Genome Biology

RESEARCH

Open Access

Loss of *Lateral suppressor* gene is associated with evolution of root nodule symbiosis in Leguminosae

Tengfei Liu[1,7†], Zhi Liu[1,7,9†], Jingwei Fan[1†], Yaqin Yuan[1,7†], Haiyue Liu[2,7†], Wenfei Xian[3,8], Shuaiying Xiang[1,7], Xia Yang[1,7], Yucheng Liu[1], Shulin Liu[1], Min Zhang[1], Yuannian Jiao[4,7], Shifeng Cheng[3], Jeff J. Doyle[5*], Fang Xie[2,7*], Jiayang Li[6,7,10*] and Zhixi Tian[1,7,10*]

†Tengfei Liu, Zhi Liu, Jingwei Fan, Yaqin Yuan, and Haiyue Liu contributed equally to this work.

*Correspondence: jjd5@cornell.edu; fxie@cemps.ac.cn; jyli@genetics.ac.cn; zxtian@genetics.ac.cn; tianzhixi@yzwlab.cn

[1] Key Laboratory of Seed Innovation, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China
[2] Key Laboratory of Plant Carbon Capture, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China
[5] School of Integrative Plant Science, Sections of Plant Biology and Plant Breeding & Genetics, Cornell University, Ithaca, NY, USA
[6] State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China
Full list of author information is available at the end of the article

Abstract

Background: Root nodule symbiosis (RNS) is a fascinating evolutionary event. Given that limited genes conferring the evolution of RNS in Leguminosae have been functionally validated, the genetic basis of the evolution of RNS remains largely unknown. Identifying the genes involved in the evolution of RNS will help to reveal the mystery.

Results: Here, we investigate the gene loss event during the evolution of RNS in Leguminosae through phylogenomic and synteny analyses in 48 species including 16 Leguminosae species. We reveal that loss of the *Lateral suppressor* gene, a member of the GRAS-domain protein family, is associated with the evolution of RNS in Leguminosae. Ectopic expression of the *Lateral suppressor* (*Ls*) gene from tomato and its homolog *MONOCULM 1* (*MOC1*) and *Os7* from rice in soybean and *Medicago truncatula* result in almost completely lost nodulation capability. Further investigation shows that Lateral suppressor protein, Ls, MOC1, and Os7 might function through an interaction with NODULATION SIGNALING PATHWAY 2 (NSP2) and CYCLOPS to repress the transcription of *NODULE INCEPTION* (*NIN*) to inhibit the nodulation in Leguminosae. Additionally, we find that the cathepsin H (CTSH), a conserved protein, could interact with Lateral suppressor protein, Ls, MOC1, and Os7 and affect the nodulation.

Conclusions: This study sheds light on uncovering the genetic basis of the evolution of RNS in Leguminosae and suggests that gene loss plays an essential role.

Keywords: Evolution of root nodule symbiosis, Leguminosae, Phylogenomic analysis, Gene loss, *Lateral suppressor*

Background

Nitrogen (N) is an essential nutrient for plant growth, but plants cannot obtain it directly from the atmosphere. Large quantities of nitrogen fertilizers are applied in agricultural practices each year, which significantly improves crop production but also results in great

economic and environmental costs. Nitrogen-fixing root nodule symbiosis (RNS) is a remarkable specialized trait that allows plants to acquire nitrogen from bacteria capable of fixing atmospheric nitrogen. Plants signal to bacteria, recognize compatible partners, allow their entry, accommodate them within novel organs (nodules, typically produced on roots), and provide them with a carbon source. Genetic engineering of RNS in non-nodulating crops has been a long-standing goal in sustainable agriculture [1, 2]. A better understanding of the genomic underpinnings governing RNS assembly and evolution could facilitate such efforts.

RNS is limited to plants belonging to the four orders that comprise the so-called Nitrogen Fixing Nodulation Clade (NFNC) that diverged approximately 115 million years ago (MYA) within the rosid clade, namely Fabales (including Leguminosae), Fagales, Cucurbitales, and Rosales [3]. Nodulation is rare in three of the four orders of NFNC with an exception of Fabales, where nodulation is found in the majority of species comprising the third largest family of flowering plants, Leguminosae [4]. It is widely (though not universally) accepted that nodulation evolved in the ancestor of the NFNC [5], which means that nodulating lineages have had approximately 115 million years to modify the original nodulation program inherited from that ancestor. Because of the rapid radiation/proliferation of the NFNC [6, 7], considerable divergence is expected in the nodulation symbiosis.

Gene loss is an important source of genetic variation that takes part in evolution [8–15]. In bacteria, to adapt various environmental conditions, gene loss events are frequently occurred [16]. In contrast, the importance of gene loss involving in phenotypic adaptations or evolution in plants and mammals is largely undisclosed although several cases were reported. For instance, it was found the loss of *S locus cysteine-rich protein* (*SCR*) [17] and/or *S locus receptor kinase* (*SRK*) [18] genes is essential for the evolution of outcrossing based on self-incompatibility (SI) to a selfing system in *Arabidopsis* [19]; loss of *ANTHOCYANIN2* (*AN2*) and *flavonoid 3'-hydroxylase* (*F3'h*) gene leads to flower color changes in *Petunia axillaris* and *Ipomoea quamoclit*, respectively, which is helpful for the pollination adaptation [20, 21]. For the evolution of RNS, it was reported that the parallel loss or loss-of-function of *NODULE INCEPTION* (*NIN*), *RHIZOBIUM-DIRECTED POLAR GROWTH* (*RPG*) and *NOD FACTOR PERCEPTION* (*NFP*) genes was associated with the evolution of RNS [22, 23]. However, it was also found that *NIN* was not completely lost and its homologs also presented in the non-NFNC species [22]. Nevertheless, the absent/present of the *Predisposition-Associated Cis-regulatory Element* (*PACE*) in the *NIN* promoter may play an important role in its spatial expression [24]. Though much progress on the evolution of RNS has been made, large remains to be learned, especially any other gene loss event that is also involved in its evolution.

Here, through comprehensive phylogenomic analyses and functional validation, we found that loss of the *Lateral suppressor* gene was essential for the evolution of RNS in Leguminosae and revealed a possible working model of the *Lateral suppressor* gene in controlling nodulation.

