

The histone modification H3K4me3 marks functional genes in soybean nodules



Qianwen Wang¹, Wai-Shing Yung¹, Zhili Wang, Hon-Ming Lam*

Center for Soybean Research of the State Key Laboratory of Agrobiotechnology and School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

ARTICLE INFO

Keywords:
H3K4me3
Epigenetic regulation
Histone modification
Transcriptome
Mature nodules
Nodule function
Soybean

ABSTRACT

Nitrogen fixation in legumes requires the development of specialized organs called root nodules. Here we characterized the high-confidence transcriptome and genome-wide patterns of H3K4me3 marks in soybean roots and mature nodules symbiotic with *Sinorhizobium fredii*. Changes in H3K4me3 levels were positively associated with the transcription levels of functional genes in the nodules. The up-regulation of H3K4me3 levels was not only present in leghaemoglobin and nodulin-related genes, but also in genes involved in nitrogen and carbon metabolic pathways. In addition, genes regulating the transmembrane transport of metal ions, phosphates, sulphates, peptides, and sugars were differentially modified. On the contrary, a loss of H3K4me3 marks was found in several key transcription factor genes and was correlated with the down-regulation of the defense-related network in nodules, which could contribute to nodule maintenance. All these findings demonstrate massive reprogramming of gene expressions via alterations in H3K4me3 levels in the genes in mature soybean nodules.

1. Introduction

Soybean (*Glycine max* [L.] Merr.) is an important source of dietary proteins. Like other legumes, soybean could assimilate atmospheric nitrogen into the organic form through symbiotic rhizobia forming bacteroids in root nodules [1]. Establishment of symbiosis is beneficial to both species, through which soybean provides carbon sources in exchange for ammonia converted from atmospheric nitrogen by the microbial partner. Inside the nodules, ammonia from bacteroids is transported across the peribacteroid membrane to the surrounding infected plant cells, where it is assimilated into glutamine by glutamine synthetase [2]. Glutamine is then incorporated into the *de-novo* purine biosynthesis pathway to produce urate. In neighbouring uninfected cells, urate is further oxidized into ureides, which are then translocated to other parts of the plant via xylems [3]. Meanwhile, sucrose produced from photosynthesis in the shoot is converted into dicarboxylates to provide energy for bacteroids [3].

In addition to the continuous exchange of nitrogen and carbon resources between bacteroids and plants, the transport of other ions is

also essential for nodule maintenance and functions [4]. Phosphate transport is essential for providing biochemical energy to maintain active cellular metabolism in nodules. Even under phosphate-limited conditions, nodules are strong sinks of phosphate [5]. Iron is required for the biosynthesis of leghaemoglobin, which captures and delivers oxygen to maintain a microaerobic environment for bacterial respiration without inhibiting the nitrogen-fixing nitrogenase, an iron-containing enzyme [1]. Nitrogenase biosynthesis requires sulfur, which is also actively metabolized in bacteroids for redox homeostasis [4]. Zinc homeostasis is critical for maintaining enzymatic reactions and signal transduction in legume nodules [1]. The exchange of amino acids was also critical for nodule functions [1,6,7]. While bacteroids were capable of synthesizing and secreting alanine in soybean nodules, their requirement of branch-chain amino acids imported from the host for development and effective nitrogen fixation was demonstrated in pea [8]. It is also important that the legume host recognizes rhizobia as benign microbes and does not trigger defense responses during symbiosis [9]. Although plant defense pathways related to nodulation have not been well studied, it has been demonstrated that the expression

Abbreviations: H3K4me3, trimethylation of histone H3 lysine 4; N, nodules; RR, remaining roots; ChIP-seq, chromatin-immunoprecipitation followed by sequencing; RNA-seq, RNA-sequencing; DEGs, differentially expressed genes; DRs, differentially enriched regions; DRGs, DR-related genes; DPI, days post-inoculation; TFs, transcription factors

* Corresponding author: Room E409, Science Centre East Block, School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China.

E-mail address: honming@cuhk.edu.hk (H.-M. Lam).

¹ Contributed equally to this work.

<https://doi.org/10.1016/j.ygeno.2020.09.052>

Received 4 July 2020; Received in revised form 7 September 2020; Accepted 22 September 2020

Available online 26 September 2020

0888-7543/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

levels of most plant immunity-related genes increased shortly after inoculation and then reduced to background levels after the establishment of symbiosis [9]. Therefore, it is assumed that the suppression of plant innate immunity plays a critical role during nodule establishment and maintenance.

To maintain the effective functioning of nitrogen-fixing nodules, coordinated transcription regulation is essential. Epigenetic mechanisms are well-known for regulating gene expressions without changing the DNA sequences, especially in biotic/abiotic stress responses and developmental processes [10]. With respect to legume-rhizobia symbiosis, studies in *Medicago* showed that DNA methylation influenced the expressions of more than half of the nodule-specific cysteine-rich genes, and that DNA demethylation modulated by DEMETER is a key epigenetic factor associated with nodule differentiation [11,12]. Recently, increased DNA methylation especially in the CHH-context was reported during nodulation in soybean [13]. Additionally, several nodule-specific genes were found to be regulated by histone modifications. Specifically the ratio of H3K27me3 to H3K9ac levels was associated with gene expression levels [11]. Moreover, statistical analyses of four histone marks (H3K9ac, H3K9me2, H3K27me1 and H3K27me3) in roots and nodules highlighted the differential distributions of these marks at symbiotic islands related to nodule development in *Medicago* [14]. Therefore, epigenetic modifications are closely linked to nodule development and functions.

Trimethylation of histone H3 lysine 4 (H3K4me3) is well-known as an activator of gene expressions in multiple biological processes in plants [15,16]. In the common bean (*Phaseolus vulgaris*), the down-regulation of a histone H3 lysine 4 trimethyltransferase gene, *PvTRX1h*, reduced nitrogen fixation activities in nodules [17]. However, a complete genomic profile of H3K4me3 for exploring its regulatory role in nodulation has been lacking. We hypothesize that the distribution of H3K4me3 represents an epigenetic program consisting of a set of genes responsible for nodule maintenance and functions. In this study, we aim to generate the first global H3K4me3 profile in nodulated soybean through chromatin-immunoprecipitation followed by sequencing (ChIP-seq) and to discover gene expression regulation associated with alterations in H3K4me3 mark with the support of RNA-sequencing (RNA-seq).

2. Materials and methods

2.1. Plant materials and growth conditions

Surface-sterilized *Glycine max* cv. C08 seeds were germinated in autoclaved vermiculite. Seedlings were inoculated with *Sinorhizobium fredii* CCBAU45436 four days after sowing and watered with Milli-Q water regularly [48]. Seedlings were grown at 28 °C under a 16 h/8 h light-dark cycle during the whole experiment.

2.2. RNA-seq and RT-qPCR analysis

To ensure reproducibility of the RNA-Seq analysis, we prepared two independent sets of samples. For the first set of RNA-seq data, nodules and the remaining roots (main roots and lateral roots after nodules had been excised, 28 DPI [days post-inoculation]) were collected in two biological replicates. Total RNA was extracted with RNAiso Plus (Takara, USA). PolyA-based strand-specific libraries were constructed and sequenced by BGI-Shenzhen (Shenzhen, China), and were designated as Set 1. For the second set of RNA-seq data, nodules (25–28 DPI) and the remaining roots (crown roots only) were harvested with three biological replicates. Total RNA was extracted with TRIzol reagent (Invitrogen, USA) following manufacturer's instructions. Ribosomal RNA-depleted strand-specific libraries were constructed and sequenced by Groken Bioscience (Hong Kong SAR, China), and designated as Set 2. Both sets were sequenced with the Illumina HiSeq 2000 platform.

Raw reads of RNA-seq were cleaned by TrimGalore (<https://github.com/FelixKrueger/TrimGalore>) and mapped to the soybean reference genome, *Glycine max Wm82.a4.v1* [49], by Hisat2 [50]. Differentially expressed genes (DEGs; *q*-values < 0.05) were identified using DEseq2 [51]. DEGs common to both sets of RNA-seq libraries were extracted as high-confidence DEGs. The FPKM (fragments per kilobase of transcript per million mapped reads) and log₂FC (fold change) values of high-confidence DEGs in Set 1 were chosen for downstream analyses.

DNase I treatment was performed as previously described following manufacturer's instructions (Invitrogen, USA) and cDNA was generated using High-Capacity cDNA Reverse Transcription Kit (Thermofisher, USA) with 18-mer oligo dT. The cDNA was diluted 10-fold with Milli-Q water before being used as the template. 1.5 μL of template cDNA was added to a reaction mixture containing 0.4 μL forward primer (10 μM), 0.4 μL reverse primer (10 μM), 5 μL iQ SYBR Green Supermix (Bio-Rad, USA) and 2.7 μL Milli-Q water. Reactions were run on the CFX384 Touch Real-Time PCR Detection System (Bio-Rad, USA). Primers used were listed in Supplemental Table S23. The thermal cycling profile was set according to manufacturer's protocol. Relative gene expression was calculated using the 2^{−ΔΔCT} method [52], with *Bic-C2*, *F-box protein2*, and *ELF1b* as the reference genes.

2.3. ChIP-seq and ChIP-qPCR analysis

At 28 DPI, nodules and the remaining roots (crown roots only) were harvested and crosslinked with 1% formaldehyde and stored at −80 °C until use [53]. Nuclei were isolated from cross-linked tissues using Nuclei Isolation Buffer as previously described [54]. Nucleus lysate was sonicated using Bioruptor UCD-200 at 'High Power', 21 cycles (30 s on/30 s off) to obtain chromatin fragments with sizes ranging from 100 bp to 1 kb. Sonicated chromatin was precleared with Dynabeads Protein A (Thermo Fisher Scientific, USA) at 4 °C for 1 h with shaking. Small aliquots of precleared chromatin were saved as INPUT control. For ChIP reactions, 4 μg anti-H3K4me3 antibodies (Cat# 07-473, Millipore, USA) was added to the 10-fold-diluted precleared chromatin (> 15 μg). The mixture was incubated overnight at 4 °C with shaking [53]. 40 μL Dynabeads Protein A was then added to the reaction mixture. After further incubation for 5 h at 4 °C, beads were collected and washed with Low Salt Wash, High Salt Wash, LiCl Wash and TE wash (2 times) subsequently [53]. After elution, ChIPed DNA was reverse-crosslinked with 200 mM NaCl at 65 °C for 6 h and purified with DNA Clean & Concentrator (Zymo Research, USA).

Libraries were prepared from ChIPed DNA and sequenced with the Illumina HiSeq 2500 platform (Macrogen, Korea). Raw data were processed by TrimGalore to discard low-quality reads and adapter sequences. Then, the clean reads which aligned to the *Sinorhizobium fredii* CCBAU45436 genome were filtered out [55] and the remaining reads were mapped to the soybean reference genome (*Glycine max Wm82.a4.v1*) [49] by Bowtie2 [56] with default parameters. Four libraries of input control were merged following a previous methodology [57], and only uniquely mapped reads were extracted for further analyses. To determine the correlation between biological repeats, Pearson correlation coefficient was computed on normalized signal intensity by deepTools [58]. deepTools was also used to convert aligned reads to the wiggle format, which could then be visualized with JBrowse (http://www.wildsoydb.org/Gsoja_W05/jbrowse/h3k4me3-nodulation) [59]. MACS2 [60] was used to perform peak calling and differential regions between nodules and remaining roots. Only the common peaks generated in at least two replicates were picked for the final annotation.

For qPCR, ChIPed DNA was 40-fold-diluted and the INPUT control was 100-fold-diluted. PCR reactions were set up as described above. Primers used were listed in Supplemental Table S23. H3K4me3 enrichment was expressed in %INPUT based on *C_T* values.

