staining and microscopy (100 ×) confirmed bacterial purity and morphological consistency with rhizobial taxa.

DNA extraction and 16S rDNA sequencing

Total genomic DNA was isolated from rhizobial cultures using the E.Z.N.A.™ Bacterial DNA Kit (Omega, USA). The 16S rDNA gene was amplified by PCR analysis. PCR reaction mixture contained 10 × PCR buffer, 2 mM dNTPs, 10 μM primers (F: CTKAAG AGTTTGATCMTGGCTCAGATTGAAC; R: TACGGYTACCTTGTTACGACTT), 50 ng template DNA, and Ex Taq polymerase (Takara, China). PCR conditions were as follows: 94 °C (4 min); 35 cycles of 94 °C (1 min), 56 °C (1 min), and 72 °C (1 min); and a final extension at 72 °C (5 min). PCR products were electrophoresed on 1% agarose gels and then purified using the Gel Extraction Kit (Qiagen, China). The resulting products were sequenced by Sangon Biotech (Shanghai, China). The obtained 16S rDNA sequences were aligned against reference strains from NCBI GenBank. Multiple sequence alignment of 16S rDNA was performed using MUSCLE (v5.3; https://github.com/rcedgar/muscle/tree/v5.3) [106]. The resulting alignment was subsequently trimmed with trimAl (v1.5.0; https://github.com/inab/trimal) [107] to remove ambiguously aligned regions. Phylogenetic relationships were inferred using the maximum-likelihood method in IQ-TREE (v2.3.6; https://github.com/iqtree/iqtree2) [108], and branch support was assessed with 1,000 iterations.

Stylosanthes guianensis growth under field conditions

A field experiment was conducted in 2023 on a typical acidic red soil at the TCGRI, CATAS, Danzhou City, Hainan Province, China (19°30′N, 109°30′E). Soil characteristics were as follows: pH 4.2; organic matter, 8.6 $g{\cdot}kg^{-1}$; available P, 2.18 $mg{\cdot}kg^{-1}$; available N, 11.86 $mg{\cdot}kg^{-1}$; available K, 15.86 $mg{\cdot}kg^{-1}$. Seeds of *Stylosanthes guianensis* were soaked in hot water (80 °C) for 3 min and then rapidly cooled to room temperature to facilitate germination. The pretreated seeds were germinated on wet filter paper overnight in the dark at 28 °C and then transferred to nursery bags filled with sterilized soil. After 30 days of growth, the roots were inoculated with rhizobial strain LZ3-2 for 1 h. Subsequently, the nursery bags were carefully removed and seedlings with intact soil-root balls were transplanted into a prepared field with no history of *Stylosanthes guianensis* cultivation. Non-inoculated plants served as controls. No fertilizers were applied to the field. The experimental area comprised six plots (1.8 m × 1.2 m), with five plants harvested per plot. Sixty days post-inoculation, shoots, roots, and nodules were separately harvested for physiological and biochemical measurements, as well as transcriptomic and metabolomic analyses. The samples were oven-dried at 75 °C to constant weight to determine dry weight. Dried samples were ground and digested with an H_2SO_4-H_2O_2 reagent. N and P contents were then determined, with P measured using the molybdenum blue

colorimetric method spectrophotometrically detected at 700 nm and N determined by the semimicro-Kjeldahl method with a nitrogen analyzer [109, 110].

Transcriptome sequencing and analyses

Total RNA was isolated from 14 distinct tissues collected from *Stylosanthes guianensis* precultured as described above. These tissues included mature leaves, tender leaves, petioles, stipules, bracts, inner bracteoles, standard petals, keel petals, wing petals, pedicels, stems, nodules, root−Br (root without rhizobial inoculation), and root+Br (root with rhizobial inoculation). For root−Br, root+Br, and nodule samples, three independent biological replicates were included. RNA extraction, library construction, and sequencing were performed by BGI-Shenzhen (Shenzhen, China).

RNA-seq libraries were sequenced on the DNBSEQ-T7 platform using a paired-end 150 bp strategy. Raw reads were filtered using SOAPnuke [90]. Clean reads were used for downstream analyses. Transcriptome de novo assembly was performed using Trinity (v2.15.1; https://github.com/trinityrnaseq/trinityrnaseq) [111] with default parameters. Additionally, clean data were aligned to the *Stylosanthes guianensis* genome using HISAT2 (v2.2.1; https://daehwankimlab.github.io/hisat2/) with the parameters "–phred33 –sensitive –dta -p 40", followed by reference-guided transcript assembly with StringTie (v2.2.1; https://github.com/gpertea/stringtie) [112] with default settings. ORFs were predicted from both de novo assembled transcripts and reference-guided transcripts using TransDecoder (v5.7.0; https://github.com/TransDecoder/TransDecoder) with default parameters.

Gene-level read counts were quantified from the genome-aligned RNA-seq data using featureCounts (v2.0.8; https://github.com/ShiLab-Bioinformatics/subread) [113] with default parameters. Raw read counts were used for differential expression analysis, while transcripts per million (TPM) values were calculated for expression visualization and co-expression analyses. Genes with TPM > 0 in at least one sample were considered expressed.

DEGs were identified using DESeq2 R package (v3.19) [114] based on raw read counts. Genes with an absolute $\log_2$ fold change ≥ 1 and a false discovery rate (FDR) < 0.05 were defined as significantly differentially expressed. KEGG pathway enrichment analyses of DEGs were performed in R using a hypergeometric test implemented with the *phyper* function. The background gene set consisted of all expressed genes included in the differential expression analysis, and 139 KEGG pathways were tested. *P* values were adjusted for multiple testing using the Benjamini–Hochberg procedure, and pathways with FDR < 0.05 were considered significantly enriched.

Gene expression patterns were visualized using heatmaps with the pheatmap package (v1.0.12; https://github.com/raivokolde/pheatmap). WGCNA was performed to identify gene modules associated with tissue specificity. DEGs identified from pairwise comparison among nodule, root+Br, and root−Br groups were used as input, and their expression profiles across an additional 11 *Stylosanthes guianensis* tissues were included to improve network robustness. Genes with detectable expression (TPM > 0) in at least one sample were retained. WGCNA was conducted following published methods, and module–tissue associations were evaluated using Pearson correlation analysis [115]. Modules

with correlation correlation ≥ 0.75 and p-value < 0.01 were considered significantly associated with specific tissues.