Results

Lateral suppressor and *SULFUR DEFICIENCY INDUCED 2* were lost in Leguminosae

To investigate the candidate gene loss events potentially involved in the evolution of RNS in Leguminosae, we conducted phylogenomic analyses using 48 representative species

among angiosperms with high-quality genome sequences (Fig. 1a, Additional file 2: Table S1). In this study, 16 species from three of the six subfamilies of Leguminosae [25] were sampled, including members of both subfamilies with nodulating species (Caesalpinioideae and Papilionoideae) and *Cercis canadensis,* representing the non-nodulating Cercidoideae, which is likely sister to the remaining subfamilies. Caesalpinioideae was represented by a single species each from the nodulating genus, *Chamaecrista*, and the non-nodulating *Senna*, sister to the *Chamaecrista* clade (Fig. 1a). The largest and best-known legume subfamily, Papilionoideae [4], was represented by 13 species, all of them from nodulating genera except for *Nissolia*, which has lost the ability to nodulate (Fig. 1a).

Phylogenetic analysis discovered 2 orthologous groups, SULFUR DEFICIENCY INDUCED 2 (SDI2) and Lateral suppressor (Ls), which were found in all non-Leguminosae but absent, and therefore presumably lost, in Leguminosae (Fig. 1a, Additional file 2: Table S2). SDI2 belongs to the tetratricopeptide repeat (TPR)-like superfamily and the gene in *Arabidopsis thaliana* is designated as *SDI2* which was reported to play an important role in sulfur metabolism by reducing glucosinolates biosynthesis under sulfur-deprived conditions [26]. The *Lateral suppressor* gene is a member of the GRAS-domain protein family of plant-specific transcription regulators [27–29], which can be further classified into several subfamilies [30]. The *Lateral suppressor* (*Ls*) gene in tomato [27], *LATERAL SUPPRESSOR* (*LAS*) gene in *Arabidopsis thaliana* [28], *MONOCULM 1* (*MOC1*) gene in rice [29], *ERAMOSA* in *Antirrhinum* [31], *AaLAS* in *Arabis alpine* [32], *LOSS OF AXILLARY MERISTEMS* (*LAM*) in strawberry [33], and *ClLs* in watermelon [34] are members of the Ls subfamily controlling axillary branching. We found that except for the Ls or SDI2 subfamily, other major GRAS or TPR subfamilies were universally present among Leguminosae and non-Leguminosae plant species (Fig. 1b, Additional file 1: Fig. S1a, Additional file 2: Table S3).

To confirm the loss of *Lateral suppressor* gene and *SDI2* in Leguminosae, we further performed synteny analysis and blasted the nucleotide sequences at a whole-genome level in representative Leguminosae and non-Leguminosae species. Synteny analysis revealed that although the neighboring genes of the *Lateral suppressor* gene or *SDI2* showed preservation in non-Leguminosae and Leguminosae, the *Lateral suppressor* gene and *SDI2* were absence at the expected location in Leguminosae (Fig. 1c, Additional file 1: Fig. S1b). Nucleotide sequences blast showed that no gene predicted in the best hit sequences in the Leguminosae species (Fig. 1d, Additional file 1: Fig. S1c). These results suggested that the *Lateral suppressor* gene and *SDI2* were lost in the Leguminosae species we investigated.

Ectopic expression of the non-Leguminosae *Lateral suppressor* gene inhibits nodulation in Leguminosae

To determine whether the loss of the *Lateral suppressor* gene and *SDI2* was associated with the evolution of RNS in Leguminosae, we conducted hairy root transformations. Firstly, we introduced one *SDI2* gene from rice, a non-Leguminosae species, into soybean. Although the gene was expressed, we did not observe significant nodulation differences (Additional file 1: Fig. S2a–c), suggesting that loss of *SDI2* is unlikely to be involved in the evolution of RNS in Leguminosae and it may be associated with other traits in