2.4. Western blot analysis

Tissues harvested the same way as the ChIP experiments above were

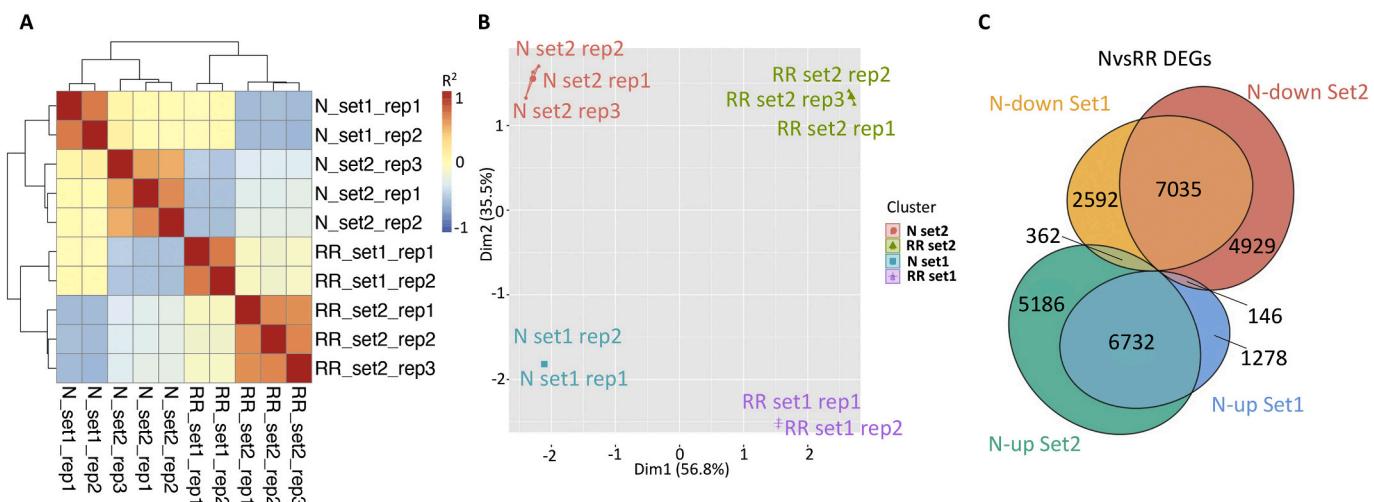


Fig. 1. Transcriptomic analysis of mature nodules and remaining roots of soybean inoculated with *Sinorhizobium fredii* 28 days post-inoculation. (A) Pearson correlation coefficients of all RNA-seq libraries. (B) Principal component analysis (PCA) of all RNA-seq libraries. (C) Venn diagram showing the differentially expressed genes (DEGs) in nodules compared to the remaining roots in two independent sets of RNA-seq data (Set1 and Set2) with different RNA processing protocols. N-up = up-regulated in the nodules, N-down = down-regulated in the nodules, N = nodules, RR = remaining roots.

ground into fine powder in liquid nitrogen. Nucleus extract was re-suspended in 0.4 M HCl at 4 °C overnight. 100% trichloroacetic acid (TCA) solution was then added to the supernatant to a final concentration of 33% for protein precipitation at 4 °C overnight. After centrifugation at 12,000 xg for 10 min, the pellet was washed with ice-cold acetone thrice. RIPA buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, 1 × protease inhibitor cocktail) was then used to dissolve the protein pellet. 3 × protein loading dye containing β-mercaptoethanol was added to the samples. After boiling at 99 °C for 10 min, samples were loaded into a 15% SDS-PAGE gel. Proteins were separated and transferred to a polyvinylidene difluoride (PVDF) membrane. The membrane was blocked with 2% skimmed milk in TBST (20 mM Tris, 150 mM NaCl, 0.1% w/v Tween®20, pH 7.6) and incubated with anti-H3K4me3 antibodies (Cat# 07–473, Millipore, USA) (1:5000, 2% skimmed milk in TBST) at 4 °C overnight with shaking. After washing with TBST thrice, anti-rabbit IgG (1:10,000, 2% skimmed milk in TBST) was added for incubation at room temperature for 1 h. After washing with TBST thrice, Clarity Western ECL Substrate (Bio-Rad, USA) was used for signal development. Signal was detected using the ChemiDoc Imaging System (Bio-Rad, USA).

2.5. GO and KEGG enrichment analysis

Gene Ontology (GO) were downloaded from the PlantTFDB-v5.0 database [20] and GO enrichment analysis was performed using the web-based agriGO tools [61]. Enrichment test was performed using Fisher's exact test and the Benjamini-Yekutieli False Discovery Rate *p*-value normalization. Background terms were set to all GO terms in *Glycine max*. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment was analyzed by KOBAS 3.0 [62].

2.6. Sequence alignment and phylogenetic analysis

Multiple alignments of coding sequences of the selected transporter genes were performed by MUSCLE-v3.8.31 with default parameters [63]. Phylogenetic trees were constructed using the neighbor-joining method in the same software.

2.7. Transcription factor (TF) classification and enrichment analysis

We downloaded all TFs in *Glycine max* from PlantTFDB-v5.0 [20]

(<http://planttfdb.cbi.pku.edu.cn/index.php?sp=Gma>). The hypergeometric distribution test was applied for computing the significance and enrichment of transcription factors in different gene sets [64].

2.8. Motif analysis

Meme-Chip from the MEME suit tools [65] was used to identify the motif at the promoters (2 kb upstream) of the up- and down-regulated DEG sets.

2.9. Construction of regulatory network

The plant transcriptional regulatory map predicted by three methods (motif, motif_CE, FunTFBS) in PlantTFDB-v5.0 [20] was searched to identify the regulatory network among down-regulated DEGs in nodules (<http://plantregmap.cbi.pku.edu.cn/download.php#networks>). Fisher's exact test was conducted to check whether the TF along with its target genes are significantly enriched (*q*-value < 0.001), and all regulations predicted in the database were used as background. Only the TFs that were found among the set of overlapping DEGs and DRGs were selected as the hub transcription factors. The network was visualized by Cytoscape-v3.7.2 [66].

3. Results

3.1. Differentially expressed genes (DEGs) are related to root nodule functions

To perform transcriptomic analysis in nodulated soybean plants, nodules and the remaining roots were collected 28 days post-inoculation (DPI). Two independent sets of samples were analyzed and the results were highly consistent (see below). A total of 18,145 and 24,390 DEGs between nodules and the remaining roots were identified in Set 1 and Set 2, respectively. In total, around 20 million paired-end mapped reads were obtained from each library (Supplemental Table S1). Pearson correlation and principal component analysis (PCA) among all RNA-seq libraries were plotted in Fig. 1, A and B. In general, the expressed genes (FPKM > 0) from the nodules and the remaining roots were distinctly separated into two groups. Moreover, replicates from the same tissue were clustered together. Fifty-two selected genes were validated with RT-qPCR, indicating that these data were highly reproducible (Supplemental Fig. S1). Those DEGs common to both sets

were designated as high-confidence DEGs. In total, 6732 up-regulated and 7035 down-regulated high-confidence DEGs were identified in nodules versus the remaining roots (Fig. 1C, Supplemental Table S2). Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis indicated that pathways such as “purine metabolism”, “carbon metabolism”, “biosynthesis of amino acids”, which were known to be induced in nodules, were enriched among up-regulated DEGs in nodules (Supplemental Table S3). On the other hand, pathways involved in “plant hormone signal transduction” and “plant-pathogen interaction” were significantly enriched among down-regulated DEGs in nodules (Supplemental Table S4).

3.2. Genome-wide profile of H3K4me3 marks was associated with differential gene expressions in nodules

Total histone was extracted from the isolated nuclei of nodules and the remaining roots 28 DPI. Western blot analysis showed that the total H3K4me3 level in nodules was significantly higher than that in the remaining roots (Supplemental Fig. S2). In light of the drastic difference between the two tissues, we performed chromatin immunoprecipitation-sequencing (ChIP-seq) to profile the enrichment of H3K4me3 marks at the level of individual genes in these two tissues. Over 20 million uniquely mapped reads were obtained per library (Supplemental Table S5). Two biological replicates were performed, with a Pearson correlation coefficient of 0.97 for nodules and 0.98 for the remaining roots between the replicates, indicating that the ChIP results were highly reproducible (Supplemental Fig. S3A). In total, around 37,000–39,000 peaks (differences between sample and input control) were called from each tissue (Fig. 2A, Supplemental Table S6). We observed higher levels of H3K4me3 in transcriptionally active chromatin (euchromatin; Fig. 2A). The distribution of H3K4me3 peaks showed similar patterns of preferential enrichment in genic regions in both tissues (Fig. 2B). The distribution of H3K4me3 marks along genic regions showed that the highest H3K4me3 level was observed just after the transcriptional start site (TSS) and the level decreased gradually along the gene body in both tissues (Fig. 2C).

Differential analysis of peaks between nodules and the remaining roots was performed and differentially enriched regions (DRs) were plotted (Fig. 2A). In total, there were 6038 peaks with higher H3K4me3 enrichment and 3367 peaks with lower H3K4me3 enrichment in nodules compared to the remaining roots. A higher number of up-regulated regions were found in nodules, which is consistent with our western blot result (Supplemental Fig. S2). We then defined the genes containing DRs within 2 kb upstream of TSS and/or gene bodies as differentially enriched region-related genes (DRGs). A total of 3797 DRGs with higher H3K4me3 enrichment and 2092 DRGs with lower enrichment in nodules compared to the remaining roots were identified (Supplemental Table S7). The ChIP-Seq results were individually validated at selected DRGs by ChIP-qPCR (Fig. 2E, Supplemental Fig. S3B), and the results were highly consistent with the ChIP-Seq data.

Based on the global H3K4me3 profile of expressed genes, higher expression levels were found among the genes with H3K4me3 enrichment than those without in both tissues (Supplemental Fig. S4A). Spearman correlation analysis of the genes with H3K4me3 enrichment revealed a positive correlation (Spearman correlation coefficient = 0.39, p -value < 2.2^{e-16}) between the levels of H3K4me3 and transcription in both tissues (Supplemental Fig. S4B).

To further investigate the relationship between transcriptional regulation and the H3K4me3 profile, we compared the DRGs against the DEGs (Supplemental Table S8). In total, we identified 1966 (51.8%) up-regulated DRGs overlapping with up-regulated DEGs and 1267 (60.6%) down-regulated DRGs overlapping with down-regulated DEGs, when comparing nodules to the remaining roots (Fig. 3A). As expected, the majority of DRGs overlapped with DEGs in the same regulatory direction (DEGs + DRGs; Supplemental Table S9). Scatter plot and heat map revealed a strong positive relationship between the fold change

levels of H3K4me3 and transcription (Spearman correlation coefficient = 0.81, p -value < 2.2^{e-16}) (Fig. 3, B and C). These are consistent with a previous finding that H3K4me3 positively affects gene expression [16].