For full-length transcriptome analysis, RNA from above tissues was pooled to construct a PacBio Iso-Seq library. Single-molecule real-time sequencing was performed on the PacBio RSII sequencing platform with single-end sequencing by BGI-Shenzhen (Shenzhen, China). IsoSeq3 software (v0.7.2; https://github.com/ylipacbio/IsoSeq3) was adopted to analyze the full-length non-chimeric transcripts, with CCS parameter set to "–num-passes 1". ORFs were predicted using TransDecoder software.

Genome annotation

The de novo prediction of transposable elements (TEs) was performed using EDTA software (v2.2.2; https://github.com/oushujun/EDTA) [116], while tandem repeat sequences were annotated using TRF software (v4.09.1; https://github.com/Benson-Genomics-Lab/TRF). Known repeat sequences were annotated using RepeatMasker (v 4.1.1; https://github.com/Dfam-consortium/RepeatMasker) [117] using the Repbase database.

Protein-coding genes were predicted through a combination of ab initio, homology-based, and transcriptome-based approaches. Ab initio predictions were performed using GALBA (v1.0.8; https://github.com/Gaius-Augustus/GALBA) [118] and SNAP software (v2013-02–16; https://github.com/KorfLab/SNAP) with default parameters. Homology-based gene prediction was carried out using miniprot (v0.14; https://github.com/lh3/miniprot; parameters: "-t 20 -I –gff") [119], with amino acid sequences from *Glycine max*, *Medicago truncatula*, *Cajanus cajan*, *Stylosanthes angustifolia*, *Arabidopsis thaliana*, and the Swiss-Prot database (release 20,250,619). Open reading frames (ORFs) derived from RNA-seq and full-length transcripts were incorporated as transcript-based evidence. These predictions were integrated using EvidenceModeler (EVM) (v2.1.0; https://github.com/EVidenceModeler/EVidenceModeler) [120] to generate a non-redundant set of protein-coding genes. Finally, the gene set was further refined and untranslated regions (UTRs) were annotated using PASA software (PASApipeline v2.5.3; https://github.com/PASApipeline/PASApipeline) [121] based on transcriptome data.

Gene functional annotation was performed using eggNOG-mapper software (v2.1.9; https://github.com/eggnogdb/eggnog-mapper) based on the EggNOG v5.0 database [122, 123]. Additional annotation analyses were conducted using the NCBI NR, SwissProt, and KOG databases. Pfam domains were identified using InterProScan software (v5.66–98.0; https://github.com/ebi-pf-team/interproscan) with the Pfam database. GO annotation was derived based on the NR and EggNOG annotation results. Non-coding RNA (ncRNA) annotation was conducted following previously published methods [34].

Comparative genomic analysis

Gene family identification was conducted for *Stylosanthes guianensis* and 10 representative leguminous species, with *Arabidopsis thaliana* used as an outgroup. To ensure the reliability of comparative analyses, the completeness of the reference genomes and gene sets was first evaluated using BUSCO. The results showed that the BUSCO completeness of the genomes for the 10 selected species was greater than 99%, except for *Arachis ipaensis* (98.6%). For the gene sets, eight species exhibited BUSCO completeness greater

than 96%, while the remaining species were *Arachis ipaensis* (85.9%), *Arachis duranensis* (88.4%), and *Senna tora* (89.8%) (Additional file 1: Figure S22).

OGs were identified using OrthoFinder software with protein sequences from all plant species (parameters: -S diamond -M msa -A mafft). Based on single-copy orthologs, OrthoFinder software also inferred a species phylogenetic tree using FastTree software (v2.1.11) [124]. To estimate species divergence times, we obtained calibration points from TimeTree (http://www.timetree.org/) for the divergence times between *Glycine max* and *Cajanus cajan*, as well as between *Glycine max* and *Medicago truncatula*. Using these calibration points, we calculated the divergence times for the studied species based on the fourfold degenerate synonymous sites (4DTv) in single-copy orthologs, applying MCMCTree (4.9j; https://github.com/dosreislab/mcmctree) [125].

Furthermore, CAFÉ5 (v1.1; https://github.com/hahnlab/CAFE5) [126] was performed to analyze gene family expansion and contraction analysis. Gene families showing significant expansion in *Stylosanthes guianensis* were subjected to KEGG pathway enrichment analysis using a hypergeometric test. The background gene set consisted of all *Stylosanthes guianensis* genes assigned to orthologous groups, and a total of 131 KEGG pathways were tested. Multiple testing correction was applied using the Benjamini–Hochberg method (*p.adjust* (method = "BH")), and pathways with FDR < 0.05 were considered significantly enriched.

Intra- and inter-species collinearity analyses were performed among *Stylosanthes guianensis*, *Glycine max*, *Cajanus cajan*, and *Stylosanthes angustifolia*. Protein-coding genes were first aligned within and between species using DIAMOND (v2.1.1; https://github.com/bbuchfink/diamond) [127], and collinear gene blocks were subsequently identified with MCScanX (https://github.com/wyp1125/MCScanX) [128]. For intra-species collinearity analysis, we employed the 'duplicate_gene_classifier' tool implemented in MCScanX to classify gene expansion types, which include singleton, proximal, tandem, dispersed, segmental, and WGD. Nonsynonymous (Ka) and synonymous (Ks) substitution rates for collinear gene pairs were calculated using ParaAT (v2.0; https://github.com/wonaya/ParaAT) with the NG model.

Identification of PSR and SNF genes

Homology-based annotation of PSR and SNF genes in *Stylosanthes guianensis* was conducted using the PSR genes set of *Arabidopsis thaliana* (Additional file 2: Table S29) and the SNF reference gene set (Additional file 2: Table S22) constructed from four legumes (*Medicago truncatula*, *Lotus japonicus*, *Glycine max*, and *Phaseolus vulgaris*), respectively. The amino acid sequences of the genes from *Stylosanthes guianensis* were aligned with the PSR and SNF gene sets using DIAMOND software [127]. Genes with identity and coverage both ≥ 50% were selected as homologous genes.