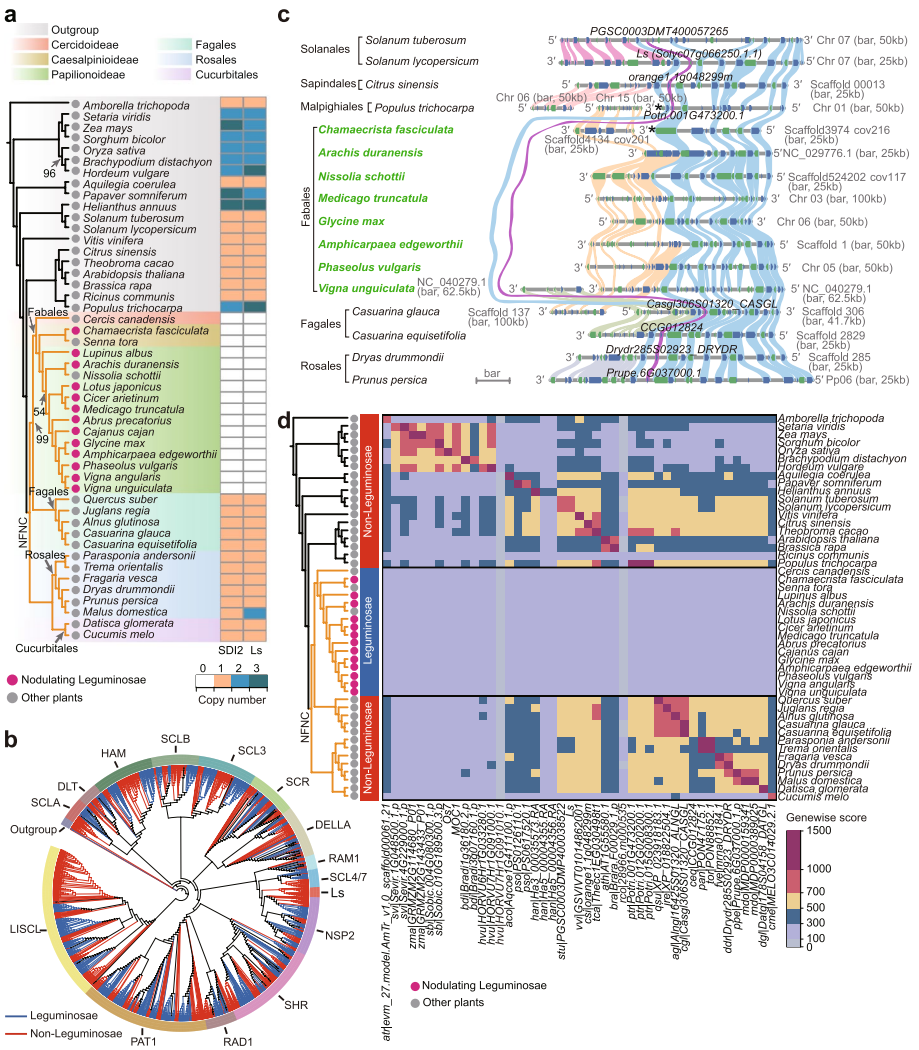

Fig. 1 *Lateral suppressor* was lost in Leguminosae. **a** The presence and absence of variation of Leguminosae-loss genes identified by phylogenomic analyses of 48 species. The species tree was constructed using 267 conserved and low-copy orthologous genes. Bootstrap values (1000 replicates) lower than 100% are indicated on the nodes. NFNC, Nitrogen Fixing Nodulation Clade; SDI2, SULFUR DEFICIENCY INDUCED 2; Ls, Lateral suppressor. **b** Phylogenetic analyses of the GRAS families from 33 species (16 Leguminosae and 17 non-Leguminosae species). A human (*Homo sapiens*) signal transducers and activators of transcription (STAT) protein (AAH02704) were used as an outgroup. Subfamilies (Ls, DELLA, etc.) are color-coded. The genes of Leguminosae and non-Leguminosae species are indicated with blue and red branches, respectively. SCL, SCARECROW like; DLT, DWARF AND LOW TILLERING; HAM, Hairy Meristem; SCR, SCARECROW; RAM1, Reduced Arbuscular Mycorrhization 1; NSP2, NODULATION SIGNALING PATHWAY 2; SHR, SHORTROOT; RAD1, REQUIRED FOR ARBUSCULE DEVELOPMENT 1; PAT1, phytochrome A signal transduction 1; LISCL, Lilium longiflorum SCARECROW like. **c** Synteny analysis of *Lateral suppressor* gene in representative non-Leguminosae and Leguminosae species. Leguminosae species names are highlighted in green. Homologous genes in each specific block are connected by lines of specific colors. The homologous genes of the *Lateral suppressor* gene in non-Leguminosae species are connected by magenta lines. Asterisks indicate the end of the chromosome or scaffold. **d** The heatmap of the genewise score of Ls orthologous groups. Full-length protein sequences of Ls orthogroups were used as baits to perform tblastn against the genome sequences of the 48 species. The best hit sequences were used to perform gene prediction using genewise, and the scores were used to plot the heatmap. The *x*-axis represents Ls orthogroups and *y*-axis represents the best hit of 48 species

Leguminosae, such as seed component accumulation as it has been reported that *SDI2* may affect seed protein accumulation under sulfur-deprived conditions [26, 35].

Then, we introduced *MOC1* and its homolog *Os7* in rice, and *Ls* gene from tomato, into soybean using hairy root transformation driven by the cauliflower mosaic virus (CaMV) 35S promoter. We found that the transgenic soybean roots carrying each of the three orthologous genes almost completely lost nodulation capability (Fig. 2a–d, Additional file 1: Fig. S3a). The same results were obtained when the three genes were individually introduced into *Medicago truncatula* (*M. truncatula*) using hairy root transformation (Fig. 2e–h, Additional file 1: Fig. S3b). The infection events were impaired when *Os7* or *MOC1* was introduced into *M. truncatula* (Additional file 1: Fig. S4a, b). Cultured with adequate N (treated with 5 mM KNO_3), no significant differences in growth habits were observed between the transgenic plants carrying empty vector and *35S:MOC1* (Additional file 1: Fig. S4c). In contrast, in the absence of N (treated with 5 mM KCl), the *35S:MOC1* transgenic plants exhibited significantly dwarfed and yellowed leaves compared to those carrying the empty vector after 4 weeks of inoculation with *Bradyrhizobium diazoefficiens* strain USDA110 (Additional file 1: Fig. S4c). These results suggested that the expression of *Lateral suppressor* gene repressed nodulation but did not affect plant growth in the presence of N nutrients. However, overexpressing members of the other GRAS subfamilies in soybean did not yield the same nodulation inhibitory effect as the *Lateral suppressor* gene (Fig. 2i–l, Additional file 1: Fig. S3c). Above results indicated that ectopic expression of *Lateral suppressor* gene rather than other GRAS members was responsible for inhibiting nodulation in Leguminosae.

Lateral suppressor protein inhibits *NIN* transcription by interacting with NSP2 and CYCLOPS

NIN is the master transcription factor that regulates Leguminosae nodulation [36, 37]. We found that the transcription of reported *MtNIN* was significantly suppressed in the *Os7*, *MOC1*, or *Ls* gene overexpression hairy roots in *M. truncatula* (Additional file 1: Fig. S3d, e). The soybean genome contains four *NIN* homologs that are all expressed on induction by rhizobia, although at very different levels [38]. Knockdown of the four *NIN* genes almost completely eliminated nodulation ability in soybean (Additional file 1: Fig. S2d–f), confirming their roles in soybean nodulation. Our findings revealed that transformation with *Os7*, *MOC1*, or *Ls* gene also significantly reduced the expression levels of all four *NIN* genes in soybean (Additional file 1: Fig. S3f, g), suggesting that the *Lateral suppressor* gene inhibits RNS by suppressing *NIN* transcription.

NIN transcription in *M. truncatula* can be induced by NSP1 through interaction with NSP2 [39]. Yeast two-hybrid, split firefly luciferase complementation (SFLC) assays and coimmunoprecipitation (Co-IP) assays determined that Os7, MOC1, or Ls protein could interact with MtNSP2 (Fig. 3a–c). Further investigation showed that Os7, MOC1, or Ls protein could repress the transcriptional activity of the MtNSP1/MtNSP2 complex on *MtNIN* (Fig. 3d). In soybean, there are two NSP2 members, GmNSP2a and GmNSP2b, and two NSP1 members, GmNSP1a and GmNSP1b [40]. We found Os7, MOC1 or Ls protein could interact with both GmNSP2a and GmNSP2b (Additional file 1: Fig. S5a–f). In addition, we determined that MOC1 and Os7 could also interact with OsNSP2 in