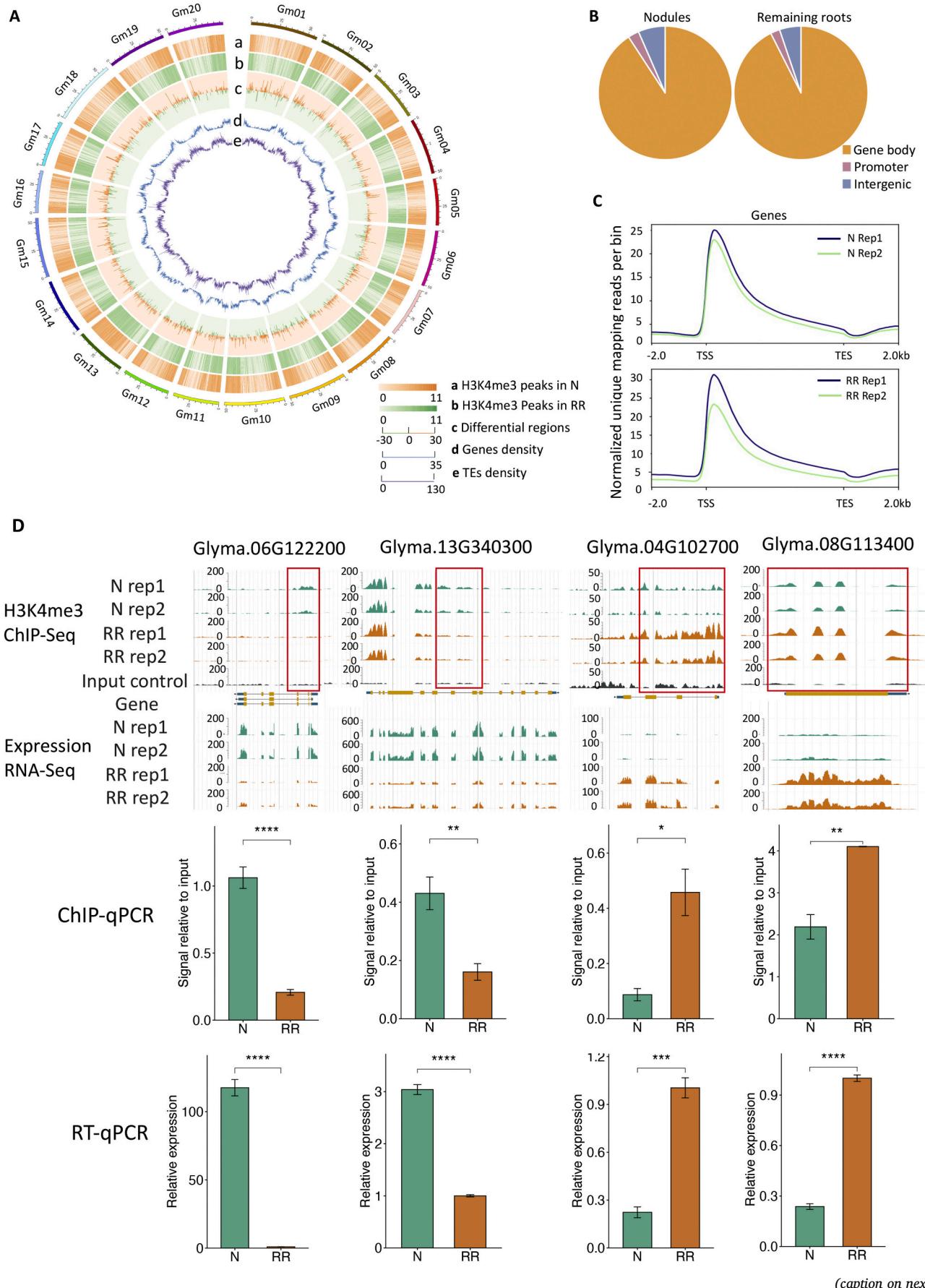
Through KEGG enrichment analysis, we found the pathways of “plant hormone signal transduction” and “plant-pathogen interaction” were enriched among down-regulated DRGs, and “purine metabolism”, “carbon metabolism”, and “biosynthesis of amino acids” were enriched among up-regulated DRGs in nodules compared to the remaining roots (Supplemental Tables S10 and S11). In GO enrichment analysis, genes with higher H3K4me3 enrichment in nodules were enriched in terms such as “nodulation”, “nitrogen fixation”, “transporter activity” and “iron ion binding” (Fig. 3D, Supplemental Tables S12 and S13). It is not surprising that genes encoding all four functional leghemoglobin (Lba, Lbc1, Lbc2 and Lbc3) were highly marked with H3K4me3 ($\log_2FC > 4$) in nodules, along with other nodulin genes including *nodulin-16, -20, -22, -24, -44* and *ENOD93* ($\log_2FC > 3.5$). For genes with reduced H3K4me3 levels in nodules, they were enriched with GO terms such as “adenyl nucleotide binding”, “defense response” and “transcription factor activity”. When we focused on transcription factors (TFs) among the DEGs and DRGs, we found the NIN-like (Nodule Inception), SRS (SHI-related sequence), B3, bZIP (basic leucine zipper), TALE (three amino acid loop extension), and Trihelix families were overrepresented in all up-regulated data sets (DEGs, DRGs and DEGs + DRGs; Supplemental Table S14). On the other hand, the WRKY and homeodomain-leucine zipper TF families were significantly enriched in all down-regulated data sets (Supplemental Table S15).

3.3. Nitrogen and carbon metabolic pathways are highly associated with increased H3K4me3 levels in nodules

The exchange of carbon and nitrogen between rhizobia and hosts is a fundamental process during symbiotic nitrogen fixation. To further demonstrate the possible role of H3K4me3 in regulating essential metabolic activities in nodules, genes involved in nitrogen assimilation and carbon metabolic pathways were analyzed.

Differential enrichment in H3K4me3 in nodules was found in most of the genes encoding metabolic enzymes within the purine and ureide biosynthesis pathway (Fig. 4, Supplemental Table S16). In nodules, ammonia exported from bacteroids was first assimilated into glutamine by glutamine synthetase, including *GS1a*, *GS1γ1* and *GS1γ2*, which also had higher H3K4me3 levels. Two genes encoding phosphoribosylpyrophosphate synthetase, responsible for the generation of precursors for *de-novo* purine biosynthesis, were also marked with higher H3K4me3 levels. Increase in H3K4me3 levels was also found in both *inosine monophosphate dehydrogenase (IMPD)* genes, which govern the rate-limiting step in purine biosynthesis, generating xanthosine monophosphate for ureide production. Although *purine-nucleoside phosphorylase (PNP)* was down-regulated at both H3K4me3 and transcriptional levels, two *uridine nucleosidase (URH)* genes, involved in inosine and xanthosine hydrolysis in *Arabidopsis* [18], were up-regulated. Within the downstream ureide biosynthesis pathway, higher H3K4me3 levels were also observed in genes encoding xanthine dehydrogenase.

In return for the fixed nitrogen, plants provide bacteroids with dicarboxylates. The majority of genes involved in the carbon metabolism pathway were differentially modified with H3K4me3 (Fig. 4, Supplemental Table S16). Sucrose is metabolized by sucrose synthases (SS) or alkaline invertase (AI) to hexoses (glucose and fructose). Most of the genes annotated with SS had lower expression levels and one of them had a lower H3K4me3 level in nodules when compared to the remaining roots. Conversely, all the DEGs encoding AI were highly expressed in nodules. Hexoses were phosphorylated by hexokinase (HK) or fructokinase (FK) and enter glycolysis. Five of the FK genes had higher H3K4me3 levels while one HK gene had a lower H3K4me3 level and was down-regulated in nodules. Most of the genes involved in glycolysis were highly expressed with higher H3K4me3 levels.



(caption on next page)

Fig. 2. Distribution of H3K4me3 marks in mature nodules and remaining roots of soybean inoculated with *Sinorhizobium fredii* 28 days post-inoculation. (A) Genome-wide distribution of H3K4me3 peaks in (a) nodules and (b) remaining roots, along with (c) fold-changes of differential peaks with higher levels of H3K4me3 nodules than in remaining roots (in orange) and those with higher levels in remaining roots than in nodules (in green), (d) distribution of genes (in blue) and (e) distribution of transposable elements (TEs) (in purple) in the soybean genome. (B) Distribution of H3K4me3 marks within different genomic regions in nodules and remaining roots. (C) The H3K4me3 profile from 2 kb upstream to 2 kb downstream of all soybean genes in nodules (upper panel) and remaining roots (lower panel). The ChIP signal was generated using the normalized sequencing read density of H3K4me3. (D) Gene expression levels (by RT-qPCR) and H3K4me3 levels (by ChIP-qPCR) of two DEGs up-regulated in nodules (N-up DEGs) and two DEGs up-regulated in remaining roots (RR-up DEGs) are displayed in JBrowse. Regions differentially modified with H3K4me3 are boxed. *Bic-C2*, *ELF-1b* and *F-box protein2* were used as housekeeping genes in RT-qPCR analysis. Error bars indicate SD, $n = 3$ (* p -value < 0.05; ** p -value < 0.01; *** p -value < 0.001; **** p -value < 0.0001, as determined using t -test). The same experiment was repeated in another biological replicate and a consistent trend was observed. N = nodules, RR = remaining roots. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Ultimately, phosphoenolpyruvate (PEP) is processed by phosphoenolpyruvate carboxylase (PEPC) and malate dehydrogenase (MDH) into malate, the major carbon source supplied to bacteroids. While *GmPEPC17* was down-regulated with a lower H3K4me3 level in nodules, *GmPEPC7* and *GmPEPC15* were highly expressed in nodules. In addition, higher expression levels were found in almost all the genes encoding MDH.

3.4. Regulation of transporter genes are correlated with H3K4me3 levels in nodules and the remaining roots

The epigenetic regulation of transporter genes in nodules was complicated, since GO terms related to transporter activities were discovered among both up- and down-regulated genes in nodules. Numerous potential sugar, peptide, amino acid, and ion transporter-encoding genes were found within the DEGs + DRGs dataset (Supplemental Table S17). Among the 16 sugar transporter genes with both higher H3K4me3 and expression levels in nodules compared to the remaining roots, two *SWEET-like* genes (*Glyma.06G122200* and *Glyma.19G232200*) were identified. However, eight other sugar transporter-encoding genes showed lower H3K4me3 and expression levels in nodules. Also, two putative tonoplast dicarboxylate transporter-encoding genes (*Glyma.07G265900* and *Glyma.15G115200*) were up-regulated in nodules.

Since differential H3K4me3 enrichment was observed in 21 potential soybean members of the oligopeptide transporter (OPT) and the nitrogen/phosphate transporter (NRT/PTR) families, a phylogenetic tree was generated using nucleotide sequences from these 21 genes and those from key *NRT/PTRs* identified in *Arabidopsis* (Fig. 5A). *Glyma.04G036100* and *Glyma.13G323800*, which showed similarity to *AtNRT2.1* and *AtNPF6.3* respectively, were down-regulated in both H3K4me3 and transcription levels in nodules. While two closely related genes (*Glyma.02G224600* and *Glyma.05G136400*) showing similarity to *Arabidopsis PTRs* were regulated in opposite directions in nodules, another branch with seven up-regulated genes with similarity to *AtPTRs* and *AtNPF6.3* was also observed. Differential regulation in both directions (one up-regulated and two down-regulated soybean genes in nodules) was also found among genes clustered with *AtOPTs* (Fig. 5A).

Differential H3K4me3 enrichment and expression were observed in 18 genes that potentially encode amino acid transporters. Within the cluster containing amino acid permeases (*AtAAP1-5*), three soybean genes were up-regulated in nodules (Fig. 5B). *Glyma.12G084500* and *Glyma.20G188800*, both down-regulated, are related to the gamma-aminobutyric acid (GABA) transporter gene, *AtGAT1* [19]. While two genes showing similarity to *AtCAT1* were up-regulated, two genes (*Glyma.05G155500* and *Glyma.08G113400*), related to *Arabidopsis polyamine uptake transporter* (*AtPUT3*), were down-regulated in nodules. Differential regulation was also observed in genes related to *Arabidopsis lysine histidine transporter* (*AtLHT1*) and *bidirectional amino acid transporter* (*AtBAT1*) (Fig. 5B).

The potential solutes for ion transport-related genes include sulfate, phosphate, iron, zinc, copper, calcium and potassium ions (Supplemental Table S17). When comparing nodules against the remaining roots, H3K4me3 and expression levels of two genes encoding

sulfate transporters were increased, whereas the levels in two others were decreased. While *GmPT5*, encoding a phosphate transporter known to regulate nodulation, was up-regulated and possessed higher H3K4me3 levels, two other *GmPTs* (*GmPT1* and *GmPT4*) had lower levels of both expression and H3K4me3 in nodules. In addition, the ferrous ion transporter-encoding genes *GmVTL1a*, *GmDMT1*, and *GmNRAMP3a* (*Glyma.04G044000*), were modified by higher H3K4me3 levels in soybean nodules. The same trend in H3K4me3 level was also observed in *GmZIP1* and *GmCOPT2* (*Glyma.01G106700*), which encode zinc and copper transporters, respectively (Supplemental Table S17).