The expansion of *CHR* in *Stylosanthes guianensis* was systematically analyzed. All CHR protein sequences from *Stylosanthes guianensis* were aligned to the *Glycine max* (T2T genome) and *Medicago truncatula* genomes using miniprot for gene prediction. Based on these predictions, each *Stylosanthes guianensis* CHR copy was manually curated to identify its corresponding loci in the *Glycine max* and *Medicago truncatula* genomes. The syntenic genomic positions were subsequently cataloged. Microsynteny analysis of the CHR-expanded loci was performed using the JCVI toolkit [129].

Phylogenetic analysis and gene expression clustering

For gene phylogenetic tree construction, multiple sequence alignment of the amino acid sequences was performed using MUSCLE (v5.3; https://github.com/rcedgar/muscle/tree/v5.3) [106], followed by alignment trimming with trimAl (v1.5.0; https://github.com/inab/trimal) [107]. A maximum likelihood phylogenetic tree was then generated using IQ-TREE (v2.3.6; https://github.com/iqtree/iqtree2) [108] with 1,000 iterations. Finally, the phylogenetic tree was visualized and annotated using iTOL (v7; https://itol.embl.de/) [130]. Additionally, differentially expressed PSR genes were visualized through a clustered heatmap generated with ClusterGVis (v0.1.2; https://github.com/junjunlab/ClusterGVis), where gene expression patterns were clustered using the mfuzz algorithm.

Metabolomic analysis and quantitative detection of central carbon metabolites

Widely-targeted flavonoid metabolomics was conducted as previously described [131], whereas untargeted metabolomics was performed as previously reported [132]. Differentially accumulated metabolites (DAMs) were identified by pairwise comparisons between groups using the following criteria: $|\log_2$ fold change (FC)$| \geq 1$, variable importance in projection (VIP) ≥ 1, and a Student's *t*-test *P* value < 0.05 [133]. Central carbon metabolites (ATP, PEP, 3-PGA, G6P, DHAP, F6P, F-1,6-P_2, citrate, pyruvate, and malate) were quantitatively determined using a high-performance ion chromatography–tandem mass spectrometry (HPIC–MS/MS) system in conjunction with corresponding metabolite standard solutions, following a previously reported method [134]. The metabolomic analyses and metabolite quantification described above were conducted by Biotree Biotech Co., Ltd. (Shanghai, China).

Activity analysis of MDH, PEPC, SUS, and pyrophosphatase

Approximately 0.15 g fresh samples were extracted in 1.5 mL of 100 mM Tris–HCl (pH 7.5) at 4 °C. After centrifugation at 12,000 rpm for 15 min at 4 °C, the supernatant was collected to determine MDH and PEPC activities. For MDH activity determination, 0.3 mL of supernatant was incubated in 100 mM Tris–HCl (pH 8.0) buffer containing 5 mM ethylenediaminetetraacetic acid, 0.2 mM nicotinamide adenine dinucleotide (NADH) and 1 mM oxaloacetate (OAA). The reaction was started by enzyme addition. MDH activity was detected spectrophotometrically at 340 nm by monitoring NADH oxidation within 2 min using a spectrophotometer (UV-2010, Hitachi, Japan) and is expressed as the amount of enzyme catalyzing the oxidation of 1 μmol NADH (extinction coefficient 6.22 mM^{-1} cm^{-1}) per minute. For PEPC activity determination, 0.4 mL of supernatant was incubated in 50 mM Tris–HCl (pH 8.5) buffer containing 10 mM $NaHCO_3$, 2 mM $MgCl_2$, 2 units of MDH, 0.2 mM NADH and 2 mM PEP. PEPC activity was measured at 340 nm by monitoring NADH oxidation within 2 min and is expressed as the amount of enzyme catalyzing the oxidation of 1 μmol NADH. Sucrose synthase activity was measured in the cleavage direction using a commercially available assay kit (Product Code: ADS-W-ZT011, Jiangsu Addison Biotechnology, China), as previously described [135]. Pyrophosphatase activity was measured according to established protocols [136]. Three biological replicates were included for all experiments.

Statistical analysis

Statistical analyses, including Student's *t*-test, were performed using R (v4.5.0). Box plots were generated in R using the ggplot2 package.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-026-04130-x.

Additional file 1.

Additional file 2.

Acknowledgements
We thank Drs. Xiaohui Mo, Jifu Li, and Xu Hu for discussions and comments on the experiments.

Peer review information
Sophie de Vries and Wenjing She were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions
Z.C. and P.L. planned the project and designed the study. P.L. performed field experiments. C.L., P.L., and Z.C. performed data analysis. R.D. and G.L. collected *Bradyrhizobium* strains and provided *S. guianensis* germplasm. W.P. provided big data server resources and contributed to part of the data analysis. J.Z., R.X., L.L. (Liting Liu), and J.L. collected samples. R.H., L.J., H.H., and L.L. (Lijuan Luo) provided technical support and suggestions. P.L., Z.C., and C.L. wrote the manuscript. All authors read and approved the final manuscript.

Funding
The research was financially supported by the National Natural Science Foundation of China (32441034, 32471766, 32371769), the National Key R&D Program of China (2024YFD1301203), the Natural Science Foundation of Hainan Province (323CXTD387), the Earmarked fund for CARS-Forage and Grass (CARS-34), the Central Public-interest Scientific Institution Basal Research Fund for CATAS (1630032025014, 1630032024016, 1630032022011), and CATAS for Science and Technology Innovation Team of National Tropical Agricultural Science Center (CATASCXTD202415).