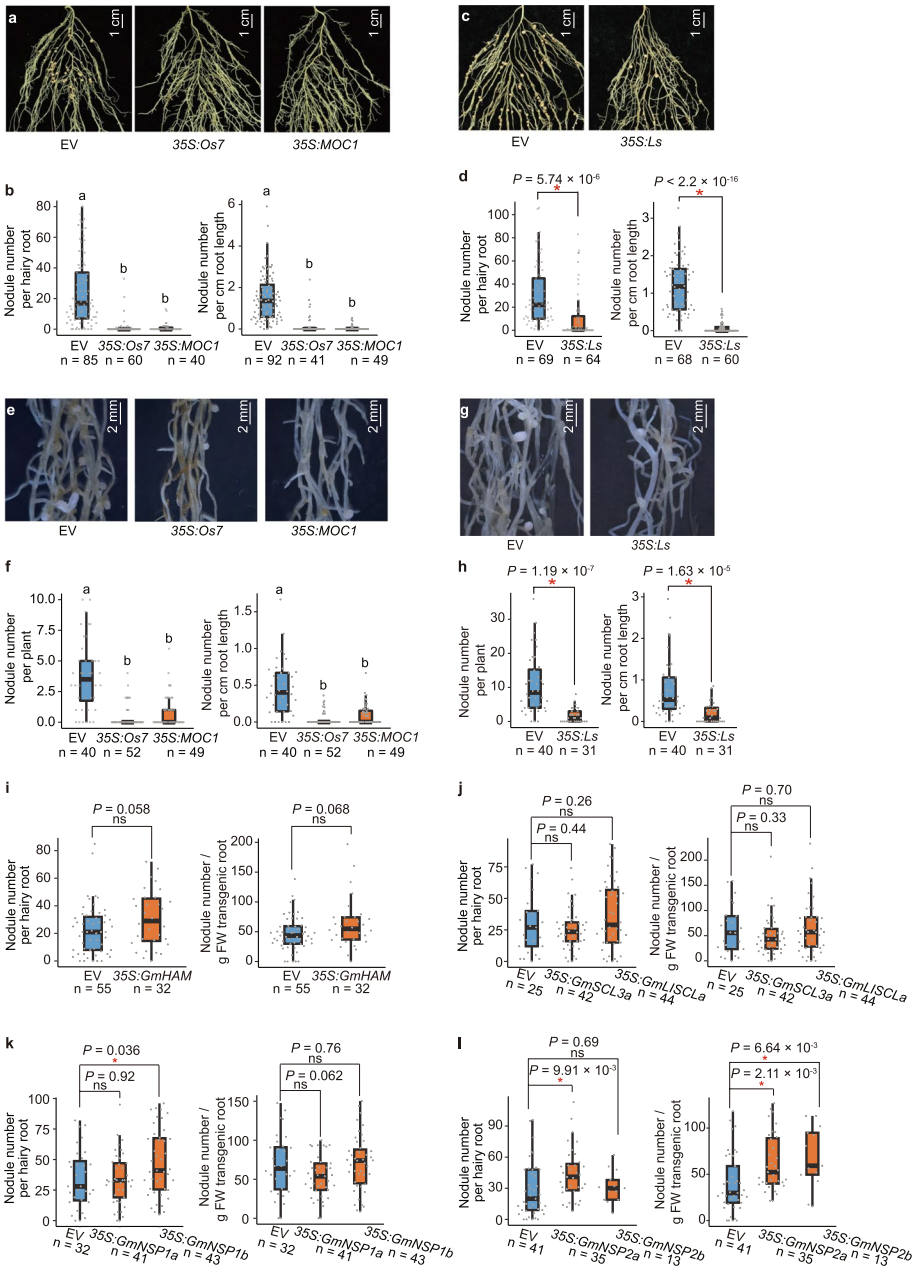

Fig. 2 Effects of Leguminosae-loss *Lateral suppressor* gene on RNS. **a**, **c** Nodulation phenotypes produced by transformation of *Os7* and *MONOCULM 1* (*MOC1*) (**a**) and *Lateral suppressor* (*Ls*) (**c**) in soybean hairy root. Bar, 1 cm. **b**, **d** Number of nodules per hairy root and number of nodules per root length (cm) after the transformation of *Os7* and *MOC1* (**b**) and *Ls* (**d**) in soybean. **e**, **g** Nodulation phenotypes produced by transformation of *Os7* and *MOC1* (**e**) and *Ls* (**g**) in *M. truncatula* hairy root. Bar, 2 mm. **f**, **h** Number of nodules per plant and number of nodules per root length (cm) after transformation of *Os7* and *MOC1* (**f**) and *Ls* (**h**) in *M. truncatula*. **i**–**l** Number of nodules per hairy root and number of nodules per root fresh weight (FW, g) of overexpression *GmHAM* (**i**), *GmSCL3a* and *GmLISCLa* (**j**), *GmNSP1a* and *GmNSP1b* (**k**), and *GmNSP2a* and *GmNSP2b* (**l**) in soybean. Hairy roots were collected at 21 days after inoculation (DAI) for nodulation phenotype. EV indicates transformation with an empty vector. Asterisks indicate significant differences from the empty vector controls and the letter "ns" indicates that there is no significant difference compared to EV. Statistical significance was determined by a two-sided *t*-test in **d**, **h**, and **i**–**l**. The lowercase letters in **b** and **f** indicate significant differences between samples [analysis of variance (ANOVA), least significant difference (LSD) method]