3.5. Down-regulated transcription factors (TFs) formed the hubs of downstream defense response regulatory networks in nodules

As mentioned, TF enrichment and classification analysis demonstrated an overrepresentation of the WRKY family in the down-regulated DEGs, DRGs and DEGs + DRGs datasets (Supplemental Table S15). This indicates that certain TFs could play important regulatory roles in nodules. Therefore, we constructed a transcriptional regulatory network among all down-regulated DEGs to obtain pairwise regulatory interactions between TFs and their targets from PlantTFDB5.0 [20]. In total, dozens of TFs were significantly over-represented by possessing targets among down-regulated genes in nodules, where seven down-regulated TFs were simultaneously marked by lower H3K4me3 levels (Supplemental Table S18). An integrative network linking these seven TFs and their 1315 target genes through 1963 pairwise interactions was constructed in Cystoscope (Supplemental Fig. S5, Supplemental Table S19). Among the target genes, 1028 (78.2%) did not have differential enrichment of H3K4me3, suggesting they could be regulated by different mechanisms. The network showed four hubs all made up of TFs, including four *GmWRKYs* (*GmWRKY50*, *GmWRKY90*, *GmWRKY135* and *GmWRKY149*) clustered together, plus *GmTCP18*, *GmTCP43*, and *GmMYB214*, with clusters of 227, 327, 479 and 509 target genes respectively. Besides the expected GO terms of “transcription factor activity” and related terms being significantly enriched among the target genes, those terms such as “defense response”, “response to stimulus” and “response to other organism” were also over-represented (Fig. 6A, Supplemental Table 20). Among all target genes, 17 were annotated as NBS-LRR (nucleotide binding site leucine-rich repeat) R (resistance) protein-coding genes. Moreover, 21 of the target genes encoded other WRKYs, including the homologs of *AtWRKY11*, *AtWRKY22*, *AtWRKY29* and *AtWRKY33*.

Motif analysis was also conducted at the 2000-bp upstream promoter regions of the up- and down-regulated DEGs using the MEME software suite (E -value < 0.001). Hundreds of motifs were significantly enriched in both directions including the binding motifs of TCPs and MYBs. Interestingly, TTGACY (W-box element), typical of the WRKY family binding motif, was significantly enriched among the promoters of the down-regulated DEGs set (Supplemental Table S21) when compared to the up-regulated one (Supplemental Table S22), supporting the role of WRKYs as master regulators of the repressed genes in nodules.

The sub-network comprising the first-neighbor interactions of *GmWRKYs* was shown in Fig. 6B. The predicted binding motifs of these *GmWRKYs* matched our motif scanning results (Supplemental Table

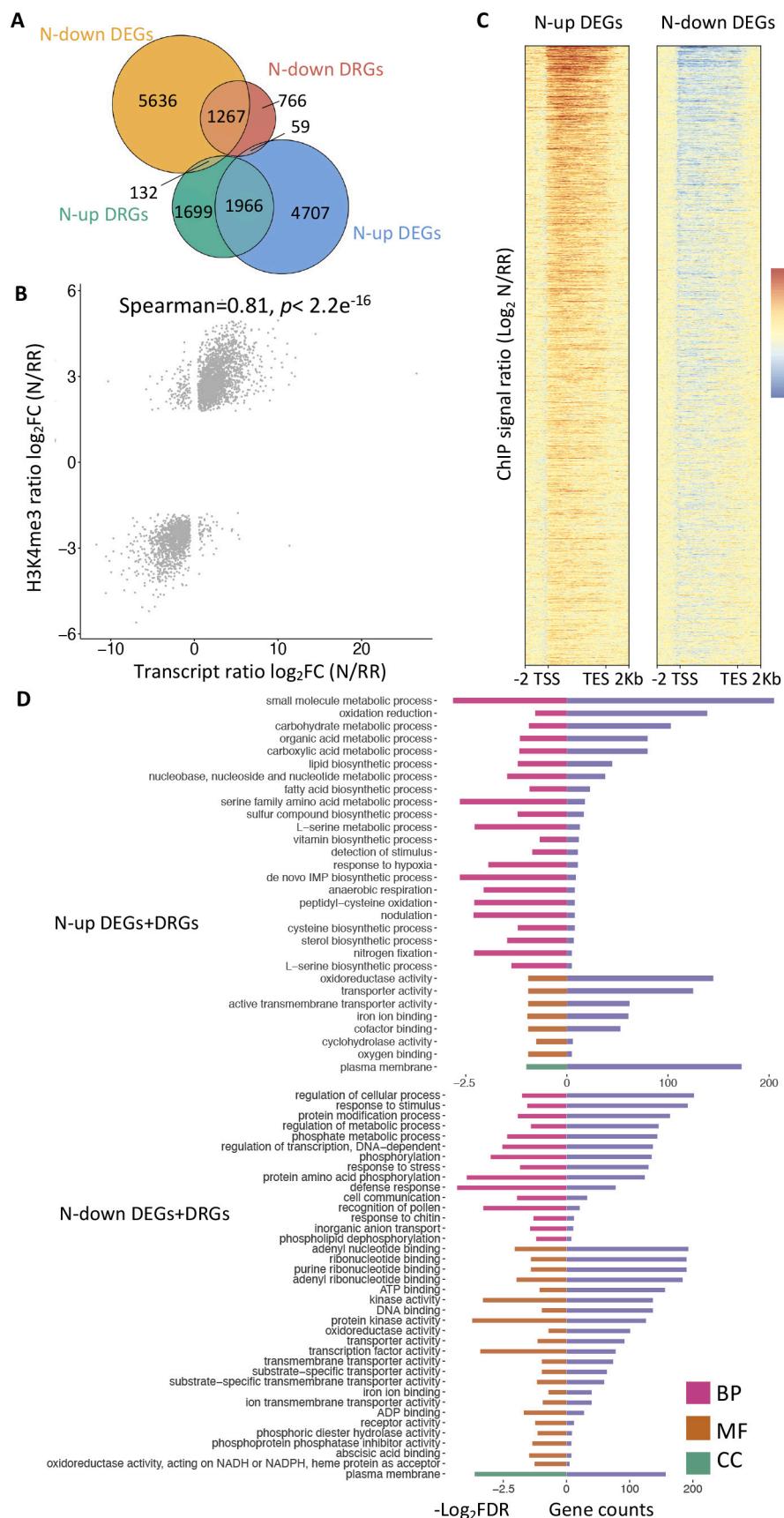


Fig. 3. Correlation between differentially expressed genes (DEGs) and H3K4me3 differentially enriched region-related genes (DRGs) in nodules and remaining roots of soybean inoculated with *Sinorhizobium fredii* 28 days post-inoculation. (A) Venn diagram of DEGs and H3K4me3 DRGs in nodules and remaining roots. (B) Spearman correlation of overlapped genes in (A). (C) Heatmaps showing the H3K4me3 signal ratios ($\log_2 N/RR$) of up-regulated genes in nodules (N-up DEGs; left panel, number of genes = 6732) and down-regulated genes in nodules (N-down DEGs; right panel, number of genes = 7035). Genes are ranked from top to bottom in each line according to the fold change in the expression level in nodules versus in remaining roots. (D) Gene Ontology (GO) enrichment analysis of the overlapping genes between N-up DEGs and N-up DRGs (N-up DEGs + DRGs; upper panel, number of genes = 1966) and those between N-down DEGs and N-down DRGs (N-down DEGs + DRGs; lower panel, number of genes = 1267). The fold of differential enrichment ($\log_2 FDR$) associated with each GO term classified under biological process (BP), molecular function (MF) and cellular component (CC) is represented in pink, orange and green respectively on the left half of the bar plots. The number of genes associated with each of these GO terms is shown in purple on the right half of the bar plots. The corresponding nonredundant GO terms are listed to the left of the bar plots. The full GO enrichment results are listed in Supplemental Tables S12 and S13. N = nodules, RR = remaining roots. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