Data availability
The genome and transcriptome sequencing data of *S. guianensis,* along with its genome assembly and annotation, have been deposited in the National Genomics Data Center (https://ngdc.cncb.ac.cn/) under the accession number PRJCA037887 [137]. The custom scripts developed and employed in this study have been deposited in publicly accessible repositories on GitHub (https://github.com/LiuChun-Lab/Sguianensis_T2T) [138] and Zenodo (https://doi.org/10.5281/zenodo.18384843) [139]. The genome assembly and the corresponding genome annotation are available at Figshare (https://doi.org/10.6084/m9.figshare.32231826) [140]. The widely-targeted flavonoid metabolomics data have been deposited in the MetaboLights [141] repository under the study identifier MTBLS14548 (https://www.ebi.ac.uk/metabolights/MTBLS14548) [142]. The untargeted metabolomics data have been deposited in the MetaboLights [141] repository under the study identifier MTBLS14552 (https://www.ebi.ac.uk/metabolights/MTBLS14552) [143].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 30 March 2025 Accepted: 26 May 2026
Published online: 15 June 2026

References

1. Udvardi M, Poole PS. Transport and metabolism in legume-rhizobia symbioses. Annu Rev Plant Biol. 2013;64:781–805.
2. Roy S, Liu W, Nandety RS, Crook A, Mysore KS, Pislariu CI, et al. Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. Plant Cell. 2020;32(1):15–41.
3. Herridge DF, Peoples MB, Boddey RM. Global inputs of biological nitrogen fixation in agricultural systems. Plant Soil. 2008;311:1–18.
4. Kebede E. Contribution, utilization, and improvement of legumes-driven biological nitrogen fixation in agricultural systems. Front Sustain Food Syst. 2021;5:767998.

5. Yang J, Lan L, Jin Y, Yu N, Wang D, Wang E. Mechanisms underlying legume-rhizobium symbioses. J Integr Plant Biol. 2022;64(2):244–67.
6. Liu T, Liu H, Xian W, Liu Z, Yuan Y, Fan J, et al. Duplication and sub-functionalization of flavonoid biosynthesis genes plays important role in Leguminosae root nodule symbiosis evolution. J Integr Plant Biol. 2024;66(10):2191–207.
7. Hawkins JP, Oresnik IJ. The *rhizobium*-legume symbiosis: co-opting successful stress management. Front Plant Sci. 2022;12:796045.
8. Yeremko L, Czopek K, Staniak M, Marenych M, Hanhur V. Role of environmental factors in legume-*rhizobium* symbiosis: a review. Biomolecules. 2025;15(1):118.
9. Guo JH, Liu XJ, Zhang Y, Shen JL, Han WX, Zhang WF, et al. Significant acidification in major Chinese croplands. Science. 2010;327(5968):1008–10.
10. Ao JH, Chen ZJ, Wu M, Lu X, Huang ZR, Liao H. Phosphorus fractions of red soils in Guangdong province of South China and their bioavailability for five crop species. Soil Sci. 2014;179:514–21.
11. Li Y, Wang C, Zheng L, Ma W, Li M, Guo Z, et al. Natural variation of *GmRj2/Rfg1* determines symbiont differentiation in soybean. Curr Biol. 2023;33(12):2478–90.
12. Lin MH, Gresshoff PM, Ferguson BJ. Systemic regulation of soybean nodulation by acidic growth conditions. Plant Physiol. 2012;160(4):2028–39.
13. Kochian LV, Piñeros MA, Liu J, Magalhaes JV. Plant adaptation to acid soils: the molecular basis for crop aluminum resistance. Annu Rev Plant Biol. 2015;66:571–98.
14. Zhu XF, Shen RF. Towards sustainable use of acidic soils: deciphering aluminum-resistant mechanisms in plants. Fundam Res. 2023;4(6):1533–41.
15. Chaudhary MI, Adu-Gyamfi JJ, Saneoka H, Nguyen NT, Suwa R, Kanai S, et al. The effect of phosphorus deficiency on nutrient uptake, nitrogen fixation and photosynthetic rate in mashbean, mungbean and soybean. Acta Physiol Plant. 2008;30:537–44.
16. Hernández G, Valdés-López O, Ramírez M, Goffard N, Weiller G, Aparicio-Fabre R, et al. Global changes in the transcript and metabolic profiles during symbiotic nitrogen fixation in phosphorus-stressed common bean plants. Plant Physiol. 2009;151:1221–38.
17. Buoso S, Zamboni A, Franco A, Commisso M, Guzzo F, Varanini Z, et al. Nodulating white lupins take advantage of the reciprocal interplay between N and P nutritional responses. Physiol Plant. 2022;174(1):e13607.
18. Singh J, Mendoza-Soto AB, Tiwari M, Acevedo-Sandoval TT, Formey D, Ané JM, et al. Phosphate deficiency reduces nodule formation through a phosphate starvation response-like protein in *Phaseolus vulgaris*. Plant Cell Physiol. 2025;66(12):1794–810.
19. Cheng F, Cao G, Wang X, Zhao J, Yan X, Liao H. Isolation and application of effective nitrogen fixation rhizobial strains on low-phosphorus acid soils in south China. Chin Sci Bull. 2009;54(3):412–20.
20. Sulieman S, Tran LSP. Phosphorus homeostasis in legume nodules as an adaptive strategy to phosphorus deficiency. Plant Sci. 2015;239:36–43.
21. Vardien W, Steenkamp ET, Valentine AJ. Legume nodules from nutrient-poor soils exhibit high plasticity of cellular phosphorus recycling and conservation during variable phosphorus supply. J Plant Physiol. 2016;191:73–81.
22. Chen ZJ, Wang LJ, Cardoso JA, Zhu SN, Liu GD, Rao IM, et al. Improving phosphorus acquisition efficiency through modification of root growth responses to phosphate starvation in legumes. Front Plant Sci. 2023;14:1094157.
23. Qin L, Jiang H, Tian J, Zhao J, Liao H. *Rhizobia* enhance acquisition of phosphorus from different sources by soybean plants. Plant Soil. 2011;349:25–36.
24. Isidra-Arellano MC, Valdés-López O. Understanding the crucial role of phosphate and iron availability in regulating root nodule symbiosis. Plant Cell Physiol. 2024;65(12):1925–36.
25. Qin L, Guo Y, Chen L, Liang R, Gu M, Xu G, et al. Functional characterization of 14 *Pht1* family genes in yeast and their expressions in response to nutrient starvation in soybean. PLoS ONE. 2012;7(10):e47726.
26. Qin L, Zhao J, Tian J, Chen L, Sun Z, Guo Y, et al. The high-affinity phosphate transporter GmPT5 regulates phosphate transport to nodules and nodulation in soybean. Plant Physiol. 2012;159(4):1634–43.
27. Li X, Zhao J, Tan Z, Zeng R, Liao H. GmEXPB2, a cell wall β-expansin, affects soybean nodulation through modifying root architecture and promoting nodule formation and development. Plant Physiol. 2015;169(4):2640–53.
28. Wang JY, Si ZY, Li F, Xiong XB, Lei L, Xie FL, et al. A purple acid phosphatase plays a role in nodule formation and nitrogen fixation in *Astragalus sinicus*. Plant Mol Biol. 2015;88:515–29.
29. Lu M, Cheng Z, Zhang XM, Huang P, Fan C, Yu G, et al. Spatial divergence of PHR-PHT1 modules maintains phosphorus homeostasis in soybean nodules. Plant Physiol. 2020;184(1):236–50.
30. Müller LM. PHO1 proteins mediate phosphate transport in the legume rhizobium symbiosis. Plant Physiol. 2021;185:26–8.
31. Zhang JY, Zhao XK, Dong RS, Liu GD, Luo LJ, Liu PD. A rapid hairy root transformation system for characterizing haloacid dehalogenase in aluminum toxicity resistance in *Stylosanthes*. Plant Physiol Biochem. 2025;229(Pt C):110563.
32. Marques A, Moraes L, dos Santos MA, Costa I, Costa L, Nunes T, et al. Origin and parental genome characterization of the allotetraploid *Stylosanthes scabra* Vogel (Papilionoideae, Leguminosae), an important legume pasture crop. Ann Bot. 2018;122:1143–59.
33. Schultze-Kraft R, Yang H, Tang J, Liu G. *Stylosanthes guianensis* CIAT 184-review of a tropical forage legume. Trop Grassl-Forrajes Trop. 2023;11(2):95–120.
34. Liu C, Zhang J, Xu R, Lv J, Qiao Z, Bai M, et al. A chromosome-scale genome assembly of the pioneer plant *Stylosanthes angustifolia*: insights into genome evolution and drought adaptation. Gigascience. 2025;14:giae118.
35. Velásquez Ramírez MG, del Castillo TD, Guerrero Barrantes JA, Vásquez Bardales J, Thomas E, Cusi Auca E, et al. Soil recovery of alluvial gold mine spoils in the Peruvian Amazon using *Stylosanthes guianensis*, a promising cover crop. Land Degrad Dev. 2021;32(18):5143–53.
36. Zhang J, Lv J, Liu C, Xu R, Liu Y, Huang R, et al. SgPAP27b, a purple acid phosphatase, confers aluminum tolerance in *Stylosanthes guianensis* through membrane lipid remodeling. J Hazard Mater. 2026;507:141716.