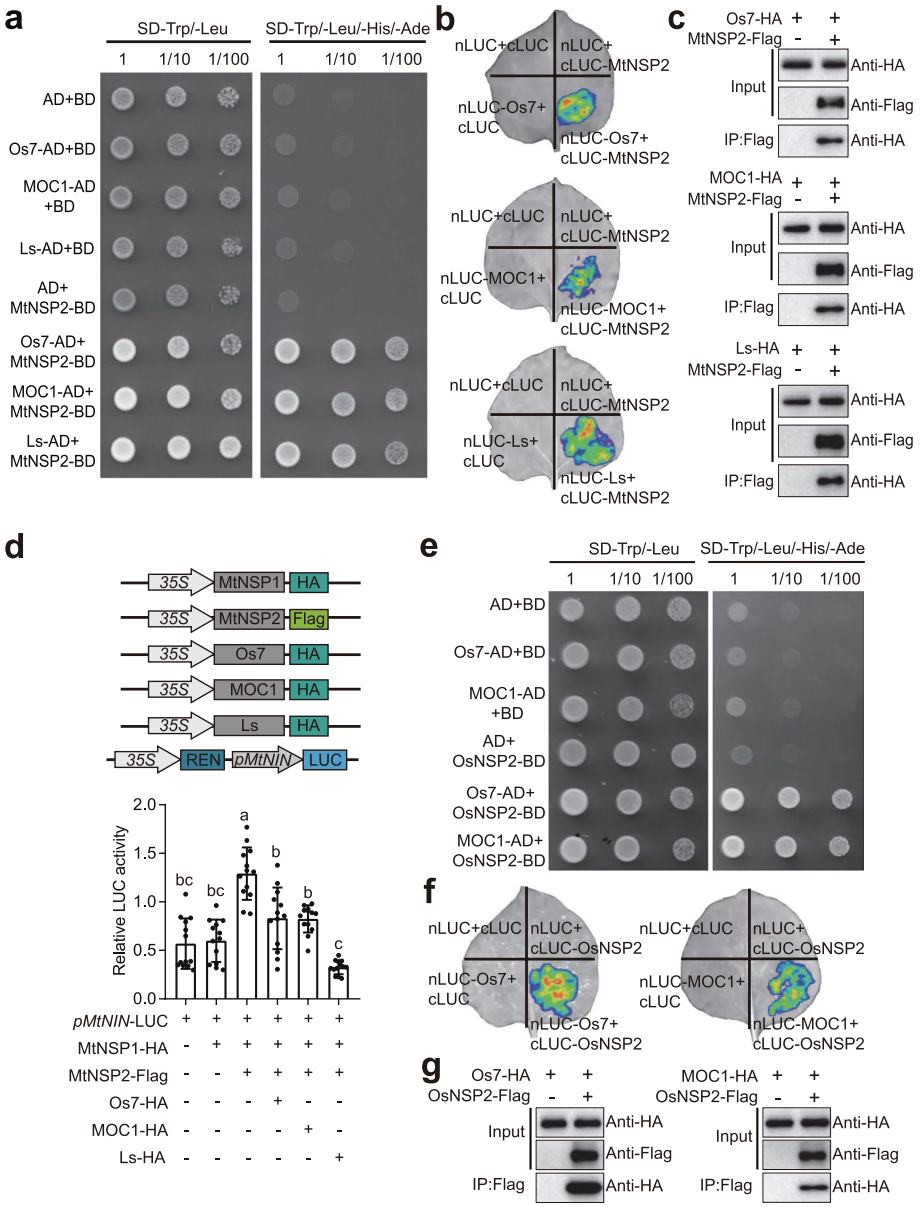

Fig. 3 Lateral suppressor protein interacts with MtNSP2 to inhibit the expression of *NIN*. **a** Yeast two-hybrid assays between MtNSP2 and Os7, MOC1 or Ls protein. AD, activation domain; BD, binding domain. **b** Split firefly luciferase complementation (SFLC) assays between MtNSP2 and Os7, MOC1, or Ls protein. **c** Coimmunoprecipitation (Co-IP) assays between MtNSP2 and Os7, MOC1 or Ls protein. **d** The effect of Os7, MOC1, or Ls protein on inhibiting MtNSP2-MtNSP1 activation to induce *MtNIN* transcription. The lowercase letters indicate significant differences between samples [ANOVA, LSD method]. REN, Renilla luciferase; LUC, firefly luciferase. **e** Yeast two-hybrid assays between OsNSP2 and Os7 or MOC1. **f** SFLC assays between OsNSP2 and Os7 or MOC1. **g** Co-IP assays between OsNSP2 and Os7 or MOC1

rice (Fig. 3e–g). However, no interaction was detected between Os7, MOC1, or Ls protein and GmNSP1 by the yeast two-hybrid system (Additional file 1: Fig. S5g).

Additionally, phosphorylated CYCLOPS, a DNA binding transcriptional activator, is known to regulate *NIN* transcription by binding to its promoter in *Lotus japonicus* [24, 41], and the replacement of phosphorylation serine sites (S50 and S154) with aspartate

(D) (S50D and S154D, CYCLOPS-DD) could result in a gain-of-function activity of CYCLOPS [41]. We discovered that Os7, MOC1, or Ls protein not only interact with CYCLOPS (Fig. 4a, b), but also exert additional repression on the transcriptional activation of *LjNIN* by CYCLOPS-DD (Fig. 4c). Hence, our results suggest that Os7, MOC1, or Ls protein interacts with both NSP2 and CYCLOPS, inhibiting *NIN* transcription and repressing nodulation in Leguminosae.

Lateral suppressor protein interacts with cathepsin H

The way in which cells adapt to gene loss can be linked to the functional effect of the lost gene [42–45]. To identify other factors with which Os7, MOC1 or Ls protein interact that affect nodulation in Leguminosae, we screened a yeast two-hybrid library in soybean using either MOC1, Os7, or Ls protein as bait, respectively. We found that FW2.2-like (FWL), nodulin-22, nodulin-20a, nodulin-C51, and cathepsin H (GmCTSH) proteins could interact with each bait (Additional file 1: Fig. S6a). Nodulin genes by definition are highly expressed in nodules [46] and the FWL family is essential for soybean nodule organogenesis [47, 48]. *GmCTSH* encodes a cysteine proteinase with cathepsin-H-like activity belonging to the plant papain-like cysteine protease (PLCPs) family [49]. Some cysteine proteinase families, such as legumain from clan CD [50], papain from clan CA [51], and cysteine proteinase with cathepsin-L-like activity [52] have been reported to be involved in nodule development, which provides insights into *GmCTSH*'s putative symbiotic functions.

We confirmed the interaction between Os7, MOC1, or Ls protein and GmCTSH using yeast two-hybrid assays, SFLC assays, and Co-IP assays (Fig. 5a–c). We also detected an interaction between Os7, MOC1, or Ls protein and CTSH in *M. truncatula* (the best hit of GmCTSH in *M. truncatula*, termed MtCTSH) (Fig. 5d, e). To investigate whether

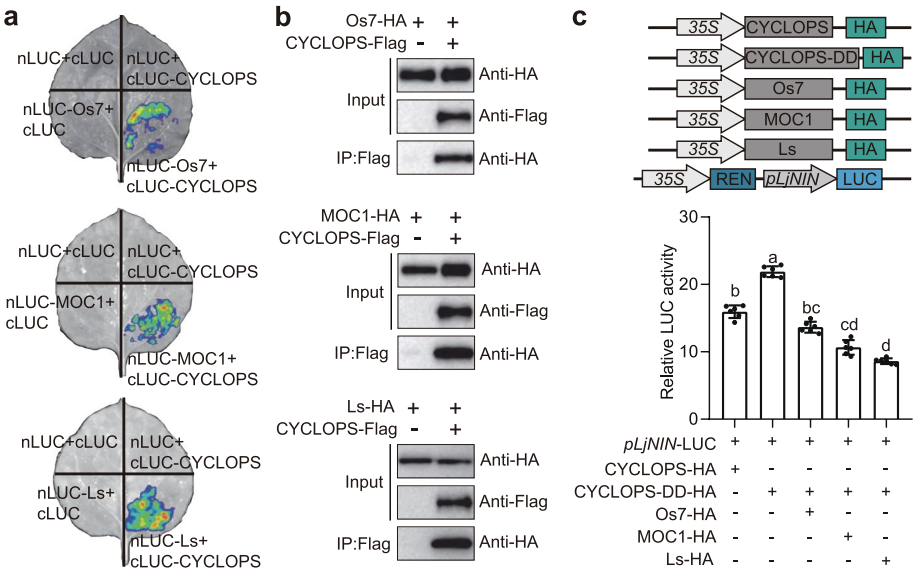

Fig. 4 Lateral suppressor protein interacts with CYCLOPS to inhibit the expression of *NIN*. **a** SFLC assays between CYCLOPS and Os7, MOC1, or Ls protein. **b** Co-IP assays between CYCLOPS and Os7, MOC1 or Ls protein. **c** The effect of Os7, MOC1, or Ls protein on the inhibition of *LjNIN* transcription activation by CYCLOPS-DD. The lowercase letters indicate significant differences between samples [ANOVA, LSD method]

CTSH is involved in nodulation, we specifically knocked down *GmCTSH* in soybean and *MtCTSH* in *M. truncatula*, and observed a significant reduction in nodule numbers in both *GmCTSH*-RNAi and *MtCTSH*-RNAi transgenic lines (Fig. 6a–f).