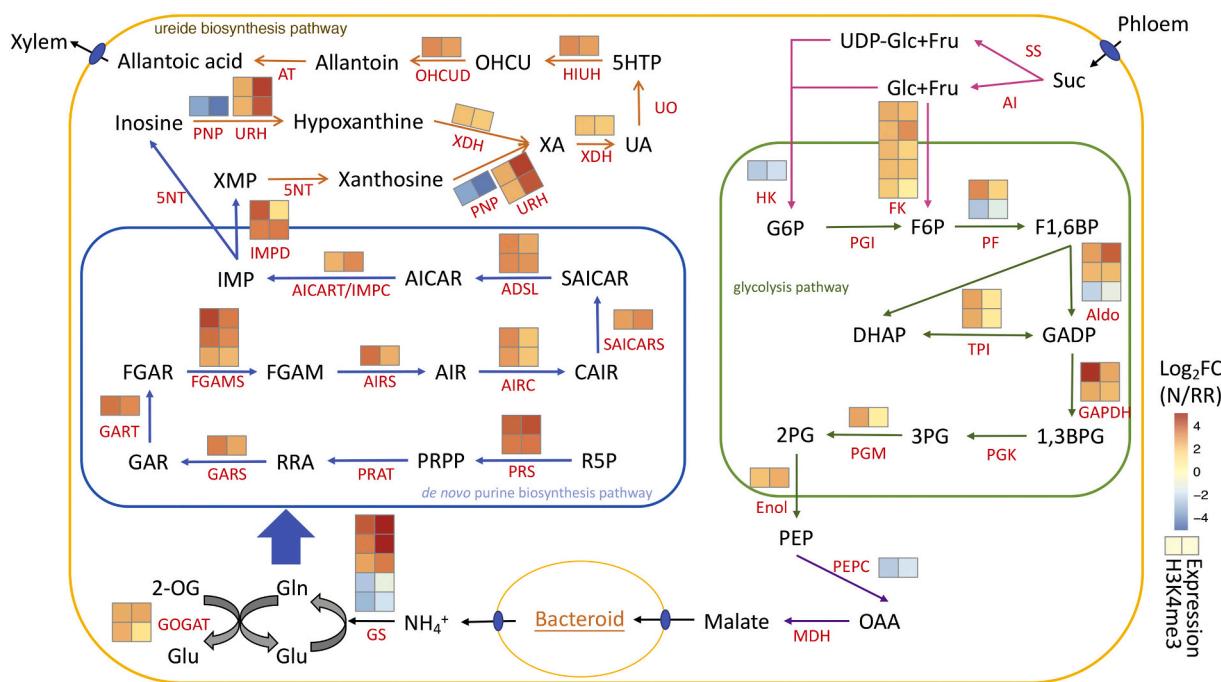


Fig. 4. Alterations of H3K4me3 in nitrogen and carbon metabolic pathway between nodules and remaining roots. Heatmaps show fold change of expression and H3K4me3 level in genes along ureide and purine biosynthesis pathway between nodules and remaining roots. Red color indicates higher level while blue color indicates lower level in nodules. 2-OG, 2-oxoglutarate; Gln, Glutamine; Glu, Glutamate; GOGAT, Glutamate synthetase; GS, Glutamine synthetase; NH₄⁺, Ammonium; R5P, Ribose 5-phosphate; PRS, Phosphoribosylpyrophosphate synthetase; PRPP, 5-phosphoribosyl-1-pyrophosphate; PPAT, Glutamine phosphoribosyltransferase; RRA, 5-phosphoribosylamine; GARS, GAR synthetase; GAR, 5-phosphoribosylglycinamide; GART, GAR transfromylase; FGAR, 5-phosphoribosyl-N-formylglycinamide; FGAMS, FGAM synthetase; FGAM, 5-phosphoribosyl-N-formylglycinamide; AIRS, AIR synthetase; AIR, 5-phosphoribosylaminoimidazole; AIRC, AIR carboxylase; CAIR, 1-(5-phosphoribosyl)-5-amino-4-carboxyimidazole; SAICARS, SAICAR synthase; SAICAR, 1-(5-phosphoribosyl)-4-(N-succinocarboxamide)-5-aminoimidazole; ADSL, adenylosuccinate lyase; AICAR, 1-(5-phosphoribosyl)-5-amino-4-imidazolecarboxamide; AICART/IMPC, AICAR transformylase/IMP cyclohydrolase; IMP, inosine monophosphate; IMPD, IMP dehydrogenase; XMP, Xanthosine monophosphate; 5NT, 5'-nucleotidase; PNP, purine-nucleoside phosphorylase; NSH2, nucleoside hydrolase 2; XA, xanthine; XDH, Xanthine dehydrogenase; UA, uric acid; UO, urate oxidase(uricase); 5HTP, 5-hydroxyisourate; HIUH, hydroxyisourate hydrolase; OHCUD, 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline decarboxylase; OHCUD, OHCUD decarboxylase; AT, allantoinase; Suc, Sucrose; SS, Sucrose synthase; AI, Alkaline invertase; Fru, Fructose; HK, hexokinase; FK, fructokinase; G6P, glucose-6-phosphate; PGI, glucose-6-phosphate isomerase; F6P, fructose-6-phosphate; PFK, 6-phosphofructokinase phosphohexokinase; F1,6BP, fructose 1,6-bisphosphate; Aldo, aldolase; GADP, glyceraldehyde 3-phosphate; TPI, triosephosphate; DHAP, dihydroxyacetone phosphate; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; 1,3BPG, 1,3-bisphosphoglyceric acid; PGK, phosphoglycerate kinase; 3PG, 3-phosphoglycerate; PGM, 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase; 2PG, 2-phosphoglycerate; Enol, enolase; PEP, phosphoenolpyruvate; PEPC, PEP carboxylase; OAA, Oxaloacetate; MDH, malate dehydrogenase. N = nodules, RR = remaining roots. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

S21). Among the target genes of GmWRKYs, there were *GmWRKY4* (*Glyma.01G128100*) with four other WRKYs and three *NBS-LRR R* genes.

4. Discussion

4.1. Possible roles of H3K4me3 in regulating gene expressions in soybean nodules

Epigenetic modifications are a key mechanism for controlling developmental processes in plants. Through regulating transcription, epigenetic marks such as histone modifications establish temporal and spatial controls on seed germination, vegetative growth and flowering. In legumes, nodules are specialized organs responsible for biological nitrogen fixation through symbiosis with rhizobia. During symbiosis, multiple metabolic processes and transport systems contribute to nodule formation and functions. In this study, the H3K4me3 peaks in nodules and the remaining roots were located in genic regions, where the most enrichment was observed just after the transcriptional start site. Such a distribution is consistent with previous reports on soybean and other plants [15,21], strongly supporting the role of H3K4me3 in transcriptional regulation. Our analysis found around 6000 differentially modified genes between nodules and the remaining roots, suggesting that the two tissues possess distinct H3K4me3 profiles.

Additionally, 51.8% of up-regulated DRGs in nodules were transcriptionally activated while 60.6% of down-regulated DRGs in nodules were transcriptionally repressed, when compared to the remaining roots. The uncoupling of differential H3K4me3 enrichment with differential gene expression in 45.5% of N-up DRGs and 36.5% of N-down DRGs can have several explanations. It was reported that the ploidy-dependent expression of a subset of nodule-specific genes was correlated with the ratio of H3K9ac to H3K27me3 [11], and a similar phenomenon in the genomic regions associated with nodule development in *Medicago truncatula* [14]. These findings suggest that multiple epigenetic marks could be involved in regulating nodule gene expressions. H3K4me3 enrichment may also reflect previous transcriptional activities, as suggested by its persistence after transcription termination [22].

Within the gene sets with both differential expression and differential H3K4me3 enrichment in the same regulatory direction, the fold-change in H3K4me3 and expression levels were highly correlated, suggesting that alterations in H3K4me3 levels exert a strong effect on the expression of these genes. Among them, nodule genes annotated with the GO terms “nodulation” and “nitrogen fixation” were enriched as expected. Furthermore, the up-regulation of other nodulin genes known to be highly expressed in nodules was also associated with drastic increases (> 3.5 log₂FC) in their H3K4me3 levels. Thus, we have demonstrated that the up-regulation of processes directly involved

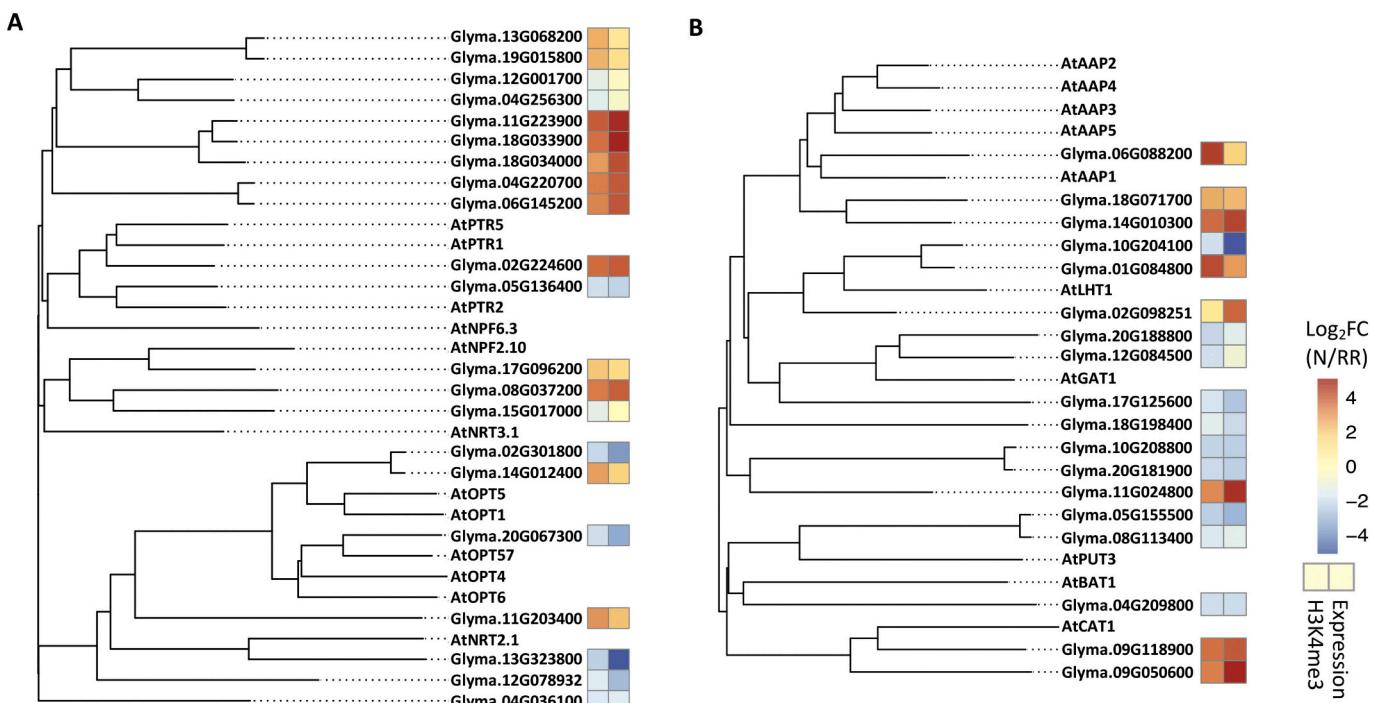


Fig. 5. Phylogenetic trees of genes related to peptide and amino acid transports. Coding sequences from (A) putative oligopeptide and nitrate/peptide transporter-encoding genes and (B) amino acid transporter-encoding genes, together with key transporter-encoding genes identified in *Arabidopsis* were used to construct the phylogenetic trees. Fold-changes in expression (right) and H3K4me3 levels (left) in the selected genes in nodules compared to remaining roots ($\text{Log}_2\text{FC} [\text{N}/\text{RR}]$) are shown in heatmaps. Red color indicates a higher level while blue color indicates a lower level in nodules. N = nodules, RR = remaining roots. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

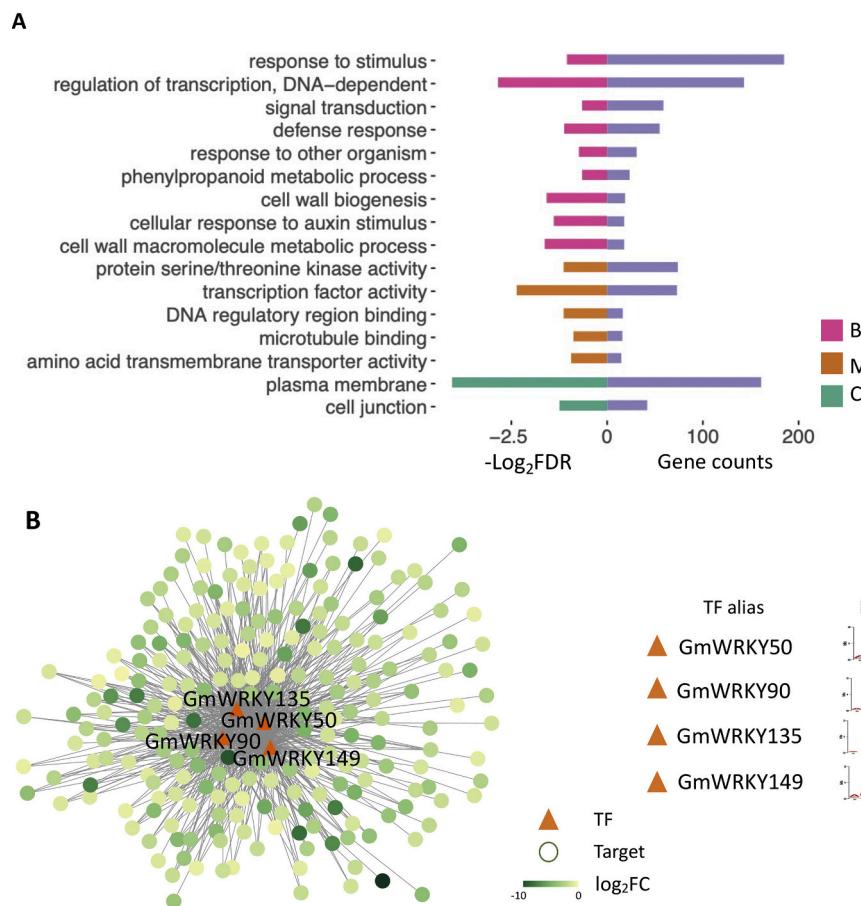


Fig. 6. Regulatory network of down-regulated DEGs in nodules. (A) Gene Ontology (GO) enrichment analysis of the target genes of transcription factors (TFs) identified in the whole regulatory network associated with nodulation (Supplemental Fig. S5). The significance levels (Log_2FDR) of selected GO terms classified under biological process (BP), molecular function (MF) and cellular component (CC) are represented in pink, orange and green, respectively, on the left half of the bar plot. The number of genes belonging to the corresponding GO terms are shown in purple on the right half of the bar plot. The corresponding nonredundant GO terms are listed to the left of the bar plot. The full results are listed in Supplemental Table S20. (B) A negative gene regulatory subnetwork based on pairwise interactions of *GmWRKY*s. Orange triangles represent four of the down-regulated TFs with a loss of H3K4me3 and circles represent their targets. The green color from darkest to lightest represents the fold-change in target gene expression from the highest to the lowest as represented by Log_2FC . Predicted binding motifs (PlantTFDB v5.0) of *GmWRKY*s are shown to the right of their names. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in nitrogen fixation is potentially mediated by H3K4me3 enrichment, consistent with the previously demonstrated role of H3K4me3 in regulating nitrogen fixation by knocking down *histone H3 lysine 4 trimethyltransferase, PvTRX1h*, in common bean [17].