37. Lin Y, Liu G, Liu P, Chen Q, Guo X, Lu X, et al. Border-like cell formation mediated by SgPG1 confers aluminum resistance in *Stylosanthes guianensis*. Plant J. 2024;120:1605–24.
38. Chen ZJ, Sun LL, Liu PD, Liu GD, Tian J, Liao H. Malate synthesis and secretion mediated by a manganese-enhanced malate dehydrogenase confers superior manganese tolerance in *Stylosanthes guianensis*. Plant Physiol. 2015;167:176–88.
39. Luo J, Chen Z, Huang R, Wu YH, Liu C, Cai ZP, et al. Multi-omics analysis reveals the roles of purple acid phosphatases in organic phosphorus utilization by the tropical legume *Stylosanthes guianensis*. Plant J. 2024;117(3):729–46.
40. Wang LJ, Li JF, Liu LT, Dong RS, Liu GD, Rao IM, et al. Phenylalanine ammonia-lyase 2 regulates secondary metabolism and confers manganese tolerance in *Stylosanthes guianensis*. Plant Physiol. 2025;197(1):1–18.
41. Huang Y, Guan E, Song S, Koo DH, Schmidt MA, Su H, et al. Genetic diversity and architectural dynamics of soybean centromeres. Genome Biol. 2026;27(1):17.
42. Wang Y, Hao X, Chen C, Wang H, Gao P, Yang X, et al. Telomere-to-telomere genome of common bean (*Phaseolus vulgaris* L., YP4). Gigascience. 2025;14:giaf001.
43. Shen L, Yi C, Liu Y, Han F, Feng J. Two complete telomere-to-telomere genome assemblies of *Medicago* reveal the landscape and evolution of its centromeres. Mol Plant. 2025;18(9):1409–12.
44. Ma Y, Xiao C, Liu J, Ren G. Nutrient-dependent regulation of symbiotic nitrogen fixation in legumes. Hortic Res. 2024;12(3):uhae321.
45. Gresshoff PM, Su C, Su H, Hastwell A, Cha Y, Zhang M, et al. Functional genomics dissection of the nodulation autoregulation pathway (AON) in soybean (*Glycine max*). J Integr Plant Biol. 2025;67(3):762–72.
46. Lin JS, Li X, Luo Z, Mysore KS, Wen J, Xie F. NIN interacts with NLPs to mediate nitrate inhibition of nodulation in *Medicago truncatula*. Nat Plants. 2018;4(11):942–52.
47. Feng J, Lee T, Schiessl K, Oldroyd GED. Processing of Nodule Inception controls the transition to nitrogen fixation in root nodules. Science. 2021;374(6567):629–32.
48. Fu M, Sun JF, Li X, Guan Y, Xie F. Asymmetric redundancy of soybean *Nodule Inception* (*NIN*) genes in root nodule symbiosis. Plant Physiol. 2022;188(1):477–89.
49. Cathebras C, Gong X, Andrade RE, Vondenhoff K, Keller J, Delaux PM, et al. A novel cis-element enabled bacterial uptake by plant cells. Nat Plants. 2026;12(1):140–51.
50. Wang X, Chen K, Zhou M, Gao Y, Huang H, Liu C, et al. GmNAC181 promotes symbiotic nodulation and salt tolerance of nodulation by directly regulating *GmNINa* expression in soybean. New Phytol. 2022;236(2):656–70.
51. Lan P, Li WF, Schmidt W. 'Omics' approaches towards understanding plant phosphorus acquisition and use. Annu Plant Rev. 2015;48:65–98.
52. Xu W, Zhang Q, Yuan W, Xu F, Muhammad Aslam M, Miao R, et al. The genome evolution and low-phosphorus adaptation in white lupin. Nat Commun. 2020;11(1):1069.
53. Yang SY, Lin WY, Hsiao YM, Chiou TJ. Milestones in understanding transport, sensing, and signaling of the plant nutrient phosphorus. Plant Cell. 2024;36(5):1504–23.
54. Gutiérrez-Luna FM, Hernández-Domínguez EE, Valencia-Turcotte LG, Rodríguez-Sotres R. Review: "Pyrophosphate and pyrophosphatases in plants, their involvement in stress responses and their possible relationship to secondary metabolism. Plant Sci. 2018;267:11–9.
55. Huang YX, Dong RS, Bai CJ, Liu GD, Liao H. Acid resistance and salt tolerance of *Stylosanthes* rhizobia strains. Chin J Trop Crops. 2010;31(6):914–9.
56. Zhang J, Dong RS, Huang YX, Wang WQ, Tang J, Liu GD, et al. The impact of rhizobium bacteria inoculation on the growth of *Stylosanthes guianensis* cv. Reyan No.5 under different pH. Chin J Trop Crops. 2012;33(4):609–612.
57. Dong RS, Zhang J, Huang YX, Bai CJ, Huan HF, Wang WQ, et al. Screen of efficient nitrogen-fixing rhizobium strains matching with *Stylosanthes guianensis* cv. Reyan II Chin J Trop Crops. 2013;34(7):1247–51.
58. Dong RS, Zhang J, Huang YX, Bai CJ, Huan HF, Wang WQ, et al. Optimum P concentration promoting the symbiotic relationship between rhizobia strains and *Stylosanthes guianensis* Reyan No.2. Chin J Trop Crops. 2014;35(5):832–836.
59. Chen ZJ, Cui QQ, Liang CY, Sun LL, Tian J, Liao H. Identification of differentially expressed proteins in soybean nodules under phosphorus deficiency through proteomic analysis. Proteomics. 2011;11(24):4648–59.
60. Dong W, Song Y. The significance of flavonoids in the process of biological nitrogen fixation. Int J Mol Sci. 2020;21(16):5926.
61. Li Y, Lu L, Wang Q, Liu X, Tian J, Zhang R, et al. Arbuscular mycorrhizal fungi promote nodulation and N_2 fixation in soybean by specific root exudates. Plant Cell Environ. 2025;48(7):5514–28.
62. Ruan Y, Dong S, Jiang S, Wang Y, Wu X, Zhuang Y, et al. The molecular basis of the binding and specific activation of rhizobial NodD by flavonoids. Science. 2026;391(6781):184–9.
63. Mameda R, Waki T, Kawai Y, Takahashi S, Nakayama T. Involvement of chalcone reductase in the soybean isoflavone metabolon: identification of GmCHR5, which interacts with 2-hydroxyisoflavanone synthase. Plant J. 2018;96(1):56–74.
64. Gao JP, Liang W, Jiang S, Yan Z, Zhou C, Wang E, et al. NODULE INCEPTION activates gibberellin biosynthesis genes during rhizobial infection. New Phytol. 2023;239(2):459–65.
65. Liu CW, Breakspear A, Guan D, Cerri MR, Jackson K, Jiang S, et al. NIN acts as a network hub controlling a growth module required for rhizobial infection. Plant Physiol. 2019;179(4):1704–22.
66. Liu T, Liu Z, Fan J, Yuan Y, Liu H, Xian W, et al. Loss of *Lateral suppressor* gene is associated with evolution of root nodule symbiosis in Leguminosae. Genome Biol. 2024;25(1):250.
67. Isidra-Arellano MC, Pozas-Rodríguez EA, Del Rocío R-S, Arroyo-Canales J, Ferrer-Orgaz S, Del Socorro S-C, et al. Inhibition of legume nodulation by Pi deficiency is dependent on the autoregulation of nodulation (AON) pathway. Plant J. 2020;103(3):1125–39.
68. Wang P, Jiang F, Xue Z, Bu F, Zhu W, Zhang Y, et al. The *Medicago* SPX1/3-PHR2 network relays phosphate signaling to orchestrate root nodulation-dependent nitrogen acquisition by controlling flavonoid biosynthesis. Plant Commun. 2026. https://doi.org/10.1016/j.xplc.2026.101695.