Phylogenetic analysis showed that CTSH was a species-conserved orthologous group (Additional file 1: Fig. S6b, c, Additional file 2: Table S4). Further investigation showed that of the 5567 conserved orthologous groups in the plant species we sampled, more than half (2799 of the 5567) were differentially expressed following infection with *Bradyrhizobium diazoefficiens* strain USDA110 in soybean [53] (Additional file 2: Table S4). This suggests that conserved orthologous groups could be candidates for subsequent investigation by reverse genetics in multiple species.

Therefore, it is likely that in Leguminosae, without *Lateral suppressor* gene, NSP1-NSP2 complex and calcium- and calmodulin-dependent protein kinase (CCaMK)-CYCLOPS complex activate *NIN* expression by binding to its promoter (Fig. 7). When *Lateral suppressor* gene is in the Leguminosae, it negatively affects nodulation through several different pathways, including interacting with NSP2, CYCLOPS and CTSH (Fig. 7).

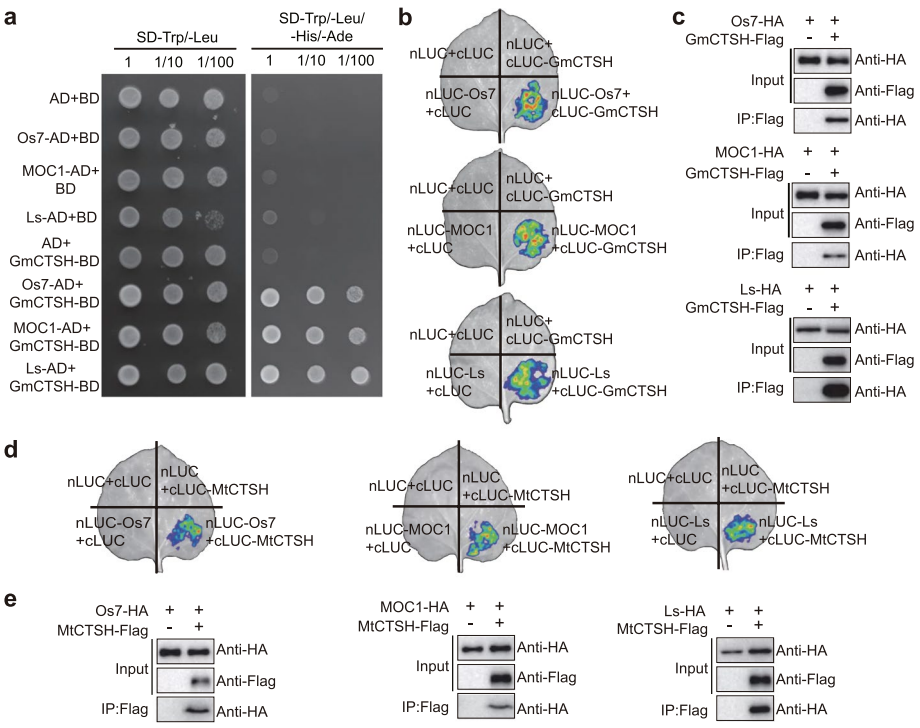

Fig. 5 Lateral suppressor protein interacts with soybean and *M. truncatula* cathepsin H (CTSH) protein. **a** Yeast two-hybrid assays between GmCTSH and Os7, MOC1 or Ls protein. **b** SFLC assays between GmCTSH and Os7, MOC1 or Ls protein. **c** Co-IP assays between GmCTSH and Os7, MOC1 or Ls protein. **d** SFLC assays between MtCTSH and Os7, MOC1 or Ls protein. **e** Co-IP assays between MtCTSH and Os7, MOC1, or Ls protein

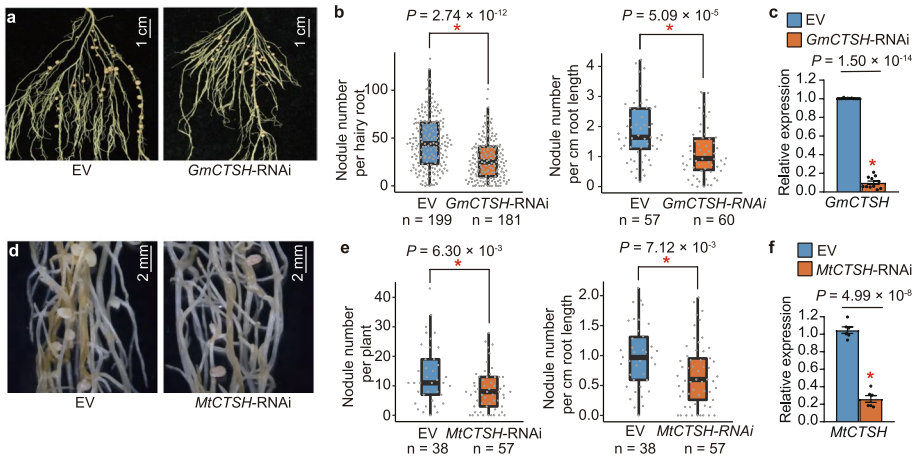

Fig. 6 *CTSH* affects RNS in Leguminosae. **a** Nodulation phenotypes produced by knocking down *GmCTSH* in soybean. Bar, 1 cm. **b** Number of nodules per hairy root and number of nodules per root length (cm) after *GmCTSH*-RNAi transformation in soybean. **c** qRT-PCR analysis of *GmCTSH* expression in transgenic soybean hairy roots carrying the EV and *GmCTSH*-RNAi. **d** Nodulation phenotypes produced by knocking down *MtCTSH* in *M. truncatula*. Bar, 2 mm. **e** Number of nodules per plant and number of nodules per root length (cm) after *MtCTSH*-RNAi transformation in *M. truncatula*. **f** qRT-PCR analysis of *MtCTSH* expression in transgenic *M. truncatula* hairy roots carrying the EV and *MtCTSH*-RNAi. EV indicates transformation with an empty vector. Asterisks indicate significant differences from the empty vector controls. Statistical significance was determined by a two-sided *t*-test in **b**, **c**, **e**, and **f**. The expression of soybean *Actin* gene or *M. truncatula Ubiquitin* gene was used as a normalization control. Data represent the mean ± s.e. in **c** and **f**. Hairy roots were collected at 21 DAI for expression experiments