H3K4me3-associated regulation was also observed in most of the genes encoding metabolic enzymes within the purine and ureide biosynthesis pathways, especially those governing the rate-limiting steps, suggesting that epigenetic regulation may help drive ammonia assimilation towards ureide production. Regarding carbon metabolism in nodules, sucrose synthases (SSs) was suggested to play a more important role than alkaline invertases (AIs) during nitrogen fixation [3]. *Nodulin-100*, which encodes a subunit of SSs, was reported to be highly expressed in soybean nodules [23]. This is consistent with our data, where the expressions of *Nodulin-100* and its homolog were increased in nodules. The nodule-specific *GmPEPC7* and its closest homolog *GmPEPC15*, which were reported to be highly expressed in nodules [24], were also induced in nodules in this study. Yet the low expression level of bacterial-type PEPC genes (*GmPEPC17*) in this study was consistent with a previous report as well [25]. Furthermore, higher H3K4me3 levels were observed in the majority of up-regulated genes involved in glycolysis, demonstrating the role of epigenetic regulation in carbon metabolism in nodules.

4.2. Concerted activation of amino acid biosynthesis facilitates efficient nodule functions

Compared to the remaining roots, nodules require more active metabolism to maintain the nutrient supply to bacteroids and to assimilate fixed nitrogen. Here the activation of amino acid biosynthesis was identified as one of the key features in nodule metabolism, consistent with previous studies [26]. Higher H3K4me3 enrichment was confined to up-regulated genes within the biosynthetic pathways of several key amino acids, including the one encoding glutamate synthetase, involved in the first step of ammonia assimilation. The up-regulation of cysteine biosynthesis can be helpful for minimizing oxidative stress in the micro-environment in nodules by promoting the subsequent production of thiol-containing peptides such as glutathione. The up-regulation of the gene for serine hydroxymethyltransferase, responsible for converting glycine to serine, could favor the incorporation of glycine into phosphoribosylamine in the second step of purine biosynthesis. Moreover, glycine is also involved in glutathione biosynthesis. Similarly, the up-regulation of histidine biosynthesis in nodules could drive the production of 5-aminoimidazole-4-carboxamide ribonucleotide, an intermediate of purine biosynthesis. Besides, the biosynthesis of branch-chain amino acids was also up-regulated, consistent with a previous finding that the supply of these amino acids to bacteroids was critical for nodule functions [8].

4.3. H3K4me3 signatures reveal distinct transportation machinery in nodules and remaining roots

To date, numerous studies have identified transporters of individual metallic ions essential for nitrogen fixation in legume nodules [1]. In contrast, there is less effort on identifying transporters of carbon and organic nitrogen compounds. Hence this study shed lights on the epigenetic regulation of various transporter families in nodules.

The regulation of sugar transport-related genes appeared to be multi-faceted. While our DEG data revealed that most of the putative members in the sugar transporter, polyol/monosaccharide transporter and inositol transporter subfamilies were up-regulated in nodules, *STP1-like*, *STP5-like*, *STP7-like* and *STP14-like* were up-regulated in the remaining roots. Interestingly, several *STP-like* genes were also up-regulated in the root-specific cluster compared to indeterminate nodules in *Medicago*, suggesting that these STPs may participate in the same mechanism governing sugar distribution [27]. It was proposed that the induction of SWEET genes, such as *MtN3* in *Medicago*, facilitates

nutrient supplies to bacteroids [28]. It was reported that *MtSWEET1b* was beneficial for the maintenance of arbuscular mycorrhizal symbiosis, and also possibly in rhizobial symbiosis given its high expression in nodules [29]. From our data, the *MtSWEET1b* homolog in soybean (*Glyma.06G122200*) was also up-regulated in nodules, supporting its potential role in nodule maintenance. It was shown that *LjALMT* (aluminum-activated malate transporter) could mediate the efflux of dicarboxylates including malate in *Xenopus oocytes* [30]. In *Lotus japonicus*, *LjALMT4* was localized in vascular bundles in mature nodules instead of the peribacteroid membrane, suggesting that another class of transporter was responsible for the direct supply of dicarboxylates to bacteroids [30]. *AttDT* (a tonoplast malate/fumarate transporter) was responsible for malate import into vacuoles in *Arabidopsis* [31]. However, the *AttDT* homologs (*Glyma.07G265900* and *Glyma.15G115200*) were up-regulated in nodules, suggesting that they may be involved in an alternative malate transporting mechanism.

Besides sugar transporters, other transporters were also differentially regulated in nodules, such as the NRT/PTR transporters. *Glyma.04G036100* and *Glyma.13G323800* were related to nitrate transporter genes (*AtNRT2.1* and *AtNPF6.3*) in *Arabidopsis*. The repression of nitrate transporters may be a strategy to limit unnecessary nitrate acquisition by the root, which can inhibit nitrogen fixation. It was reported that *LjNPF8.6* LORE1-insertion mutants had reduced nitrogen-fixing activity and displayed nitrogen deficiency symptoms. Despite *LjNPF8.6* showing nitrate uptake activity in *Xenopus laevis* oocytes, the nitrate content in mutant nodules did not significantly differ from that in the wild type. The up-regulation of *LjNPF8.6* was proposed to regulate membrane potential and energy status [32]. *AtPTR5*, the closest *Arabidopsis* ortholog of *LjNPF8.6*, is a dipeptide transporter [33]. Hence, differential regulations of *Glyma.02G224600* and *Glyma.18G064900* (> 72% predicted amino acid identity with *LjNPF8.6*) may suggest the importance of transporter capable of transporting both nitrate and dipeptide in soybean nodules. *Glyma.17G096200*, ortholog of *AtNPF2.11*, was also up-regulated in nodules. In *Arabidopsis*, *AtNPF2.11* mediates glucosinolate transport. In the colonization of *Arabidopsis* by beneficial root endophytes, the improvement in plant growth was abolished when glucosinolate biosynthesis was interrupted [34]. Since this phenotype was observed under low inorganic phosphate (Pi) conditions, the role of glucosinolate in soybean-rhizobia symbiosis awaits further investigation.

An amino acid cycling model between plants and bacteroids was proposed, in which the exchange of amino acids was beneficial to nitrogen fixation efficiency [6]. In our study, genes potentially involved in amino acid transport were also differentially regulated in nodules. In *Arabidopsis*, amino acid permeases were shown to transport a wide range of neutral amino acids, including alanine and branch-chain amino acids. In pea, the expression of a branch-chain amino acid-specific transport system complemented the impaired symbiotic performance of the bacterial mutant lacking a broad-specificity amino acid transporter [8]. Thus, the up-regulation of amino acid permeases in soybean may favor the supply of branch-chain amino acids to bacteroids for maintaining nodule functions. GABA cycling was also proposed to regulate nitrogen fixation [35]. The unexpected down-regulation of two genes potentially encoding GABA transporters may be due to the other role of GABA as a stress-signaling molecule, reflecting a delicate balance between unnecessary stress response and healthy nodule functions.

Differential regulations of ion transporters were also observed in nodules. In plants, Pi transporters from the Pht1 family display tissue-specific functions and are responsible for Pi uptake and allocation from source to sink [36]. Higher expression levels of *GmPT1* and *GmPT4* in the remaining roots were consistent with previous findings of their tissue-specific expression patterns [37]. In soybean nodules, *GmDMT1* (an NRAMP metal transporter family member) was shown to mediate ferrous ion transport [4]. In *Medicago*, another NRAMP family member, *MtNramp1*, was also capable of ferrous ion transport. It was proposed

that Nramp1 and DMT1 constitute a transport system to deliver ferrous ions to the bacteroids [30]. These two NRAMP transporters also mediated the transport of other ions, such as manganese, zinc and copper [30,38]. Thus, the up-regulation of *GmDMT1* and *GmNRAMP3a* may favor the supply and homeostasis of various ions in soybean nodules. The up-regulation of *GmVTL1a* associated with higher H3K4me3 levels was also consistent with a recent study demonstrating its function as a ferrous ion transporter important for nodule functions [39].

Consistent with previous studies, the zinc transporter *GmZIP1* was up-regulated in nodules [4]. Regarding copper ion transport, the insertion mutant *Mtcopt1* had impaired nitrogenase activities in nodules [1]. Here *GmCOPT2* (*Glyma.01G106700*) had up-regulated expressions with higher H3K4me3 levels in nodules, suggesting that it may be responsible for symbiotic copper transport in soybean. However, several genes encoding sulfate transporters were regulated in opposite directions in nodules. Although the sulfate transporter *LjSST1* was demonstrated to be essential for nodule functions [1], our results suggested a more complicated system for sulfate homeostasis in soybean nodules.

4.4. Defense response network in nodules is dampened by the loss of H3K4me3 marks in several WRKYS

Recent phylogenomic analyses of all TFs in legumes showed that the expanded TF genes during whole-genome duplication tended to be expressed in roots and nodules [40]. Also, many of the down-regulated genes in nodules in our datasets were annotated with the GO term “transcription factor activity”, leading us to analyze the enrichment of different TF families among the DEGs and DRGs. In nodules, the *Nin-like*, *bZIP* and *TALE* families were significantly enriched in three up-regulated data sets (DEGs, DRGs and DEGs + DRGs). These are involved in the positive regulation of nodulation [41–43]. In addition, three other TF families, *SRS*, *B3* and *Trihelix*, were also significantly enriched, suggesting that they also play an important role in symbiosis. On the other hand, *WRKYs* were significantly enriched in all down-regulated gene sets. These TFs were known to regulate microbe-associated molecular pattern (MAMP)-mediated host defense responses [9]. Promoter analysis revealing that W-box motifs were uniquely enriched in down-regulated DEGs further confirmed that *WRKYs* dominated among the down-regulated TFs.

By constructing a regulatory network with predicted pairwise interactions, we showed that down-regulated TFs indeed had a significant impact on the repression of their targets which were overwhelmingly annotated with defense response-related GO terms. Four *WRKYs* in particular, *GmWRKY50*, *GmWRKY90*, *GmWRKY135* and *GmWRKY149*, were clustered together in a subnetwork. It was reported that the *Atwrky11* mutant, a homolog of *GmWRKY135*, was more susceptible to nematode infection [44]. *AtWRKY22* and *AtWRKY29*, homologs of *GmWRKY149* and *GmWRKY50* respectively, are important components in the mitogen-activated protein kinase (MAPK) pathway. The over-expression of *AtWRKY29* conferred resistance to both bacterial and fungal pathogens and similarly, *AtWRKY22* [45], both being inducible by *Bacillus cereus* AR156 [46]. Moreover, we discovered five other *WRKYs* and three *NBS-LRR R* genes among the targets of these four *GmWRKYs*. Although the mechanism of *WRKY* self-regulation is still unclear, this phenomenon has been reported in several studies under various stress conditions [45]. Surprisingly, one of these target genes was *GmWRKY4* (*Glyma.01G128100*, the homolog of *AtWRKY33*), with reduced H3K4me3 enrichment in nodules. *AtWRKY33* is a key positive regulator against *Botrytis cinerea* and the repression of it may promote the disease in *Arabidopsis* [47]. Our findings suggest that loss of H3K4me3 may down-regulate key TFs, which in turn exert broader effects on the defense-responsive target gene network in nodules. The inhibition of host innate immunity appears to be essential for establishing symbiosis between rhizobia and the plant. However, it should be noted that regulation of the reported network may involve other mechanisms that remain to be elucidated. Also, as variations in host

specificity and symbiotic efficiency exist in soybean-rhizobia symbiosis, further studies comparing different germplasms and rhizobia strains can be conducted to address the consistency in defense response networks and to identify features specific to different symbiotic partners.