69. Li Y, Ma W, Zhang K, Wang X, Liu R, Tian Y, et al. Overexpression of *GmPHR1* promotes soybean yield through global regulation of nutrient acquisition and root development. Int J Mol Sci. 2022;23(23):15274.
70. Zhou Y, Olt P, Neuhäuser B, Moradtalab N, Bautista W, Uhde-Stone C, et al. Loss of LaMATE impairs isoflavonoid release from cluster roots of phosphorus-deficient white lupin. Physiol Plant. 2021;173(3):1207–20.
71. Fan S, Huang K, Yang Y, Xi Y, Ye H, Xu L, et al. OsMATE7-mediated flavonol accumulation regulates pollen tube growth in rice. Plant J. 2025;123(4):e70449.
72. Xu M, Zuo H, He M, Yang Y, Zhang L, Zhai X, et al. Functional analysis of *GmMATE* gene family in soybean phosphorus homeostasis and abiotic stress resilience. Plant Sci. 2025;359:112622.
73. Chen L, Qin L, Zhou L, Li X, Chen Z, Sun L, et al. A nodule-localized phosphate transporter GmPT7 plays an important role in enhancing symbiotic N_2 fixation and yield in soybean. New Phytol. 2019;221(4):2013–25.
74. Nguyen NNT, Clua J, Vetal PV, Vuarambon DJ, De Bellis D, Pervent M, et al. PHO1 family members transport phosphate from infected nodule cells to bacteroids in *Medicago truncatula*. Plant Physiol. 2021;185(1):196–209.
75. Liu J, Yang R, Yan J, Li C, Lin X, Lin L, et al. *VPT*-like genes modulate *Rhizobium*-legume symbiosis and phosphorus adaptation. Plant J. 2023;116(1):112–27.
76. Dissanayaka DMSB, Ghahremani M, Siebers M, Wasaki J, Plaxton WC. Recent insights into the metabolic adaptations of phosphorus-deprived plants. J Exp Bot. 2021;72(2):199–223.
77. Yang B, Tan Z, Yan J, Zhang K, Ouyang Z, Fan R, et al. Phospholipase-mediated phosphate recycling during plant leaf senescence. Genome Biol. 2024;25:199.
78. Sulieman S, Abdelrahman M, Tran LSP. Carbon metabolic adjustment in soybean nodules in response to phosphate limitation: a metabolite perspective. Environ Exp Bot. 2022;196:104810.
79. Segami S, Tomoyama T, Sakamoto S, Gunji S, Fukuda M, Kinoshita S, et al. Vacuolar H^+-pyrophosphatase and cytosolic soluble pyrophosphatases cooperatively regulate pyrophosphate levels in *Arabidopsis thaliana*. Plant Cell. 2018;30(5):1040–61.
80. Laurent F, Bartsch SM, Shukla A, Rico-Resendiz F, Couto D, Fuchs C, et al. Inositol pyrophosphate catabolism by three families of phosphatases regulates plant growth and development. PLoS Genet. 2024;20(11):e1011468.
81. Titiz O, Tambasco-Studart M, Warzych E, Apel K, Amrhein N, Laloi C, et al. PDX1 is essential for vitamin B6 biosynthesis, development and stress tolerance in Arabidopsis. Plant J. 2006;48(6):933–46.
82. Colinas M, Eisenhut M, Tohge T, Pesquera M, Fernie AR, Weber AP, et al. Balancing of B6 vitamers is essential for plant development and metabolism in Arabidopsis. Plant Cell. 2016;28(2):439–53.
83. Liu Y, Maniero RA, Giehl RFH, Melzer M, Steensma P, Krouk G, et al. PDX1.1-dependent biosynthesis of vitamin B6 protects roots from ammonium-induced oxidative stress. Mol Plant. 2022;15(5):820–39.
84. Matamoros MA, Saiz A, Peñuelas M, Bustos-Sanmamed P, Mulet JM, Barja MV, et al. Function of glutathione peroxidases in legume root nodules. J Exp Bot. 2015;66(10):2979–90.
85. Zhao Y, Nickels LM, Wang H, Ling J, Zhong Z, Zhu J. OxyR-regulated catalase activity is critical for oxidative stress resistance, nodulation and nitrogen fixation in *Azorhizobium caulinodans*. FEMS Microbiol Lett. 2016;363(13):fnw130.
86. Smith PM, Atkins CA. Purine biosynthesis. Big in cell division, even bigger in nitrogen assimilation. Plant Physiol. 2002;128(3):793–802.
87. Izaguirre-Mayoral ML, Lazarovits G, Baral B. Ureide metabolism in plant-associated bacteria: purine plant-bacteria interactive scenarios under nitrogen deficiency. Plant Soil. 2018;428:1–34.
88. Pauly N, Pucciariello C, Mandon K, Innocenti G, Jamet A, Baudouin E, et al. Reactive oxygen and nitrogen species and glutathione: key players in the legume-*Rhizobium* symbiosis. J Exp Bot. 2006;57(8):1769–76.
89. Nasr Esfahani M, Kusano M, Nguyen KH, Watanabe Y, Ha CV, Saito K, et al. Adaptation of the symbiotic *Mesorhizobium*-chickpea relationship to phosphate deficiency relies on reprogramming of whole-plant metabolism. Proc Natl Acad Sci U S A. 2016;113(32):E4610–9.
90. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience. 2018;7(1):1–6.
91. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics. 2023;39(5):btad311.
92. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18(2):170–5.
93. Cheng H, Asri M, Lucas J, Koren S, Li H. Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. Nat Methods. 2024;21(6):967–70.
94. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016;3(1):95–8.
95. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science. 2017;356(6333):92–5.
96. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100.
97. Jain C, Rhie A, Hansen NF, Koren S, Phillippy AM. Long-read mapping to repetitive reference sequences using Winnowmap2. Nat Methods. 2022;19(6):705–10.
98. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008.
99. Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, et al. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. Genome Biol. 2024;25(1):107.
100. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. PLoS Comput Biol. 2018;14(1):e1005944.
101. Zeng X, Yi Z, Zhang X, Du Y, Li Y, Zhou Z, et al. Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. Nat Plants. 2024;10(8):1184–200.
102. Xu D, Yang J, Wen H, Feng W, Zhang X, Hui X, et al. CentIER: accurate centromere identification for plant genomes. Plant Commun. 2024;5(10):101046.