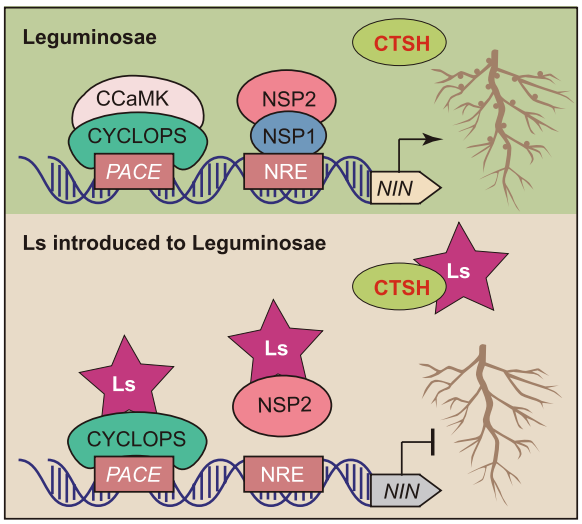

Fig. 7 Working model of *Lateral suppressor* gene inhibiting nodulation when introduced in Leguminosae. In Leguminosae, the NSP1-NSP2 complex and CCaMK-CYCLOPS complex activate *NIN* expression by binding to its promoter. When the *Lateral suppressor* gene, the Leguminosae-loss gene introduced in Leguminosae, it inhibits *NIN* transcription by interacting with NSP2 and CYCLOPS resulting in almost non-nodulation. Besides, Lateral suppressor protein can interact with proteins of other functional RNS genes, such as CTSH, to affect RNS in Leguminosae. *PACE*, *Predisposition-associated Cis-regulatory Element*; NRE, NODULATION RESPONSIVE ELEMENT

Discussion

Previous studies have determined that members of several different GRAS subfamilies play important roles in RNS, including *NODULATION SIGNALING PATHWAY 1* (*NSP1*) [54], *NSP2* [55], *DELLA* [56, 57], *Hairy Meristem 4* (*HAM4*) [58], *SHORTROOT-SCARECROW* (*SHR-SCR*) module [59], and *Scarecrow like 13 (SCL13) Involved in Nodulation* (*SIN1*) [60]. Interestingly, here we found that the ectopic expression of the *Lateral suppressor* gene, a Leguminosae-loss gene and that also belongs to GRAS subfamilies, almost completely inhibited nodulation in Leguminosae. Our results suggested that the ancestor of Os7, MOC1, or Ls protein may interact with NSP2, CYCLOPS, and CTSH to repress the transcription of *NIN* to inhibit the nodulation; whereas, loss of the *Lateral suppressor* gene in Leguminosae released the inhibition and provided a fundamental base for RNS origination (Fig. 7).

The evolutionary model of the *Lateral suppressor* gene in RNS may also be supported by the functions of the *Lateral suppressor* gene and its interaction proteins. *MOC1* can be targeted for degradation by *Tillering and Dwarf 1* (*TAD1*) [61–63], which encodes a Cdh1-type activator of APC/C that is orthologous to *CELL CYCLE SWITCH52* (*CCS52A*) in *M. truncatula*, a gene that is indispensable for symbiotic cell differentiation [64]. In addition, MOC1 is protected from degradation by binding to the DELLA protein SLENDER RICE 1 (SLR1) [65]; the *M. truncatula* DELLA protein encoded by *MtDELLA1*, *MtDELLA2*, and *MtDELLA3* forms a complex with Interacting Protein of DMI3 (IPD3, orthologue of *Lotus japonicus* CYCLOPS) and NSP2 to positively regulate rhizobial symbiosis [56, 57]. Furthermore, MOC1 and MONOCULM 3 (MOC3) physically interact to regulate tiller bud outgrowth through upregulating the expression of *FLORAL ORGAN NUMBER1* (*FON1*), the homolog of *CLAVATA1* in rice [66]; and *CLAVATA1* homologs *nodule autoregulation receptor kinase* (*NARK*) in soybean, *HYPERNODULATION ABERRANT ROOT FORMATION1* (*HAR1*) in *Lotus*, and *SUPER NUMERIC NODULES* (*SUNN*) in *M. truncatula*, are needed for control of the autoregulation of nodulation (AON) pathway [67–69]. The similar gene regulatory networks of the *Lateral suppressor* gene in Leguminosae and non-Leguminosae and the fact that the Lateral suppressor protein could interact with functional RNS genes may indeed suggest that the ancestor of the *Lateral suppressor* gene may also be involved in nodulation.

Additionally, the function of the *Lateral suppressor* gene as a transcriptional factor in RNS remains elusive. In *Arabidopsis*, SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE (SPL9) directly binds and represses the transcription of *LAS* [70, 71] while CUP-SHAPED COTYLEDON 2 (CUC2) and B-type ARABIDOPSIS RESPONSE REGULATOR1 (ARR1) function as positive regulators of *LAS* expression [70, 72, 73]. Besides, LAS binds directly to the promoter of *GA 2-oxidase 4* (*GA2ox4*) to induce its expression, which generates a low-GA content region [71]. In rice, *Oryza sativa homeobox 1* (*OSH1*) and *OsTB1* might be possible downstream genes regulated by MOC1 as their reduced expression in the *moc1* mutant [29]. Taking these results together, the *Lateral suppressor* gene integrates multiple genes to regulate axillary meristem initiation. Thus, although it has been shown that Lateral suppressor protein can interact with NSP2, CYCLOPS, and CTSH in Leguminosae to repress nodulation, how the *Lateral*

suppressor gene regulates RNS functional genes as a transcriptional factor needs to be further investigated.

Divergence of any phenotype involves changes to the gene repertoire underlying it, including the removal of some genes from participation in the phenotype. If the only role of the gene was its contribution to the phenotype in which it no longer was needed, the gene would no longer contribute to organismal fitness. This would release it from purifying natural selection and allow it to accumulate mutations, likely resulting in it being pseudogenized and/or deleted. For example, the loss or pseudogenization of the master nodulation transcription factor, *NIN*, from diverse non-nodulating lineages in the NFNC following environmentally-induced cessation of nodulation constitutes the empirical evidence for the single gain, massive parallel loss theory of nodulation evolution [22, 23]. Gene loss can also be driven by positive selection. When a phenotype is conserved across taxa but genes contributing to it are absent from a lineage, that lineage presumably has found an alternative way, involving different genes, to maintain the phenotype. The absence of genes representing the Ls class of GRAS transcription regulators in Leguminosae means that the critical roles of these genes in other flowering plants, such as control of lateral shoot formation [30], have been taken over by other genes in Leguminosae, making *Lateral suppressor* gene evolutionarily dispensable.