5. Conclusions

In summary, we uncovered the transcriptomic program with an H3K4me3 signature specialized for nodule functions, including activating amino acid biosynthesis, facilitating carbon and nitrogen metabolism to support efficient nitrogen fixation, and facilitating the exchange of materials between the host and bacteroids. We also uncovered gene candidates for future studies on alternative transport mechanisms for improving nutrient exchange and symbiotic performance. In addition, the transcriptional network revealed that the loss of H3K4me3 in several transcription factor genes may coordinate the repression of host defense responses linked to nodule maintenance.

Accession numbers

The raw sequence data of RNA-seq have been submitted to GenBank under the accession numbers PRJAN629646 (Set 1) and PRJAN626514 (Set 2). The raw data of ChIP-seq have been submitted to GenBank under the accession number PRJAN629642.

Authors' contributions

H.-M.L. organized and supervised the research; H.-M.L., Q.W. and W.-S.Y. designed the experiments; Q.W. performed the bioinformatics analyses; W.-S.Y. performed the western blot and ChIP experiments; W.-S.Y. and Z.W. performed the qPCR experiments; Q.W., W.-S.Y. and Z.W. prepared this manuscript; H.-M.L. supervised and finalized the writing.

Funding information

This work was supported by grants from the Hong Kong Research Grants Council Area of Excellence Scheme [AoE/M-403/16]; and Lo Kwee-Seong Biomedical Research Fund.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgements

This work was supported by grants from the Hong Kong Research Grants Council Area of Excellence Scheme [AoE/M-403/16]; and Lo Kwee-Seong Biomedical Research Fund. Ms. Jee Yan Chu copy-edited this manuscript. We would like to thank Dr. Yee-Shan Ku, Ms. Zeta Mui, Ms. Fuk-Ling Wong and Dr. Johanna Wong-Bajracharya for helping with RNA-seq samples collection, Mr. Yishu Gao for helping with ChIP-seq samples collection, and Dr. Man-Wah Li for his advice on the manuscript. Any opinions, findings, conclusions or recommendations expressed in this publication do not reflect the views of the Government of the Hong Kong Special Administrative Region or the Innovation and Technology Commission.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygeno.2020.09.052>.

References

- [1] S. Roy, W. Liu, R.S. Nandety, A. Crook, K.S. Mysore, C.I. Pislaru, J. Frugoli, R. Dickstein, M.K. Udvardi, Celebrating 20 Years of Genetic Discoveries in Legume

- Nodulation and Symbiotic Nitrogen Fixation, *Plant Cell* 32 (2020) 15–41, <https://doi.org/10.1105/tpc.19.00279>.
- [2] H.M. Lam, K.T. Coschigano, I.C. Oliveira, R. Melo-Oliveira, G.M. Coruzzi, The molecular-genetics of nitrogen assimilation into amino acids in higher plants, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47 (1996) 569–593, <https://doi.org/10.1146/annurev.aplant.47.1.569>.
- [3] A. Liu, C.A. Contador, K. Fan, H.M. Lam, et al., *Front. Plant Sci.* 871 (2018) 1860, <https://doi.org/10.3389/fpls.2018.01860>.
- [4] M. Udvardi, P.S. Poole, Transport and metabolism in legume-rhizobia symbioses, *Annu. Rev. Plant Biol.* 64 (2013) 781–805, <https://doi.org/10.1146/annurev.aplant-050312-120235>.
- [5] S. Sulieman, L.S.P. Tran, Phosphorus homeostasis in legume nodules as an adaptive strategy to phosphorus deficiency, *Plant Sci.* 239 (2015) 36–43, <https://doi.org/10.1016/j.plantsci.2015.06.018>.
- [6] R. Hayama, S. Yokoi, S. Tamaki, M. Yano, K. Shimamoto, Adaptation of photoperiodic control pathways produces short-day flowering in rice, *Nature* 422 (2003) 719–722, <https://doi.org/10.1038/nature01549>.
- [7] A.M. Kinnersley, F.J. Turano, Gamma-aminobutyric acid (GABA) and plant responses to stress, *CRC Crit. Rev. Plant Sci.* 19 (2000) 479–509, <https://doi.org/10.1080/07352680091139277>.
- [8] J. Prell, J.P. White, A. Bourdes, S. Bunnewell, R.J. Bongaerts, P.S. Poole, Legumes regulate rhizobium bacteroid development and persistence by the supply of branched-chain amino acids, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 12477–12482, <https://doi.org/10.1073/pnas.0903653106>.
- [9] Y. Cao, M.K. Halane, W. Gassmann, G. Stacey, The role of plant innate immunity in the legume-rhizobium symbiosis, *Annu. Rev. Plant Biol.* 68 (2017) 535–561, <https://doi.org/10.1146/annurev-aplant-042916-041030>.
- [10] V. Chinnusamy, J.-K. Zhu, Epigenetic regulation of stress responses in plants, *Curr. Opin. Plant Biol.* 12 (2009) 133–139, <https://doi.org/10.1016/j.pbi.2008.12.006>.
- [11] M. Nagymihály, A. Veluchamy, Z. Györgyplá, F. Ariel, T. Jégú, M. Benhamed, Á. Szűcs, A. Kereszt, P. Mergaert, É. Kondoros, Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression, *Proc. Natl. Acad. Sci. U. S. A.* 114 (2017) 4543–4548, <https://doi.org/10.1073/pnas.1704211114>.
- [12] C. Satgé, S. Moreau, E. Sallet, G. Lefort, M.C. Auriac, C. Remblière, L. Cottret, K. Gallardo, C. Noirot, M.F. Jardinaud, P. Gamas, Reprogramming of DNA methylation is critical for nodule development in *Medicago truncatula*, *Nat. Plants* 2 (2016) 16166, <https://doi.org/10.1038/nplants.2016.166>.
- [13] D. Niyyikiza, S. Piya, P. Routray, L. Miao, W.-S. Kim, T. Burch-Smith, T. Gill, C. Sams, P.R. Arelli, V. Pantalone, H.B. Krishnan, T. Hewezi, Interactions of gene expression, alternative splicing, and DNA methylation in determining nodule identity, *Plant J.* (2020) 0–3, <https://doi.org/10.1111/tpj.14861>.
- [14] Y. Pecrix, S.E. Staton, E. Sallet, C. Lelandais-Brière, S. Moreau, S. Carrère, T. Blein, M.F. Jardinaud, D. Latrasse, M. Zouine, M. Zahm, J. Kreplak, B. Mayjonade, C. Satgé, M. Perez, S. Cauet, W. Marande, C. Chantry-Darmon, C. Lopez-Roques, O. Bouchez, A. Bérard, F. Debelle, S. Muñoz, A. Bendahmane, H. Bergès, A. Niebel, J. Buitink, F. Frugier, M. Benhamed, M. Crespi, J. Gouzy, P. Gamas, Whole-genome landscape of *Medicago truncatula* symbiotic genes, *Nat. Plants* 4 (2018) 1017–1025, <https://doi.org/10.1038/s41477-018-0286-7>.
- [15] H. Yan, Y. Liu, K. Zhang, J. Song, W. Xu, Z. Su, Chromatin state-based analysis of epigenetic H3K4me3 marks of Arabidopsis in response to dark stress, *Front. Genet* 10 (2019) 1–12, <https://doi.org/10.3389/fgene.2019.00306>.
- [16] S.M. Lauberth, T. Nakayama, X. Wu, A.L. Ferris, Z. Tang, S.H. Hughes, R.G. Roeder, H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation, *Cell* 152 (2013) 1021–1036, <https://doi.org/10.1016/j.cell.2013.01.052>.
- [17] A. Barraza, E.L. Coss-Navarrete, J.C. Vizuet-de-Rueda, K. Martínez-Aguilar, J.I. Hernández-Chávez, J.J. Ordaz-Ortiz, R. Winkler, A. Tiessen, R. Alvarez-Venegas, Down-regulation of PvTRX1h increases nodule number and affects auxin, starch, and metabolic fingerprints in the common bean (*Phaseolus vulgaris* L.), *Plant Sci.* 274 (2018) 45–58, <https://doi.org/10.1016/j.jplantsci.2018.05.006>.
- [18] B. Jung, C. Hoffmann, T. Möhlmann, Arabidopsis nucleoside hydrolases involved in intracellular and extracellular degradation of purines, *Plant J.* 65 (2011) 703–711, <https://doi.org/10.1111/j.1365-313X.2010.04455.x>.
- [19] A. Meyer, S. Eskandari, S. Grallath, D. Rentsch, AtGAT1, a high affinity transporter for γ -aminobutyric acid in *Arabidopsis thaliana*, *J. Biol. Chem.* 281 (2006) 7197–7204, <https://doi.org/10.1074/jbc.M510766200>.
- [20] F. Tian, D.-C. Yang, Y.-Q. Meng, J. Jin, G. Gao, PlantRegMap: charting functional regulatory maps in plants, *Nucleic Acids Res.* 48 (2019) D1104–D1113, <https://doi.org/10.1093/nar/gkz1020>.
- [21] Z. Lu, A.P. Marand, W.A. Ricci, C.L. Ethridge, X. Zhang, R.J. Schmitz, The prevalence, evolution and chromatin signatures of plant regulatory elements, *Nat. Plants* 5 (2019) 1250–1259, <https://doi.org/10.1038/s41477-019-0548-z>.
- [22] Y. Ding, M. Fromm, Z. Avramova, Multiple exposures to drought “train” transcriptional responses in Arabidopsis, *Nat. Commun.* 3 (2012) 740, <https://doi.org/10.1038/ncomms1732>.
- [23] F. Thummel, D.P. Verma, Nodulin-100 of soybean is the subunit of sucrose synthase regulated by the availability of free heme in nodules, *J. Biol. Chem.* 262 (1987) 14730–14736.
- [24] S. Hata, K. Izui, H. Kouchi, Expression of a soybean nodule-enhanced phosphoenolpyruvate carboxylase gene that shows striking similarity to another gene for a house-keeping isoform, *Plant J.* 13 (1998) 267–273, <https://doi.org/10.1046/j.1365-313X.1998.00022.x>.
- [25] S. Sullivan, G.I. Jenkins, H.G. Nimmo, Roots, cycles and leaves. expression of the phosphoenolpyruvate carboxylase kinase gene family in soybean, *Plant Physiol.* 135 (2004) 2078–2087, <https://doi.org/10.1104/pp.104.042762>.
- [26] L. Brechenmacher, M.Y. Kim, M. Benitez, M. Li, T. Joshi, B. Calla, P.L. Mei, M. Libault, L.O. Vodkin, D. Xu, S.H. Lee, S.J. Clough, G. Stacey, Transcription profiling of soybean nodulation by *Bradyrhizobium japonicum*, *Mol. Plant-Microbe Interact.* 21 (2008) 631–645, <https://doi.org/10.1094/MPMI-21-5-0631>.
- [27] J. Doidy, U. Vidal, R. Lemoine, Sugar transporters in Fabaceae, featuring SUT MST and SWEET families of the model plant *Medicago truncatula* and the agricultural crop *Pisum sativum*, *PLoS One*. 14 (2019) e0223173, <https://doi.org/10.1371/journal.pone.0223173>.
- [28] P. Gamas, Use of a Subtractive Hybridization Approach to Identify New *Medicago truncatula* Genes Induced During Root Nodule Development, *Mol. Plant-Microbe Interact.* 9 (1996) 233, <https://doi.org/10.1094/MPMI-9-0-233>.
- [29] J. An, T. Zeng, C. Ji, S. de Graaf, Z. Zheng, T.T. Xiao, X. Deng, S. Xiao, T. Bisseling, E. Limpens, Z. Pan, A *Medicago truncatula* SWEET transporter implicated in arbuscule maintenance during arbuscular mycorrhizal symbiosis, *New Phytol.* 224 (2019) 396–408, <https://doi.org/10.1111/nph.15975>.
- [30] M. Tejada-Jiménez, R. Castro-Rodríguez, I. Kryvoruchko, M.M. Lucas, M. Udvardi, J. Imperial, M. González-Guerrero, *Medicago truncatula* Natural Resistance-Associated Macrophage Protein1 Is Required for Iron Uptake by Rhizobia-Infected Nodule Cells, *Plant Physiol.* 168 (2015) 258–272, <https://doi.org/10.1104/pp.114.254672>.
- [31] B. Frei, C. Eisenach, E. Martinoia, S. Hussein, X.-Z. Chen, S. Arrivault, H.E. Neuhaus, Purification and functional characterization of the vacuolar malate transporter tDT from *Arabidopsis*, *J. Biol. Chem.* 293 (2018) 4180–4190, <https://doi.org/10.1074/jbc.RA117.000851>.
- [32] V.T. Valkov, A. Rogato, L.M. Alves, S. Sol, M. Noguero, S. Léran, B. Lacombe, M. Chiurazzi, The Nitrate Transporter Family Protein LjNPF8.6 Controls the N-Fixation Node Activity, *Plant Physiol.* 175 (2017) 1269–1282, <https://doi.org/10.1104/pp.17.01187>.
- [33] U.Z. Hammes, S. Meier, D. Dietrich, J.M. Ward, D. Rentsch, Functional Properties of the *Arabidopsis* Peptide Transporters AtPT1 and AtPT5, *J. Biol. Chem.* 285 (2010) 39710–39717, <https://doi.org/10.1074/jbc.M110.141457>.
- [34] K. Hiruma, N. Gerlach, S. Sacristán, R.T. Nakano, S. Hacquard, B. Kracher, U. Neumann, D. Ramírez, M. Bucher, R.J. O’Connell, P. Schulze-Lefert, Root Endophyte *Colletotrichum tofieldiae* Confers Plant Fitness Benefits that Are Phosphate Status Dependent, *Cell* 165 (2016) 464–474, <https://doi.org/10.1016/j.cell.2016.02.028>.
- [35] S. Sulieman, Does GABA increase the efficiency of symbiotic N 2 fixation in legumes? *Plant Signal. Behav.* 6 (2011) 32–36, <https://doi.org/10.4161/psb.6.1.14318>.
- [36] L. Qin, Y. Guo, L. Chen, R. Liang, M. Gu, G. Xu, J. Zhao, T. Walk, H. Liao, Functional characterization of 14 Pht1 family genes in yeast and their expressions in response to nutrient starvation in soybean, *PLoS One*. 7 (2012) e47726, <https://doi.org/10.1371/journal.pone.0047726>.
- [37] H. Song, Z. Yin, M. Chao, L. Ning, D. Zhang, D. Yu, Functional properties and expression quantitative trait loci for phosphate transporter GmPT1 in soybean, *Plant Cell Environ.* 37 (2014) 462–472, <https://doi.org/10.1111/pce.12170>.
- [38] B.N. Kaiser, S. Moreau, J. Castelli, R. Thomson, A. Lambert, S. Bogliolo, A. Puppo, D.A. Day, The soybean NRAMP homologue, GmDMT1, is a symbiotic divalent metal transporter capable of ferrous iron transport, *Plant J.* 35 (2003) 295–304, <https://doi.org/10.1046/j.1365-313X.2003.01802.x>.
- [39] S. Liu, L.L. Liao, M.M. Nie, W.T. Peng, M.S. Zhang, J.N. Lei, Y.J. Zhong, H. Liao, Z.C. Chen, A VIT-like transporter facilitates iron transport into nodule symbiosomes for nitrogen fixation in soybean, *New Phytol.* 226 (2020) 1413–1428, <https://doi.org/10.1111/nph.16506>.
- [40] K.C. Moharana, T.M. Venancio, Polyploidization events shaped the transcription factor repertoire in legumes (Fabaceae), *Plant J.* (2020) 0–2, <https://doi.org/10.1111/tpj.14765>.
- [41] C.-W. Liu, A. Breakspear, D. Guan, M.R. Cerri, K. Jackson, S. Jiang, F. Robson, G.V. Radhakrishnan, S. Roy, C. Bone, N. Stacey, C. Rogers, M. Trick, A. Niebel, G.E.D. Oldroyd, F. de Carvalho-Niebel, J.D. Murray, NIN Acts as a Network Hub Controlling a Growth Module Required for Rhizobial Infection, *Plant Physiol.* 179 (2019) 1704–1722, <https://doi.org/10.1104/pp.18.01572>.
- [42] K. D’haeseleer, A. De Keyser, S. Goormachtig, M. Holsters, Transcription factor MtATB2: about nodulation, sucrose and senescence, *Plant Cell Physiol.* 51 (2010) 1416–1424, <https://doi.org/10.1093/pcp/pcq104>.
- [43] E. Di Giacomo, C. Laffont, F. Sciarra, M.A. Iannelli, F. Frugier, G. Frugis, KNAT3/4/5-like class 2 KNOX transcription factors are involved in *Medicago truncatula* symbiotic nodule organ development, *New Phytol.* 213 (2017) 822–837, <https://doi.org/10.1111/nph.14146>.
- [44] M. Amjad Ali, K. Wieczorek, D.P. Kreil, H. Bohlmann, The beet cyst nematode *Heterodera schachtii* modulates the expression of WRKY transcription factors in syncytia to favour its development in *Arabidopsis* roots, *PLoS One*. 9 (2014) e102360, <https://doi.org/10.1371/journal.pone.0102360>.
- [45] J. Jiang, S. Ma, N. Ye, M. Jiang, J. Cao, J. Zhang, WRKY transcription factors in plant responses to stresses, *J. Integr. Plant Biol.* 59 (2017) 86–101, <https://doi.org/10.1111/jipb.12513>.
- [46] S. Wang, Y. Zheng, C. Gu, C. He, M. Yang, X. Zhang, J. Guo, H. Zhao, D. Niu, *Bacillus cereus* AR156 Activates Defense Responses to *Pseudomonas syringae* pv. *tomato* in *Arabidopsis thaliana* Similarly to flg22, Mol, *Plant-Microbe Interact* 31 (2018) 311–322, <https://doi.org/10.1094/MPMI-10-17-0240-R>.
- [47] S. Liu, J. Ziegler, J. Zeier, R.P. Birkenbihl, I.E. Somssich, *Botrytis cinerea* B05.10 promotes disease development in *Arabidopsis* by suppressing WRKY33-mediated host immunity, *Plant Cell Environ.* 40 (2017) 2189–2206, <https://doi.org/10.1111/pce.13022>.
- [48] H.M. Rehman, W.-L. Cheung, K.-S. Wong, M. Xie, C.-Y. Luk, F.-L. Wong, M.-W. Li, S.-N. Tsai, W.-T. To, L.-Y. Chan, H.-M. Lam, High-Throughput Mass Spectrometric