103. Tegenfeldt F, Kuznetsov D, Manni M, Berkeley M, Zdobnov EM, Kriventseva EV. OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. Nucleic Acids Res. 2025;53(D1):D516–22.
104. Ou S, Jiang N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. 2018;176(2):1410–22.
105. Chen Q, Yang C, Zhang G, Wu D. GCI: a continuity inspector for complete genome assembly. Bioinformatics. 2024;40(11):btae633.
106. Edgar RC. Muscle5: high-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. Nat Commun. 2022;13(1):6968.
107. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.
108. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–4.
109. Zhu S, Chen M, Liang C, Xue Y, Lin S, Tian J. Characterization of purple acid phosphatase family and functional analysis of *GmPAP7a/7b* involved in extracellular ATP utilization in soybean. Front Plant Sci. 2020;11:661.
110. Wang Y, Yang Z, Kong Y, Li X, Li W, Du H, et al. *GmPAP12* is required for nodule development and nitrogen fixation under phosphorus starvation in soybean. Front Plant Sci. 2020;11:450.
111. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–52.
112. Shumate A, Wong B, Pertea G, Pertea M. Improved transcriptome assembly using a hybrid of long and short reads with StringTie. PLoS Comput Biol. 2022;18(6):e1009730.
113. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013;41(10):e108.
114. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
115. Tai Y, Ling C, Wang H, Yang L, She G, Wang C, et al. Comparative transcriptomic analysis reveals regulatory mechanisms of theanine synthesis in tea (*Camellia sinensis*) and oil tea (*Camellia oleifera*) plants. J Agric Food Chem. 2019;67(36):10235–44.
116. Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 2019;20(1):275.
117. Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2004;Chapter 4:Unit 4.10.
118. Brůna T, Li H, Guhlin J, Honsel D, Herbold S, Stanke M, et al. Galba: genome annotation with miniprot and AUGUSTUS. BMC Bioinformatics. 2023;24(1):327.
119. Li H. Protein-to-genome alignment with miniprot. Bioinformatics. 2023;39(1):btad014.
120. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9(1):R7.
121. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003;31(19):5654–66.
122. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. EggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38(12):5825–9.
123. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47(D1):D309–14.
124. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5(3):e9490.
125. Yang Z, Rannala B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol Biol Evol. 2006;23(1):212–26.
126. Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics. 2021;36(22–23):5516–8.
127. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59–60.
128. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.
129. Tang H, Krishnakumar V, Zeng X, Xu Z, Taranto A, Lomas JS, et al. JCVI: a versatile toolkit for comparative genomics analysis. iMeta. 2024;3(4):e211.
130. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52(W1):W78–82.
131. Shen Y, Huang Q, Yin T, Zhang H, Wu Z, Wei G, et al. An isoflavone from *Medicago truncatula*, methylnissolin, induces expression of *Sinorhizobium meliloti* nodulation genes. Plant Soil. 2025;515:1863–81.
132. Liu C, Huang R, Zhao X, Xu R, Zhang J, Li X, et al. Comparative analysis of lipid and flavonoid biosynthesis between Pongamia and soybean seeds: genomic, transcriptional, and metabolic perspectives. Biotechnol Biofuels Bioprod. 2024;17(1):86.
133. Lv J, Wu Y, Huang R, Xu R, Zhang J, Liu Y, et al. Combined analysis of the leaf metabolome, lipidome, and candidate gene function: insights into genotypic variation in phosphorus utilization efficiency in *Stylosanthes guianensis*. J Agric Food Chem. 2025;73(4):2653–68.
134. Chen J, Li Z, Yu H, Cui H, Li X. Affecting of glyphosate tolerance and metabolite content in transgenic *Arabidopsis thaliana* overexpressing EPSPS gene from *Eleusine indica*. Plants (Basel). 2024;14(1):78.
135. Xiao N, Ma H, Wang W, Sun Z, Li P, Xia T. Overexpression of ZmSUS1 increased drought resistance of maize (*Zea mays* L.) by regulating sucrose metabolism and soluble sugar content. Planta. 2024;259(2):43.
136. Jelitto T, Sonnewald U, Willmitzer L, Hajirezeai M, Stitt M. Inorganic pyrophosphate content and metabolites in potato and tobacco plants expressing *E. coli* pyrophosphatase in their cytosol. Planta. 1992;188(2):238–44.