- Analysis of the Whole Proteome and Secretome From *Sinorhizobium fredii* Strains CCBAU25509 and CCBAU45436, *Front. Microbiol.* 10 (2019) 1–19, <https://doi.org/10.3389/fmicb.2019.02569>.
- [49] J. Schmutz, S.B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D.L. Hyten, Q. Song, J.J. Thelen, J. Cheng, D. Xu, U. Hellsten, G.D. May, Y. Yu, T. Sakurai, T. Umezawa, M.K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X.C. Zhang, K. Shinozaki, H.T. Nguyen, R.A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R.C. Shoemaker, S.A. Jackson, Genome sequence of the paleaeopolyploid soybean, *Nature* 463 (2010) 178–183, <https://doi.org/10.1038/nature08670>.
- [50] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, *Nat. Methods.* 12 (2015) 357–360, <https://doi.org/10.1038/nmeth.3317>.
- [51] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol.* 15 (2014) 550, <https://doi.org/10.1186/s13059-014-0550-8>.
- [52] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta CT$ method, *Methods* 25 (2001) 402–408, <https://doi.org/10.1006/meth.2001.1262>.
- [53] M.M. Ricardi, R.M. Gonzalez, N.D. Iusem, Protocol: fine-tuning of a Chromatin Immunoprecipitation (ChIP) protocol in tomato, *Plant Methods* 6 (2010) 11, <https://doi.org/10.1186/1746-4811-6-11>.
- [54] T. Wu, T. Yuan, S.-N. Tsai, C. Wang, S.-M. Sun, H.-M. Lam, S.-M. Ngai, Mass spectrometry analysis of the variants of histone H3 and H4 of soybean and their post-translational modifications, *BMC Plant Biol.* 9 (2009) 98, <https://doi.org/10.1186/1471-2229-9-98>.
- [55] J. Jiao, M. Ni, B. Zhang, Z. Zhang, J.P.W. Young, T.F. Chan, W.X. Chen, H.M. Lam, C.F. Tian, Coordinated regulation of core and accessory genes in the multipartite genome of *Sinorhizobium fredii*, *PLoS Genet.* 14 (2018) e1007428, <https://doi.org/10.1371/journal.pgen.1007428>.
- [56] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, *Nat. Methods* 9 (2012) 357–359, <https://doi.org/10.1038/nmeth.1923>.
- [57] H.-R. Chung, C. Xu, A. Fuchs, A. Mund, M. Lange, H. Staeghe, T. Schubert, C. Bian, I. Dunkel, A. Eberharter, C. Regnard, H. Klinker, D. Meierhofer, L. Cozzuto, A. Winterpacht, L. Di Croce, J. Min, H. Will, S. Kinkley, PHF13 is a molecular reader and transcriptional co-regulator of H3K4me2/3, *Elife* 5 (2016) 1–34, <https://doi.org/10.7554/elife.10607>.
- [58] F. Ramírez, F. Dündar, S. Diehl, B.A. Grüning, T. Manke, deepTools: a flexible platform for exploring deep-sequencing data, *Nucleic Acids Res.* 42 (2014) W187–W191, <https://doi.org/10.1093/nar/gku365>.
- [59] R. Buels, E. Yao, C.M. Diesh, R.D. Hayes, M. Munoz-Torres, G. Helt, D.M. Goodstein, C.G. Elsik, S.E. Lewis, L. Stein, I.H. Holmes, JBrowse: A dynamic web platform for genome visualization and analysis, *Genome Biol.* 17 (2016) 1–12, <https://doi.org/10.1186/s13059-016-0924-1>.
- [60] Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nussbaum, R.M. Myers, M. Brown, W. Li, X.S. Shirley, Model-based analysis of ChIP-Seq (MACS), *Genome Biol.* 9 (2008) R137, <https://doi.org/10.1186/gb-2008-9-9-r137>.
- [61] Z. Du, X. Zhou, Y. Ling, Z. Zhang, Z. Su, agriGO: a GO analysis toolkit for the agricultural community, *Nucleic Acids Res.* 38 (2010) W64–W70, <https://doi.org/10.1093/nar/gkq310>.
- [62] C. Xie, X. Mao, J. Huang, Y. Ding, J. Wu, S. Dong, L. Kong, G. Gao, C.-Y. Li, L. Wei, KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases, *Nucleic Acids Res.* 39 (2011) W316–W322, <https://doi.org/10.1093/nar/gkr483>.
- [63] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.* 32 (2004) 1792–1797, <https://doi.org/10.1093/nar/gkh340>.
- [64] A. Verkest, M. Byzova, C. Martens, P. Willems, T. Verwulgen, B. Slabbinck, D. Rombaut, J. Van de Velde, K. Vandepoele, E. Standaert, M. Peeters, M. Van Lijsebettens, F. Van Breusegem, M. De Block, Selection for Improved Energy Use Efficiency and Drought Tolerance in Canola Results in Distinct Transcriptome and Epigenome Changes, *Plant Physiol.* 168 (2015) 1338–1350, <https://doi.org/10.1104/pp.15.00155>.
- [65] W. Ma, W.S. Noble, T.L. Bailey, Motif-based analysis of large nucleotide data sets using MEME-ChIP, *Nat. Protoc.* 9 (2014) 1428–1450, <https://doi.org/10.1038/nprot.2014.083>.
- [66] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks, *Genome Res.* 13 (2003) 2498–2504, <https://doi.org/10.1101/gr.123930>.