137. Liu P, Liu C, Pu W, Zhang J, Xu R, Liu L, Luo J, Huang R, Jiang L, Huan H, Luo L, Liu G, Dong R, Chen Z. Telomere-to-telomere genome of *Stylosanthes guianensis* uncovers symbiotic adaptation to phosphorus-deficient soils. NGDC Genome Sequence Archive. 2026. https://ngdc.cncb.ac.cn/gsa/search?searchTerm=PRJCA037887.
138. Liu P, Liu C, Pu W, Zhang J, Xu R, Liu L, Luo J, Huang R, Jiang L, Huan H, Luo L, Liu G, Dong R, Chen Z. Telomere-to-telomere genome of *Stylosanthes guianensis* uncovers symbiotic adaptation to phosphorus-deficient soils. Github. 2026. https://github.com/LiuChun-Lab/Sguianensis_T2T.
139. Liu P, Liu C, Pu W, Zhang J, Xu R, Liu L, et al. Telomere-to-telomere genome of *Stylosanthes guianensis* uncovers symbiotic adaptation to phosphorus-deficient soils. 2026. Zenodo. https://doi.org/10.5281/zenodo.18384843.
140. Liu P, Liu C, Pu W, Zhang J, Xu R, Liu L, et al. Telomere-to-telomere genome of *Stylosanthes guianensis* uncovers symbiotic adaptation to phosphorus-deficient soils. 2026. FigShare. https://doi.org/10.6084/m9.figshare.32231826.
141. Yurekten O, Payne T, Tejera N, Amaladoss FX, Martin C, Williams M, et al. MetaboLights: open data repository for metabolomics. Nucleic Acids Res. 2024;52(D1):D640–6.
142. Liu P, Liu C, Pu W, Zhang J, Xu R, Liu L, Luo J, Huang R, Jiang L, Huan H, Luo L, Liu G, Dong R, Chen Z. Telomere-to-telomere genome of *Stylosanthes guianensis* uncovers symbiotic adaptation to phosphorus-deficient soils. MetaboLights. 2026. https://www.ebi.ac.uk/metabolights/MTBLS14548.
143. Liu P, Liu C, Pu W, Zhang J, Xu R, Liu L, Luo J, Huang R, Jiang L, Huan H, Luo L, Liu G, Dong R, Chen Z. Telomere-to-telomere genome of *Stylosanthes guianensis* uncovers symbiotic adaptation to phosphorus-deficient soils. MetaboLights. 2026. https://www.ebi.ac.uk/metabolights/MTBLS14552.

Publisher's Note