Our study demonstrated that loss of the *Lateral suppressor* gene was associated with the evolution of RNS in Leguminosae, however, there remain many other unanswered questions. For example, genes belonging to the Ls subfamily are found in members of the other three orders of the NFNC in both nodulating and non-nodulating species. Interestingly, when the expression of the *Lateral suppressor* gene was investigated using reported transcriptome data [23, 74] (Additional file 1: Table S5) in the actinorhizal plant *Datisca glomerata* and *Parasponia andersonii*, it was found that although they showed low expression in most tissues, but exhibited significantly reduced expression value in inoculated nodules comparing with the root (Additional file 1: Tables S6, 7), indicating the *Lateral suppressor* gene may suppress the release of rhizobia from infection threads into symbiotic cells. However, further detailed functional investigation works are needed to better understand the role of the *Lateral suppressor* gene in nodulation in actinorhizal plants. Besides, although ectopic expression of *SDI2* and other GRAS members in soybean did not cause the reduction number of nodules (Fig. 2i–l, Additional file 1: Fig. S2a–c), it is possible that ectopic expression of the *Lateral suppressor* gene would lead to dysfunction. These questions provide opportunities for further research on the mechanism of *Lateral suppressor* gene function in Leguminosae and non-Leguminosae. Moreover, gene gains and losses are not the only drivers, as gene expression variations caused by mutations of *cis* or *trans* elements in the promoters and domain changes caused by amino acid mutations also result in significant phenotypic changes [59, 75, 76]. For example, NIN-binding nucleotide sequences were only found in leguminous *ASYMMETRIC LEAVES 2-LIKE 18* (*ASL18*) introns, not in homologues of non-Leguminosae [75, 76]. Besides, an AT1-box and an enhancer, driving the cortical expression of *SCR*, were typically found within 100 bp of each other in the *SCR* promoters from Leguminosae, whereas they were further from each other or one was absent in the *SCR* promoters from species outside the NFNC [59]. Thus, the genes triggering the evolution of RNS may be more complicated than we have found to date.

Conclusions

In our study, we disclosed the role of gene loss in the evolution of RNS in Leguminosae. First, we determined that two orthogroups were lost in the evolution of Leguminosae. Subsequently, we validated the roles of gene loss in RNS by transgenic experiments. Notably and surprisingly, when the *Lateral suppressor* gene, the lost gene in Leguminosae during evolution, was introduced from non-RNS species (rice and tomato) into soybean and *Medicago* (two typical Leguminsoae plants), the transgenic plants completely lost the RNS ability. Furthermore, through a systematic functional analysis, we revealed the clear working model of the *Lateral suppressor* gene in inhibiting RNS: interacts with NSP2 and CYCLOPS and suppresses the transcript activity of NSP2/NSP1 and CCaMK/CYCLOPS complex on *NIN*. We also revealed a novel RNS functional gene, *CTSH*, by screening Lateral suppressor protein-interacting proteins. In short, our study dissected gene loss events in the evolution of RNS in Leguminosae and suggests that loss of the *Lateral suppressor* gene plays a vital role in Leguminosae harboring the RNS ability.

Methods

Plant and materials

Glycine max [L.] Merr., cv. Williams 82 was used for RNA isolation and hairy root transformation. *Bradyrhizobium diazoefficiens* strain USDA110 was used for soybean root inoculation. *M. truncatula* Jemalong A17 and *Sinorhizobium meliloti* (*S. meliloti*) strain 1021 carrying pXLGD4 were used in the *M. truncatula* inoculation assays. *Agrobacterium rhizogenes* (*A. rhizogenes*) strains K599 and AR1193 were used for hairy root transformation in soybean and *M. truncatula* respectively. *Saccharomyces cerevisiae* strain AH109 was used for the yeast two-hybrid assay and the yeast two-hybrid mating system. *Agrobacterium tumefaciens* (*A. tumefaciens*) GV3101 was used for expression in *Nicotiana benthamiana* (*N. benthamiana*).

Growth condition

All the soybean materials are grown in a chamber with a 16-/8-h day/night cycle at 28 °C and 60% relative humidity. The inoculated plants were supplied with B&D nutrient solution ($CaCl_2$, 1 mM; KH_2PO_4, 0.5 mM; ferric citrate, 10 μM; $MgSO_4$, 0.25 mM; K_2SO_4, 0.25 mM; $MnSO_4$, 1 μM; H_3BO_3, 2 μM; $ZnSO_4$, 0.5 μM; $CuSO_4$, 0.2 μM; $CoSO_4$, 0.1 μM; Na_2MoO_4, 0.1 μM) [77] or tap water every other time.

M. truncatula materials were grown in pots filled with a 1:1 mixture soil of vermiculite and perlite. The plants were incubated in controlled-environment rooms at 22 ℃/18 ℃ (day/night), with a light period of 16 h/8 h (day/night), a light intensity of 250 $\mu mol \cdot m^{-2} \cdot s^{-1}$, and a relative humidity of 70%. The inoculated plants were supplied with nitrogen-free water.

Identification of orthologous groups

To cluster gene families and investigate the phylogenetic relationships of these studied species, we selected 48 species with high-quality genomes requiring that their scaffold N50 was longer than 95 kb and that the assembled genomic sequences cover more than 65% of the estimated genome size (Additional file 2: Table S1). For the phylogenomic

analyses, we first filtered the protein sequences with the following criteria: if a gene had multiple encoded protein isoforms, only the longest protein isoform was retained; if the protein had less than 10 amino acids, it was removed. An all-against-all BLASTP (v2.7.1 +) [78, 79] was performed using the protein sequences from the 48 species, and pairwise similarities were obtained. The results were further filtered: query-subject-pairs with sequence similarities below 20% identity were removed; query-subject-pairs with BLAST e-values higher than 1e-5 were removed; and alignment coverage less than 30% of either the total query or subject sequence length in the high-scoring pairs (HSPs) were removed [80–83]. Subsequently, by using OrthoFinder (v2.2.6) [84, 85] under default parameters, we identified 40059 orthologous groups (each group contained at least 2 genes). The previously identified orthologous groups were further clustered using OrthoMCL (v2.0.9; MySQL v5.7.12) [86] with the default parameters (percent match cutoff: 50%; *e* value cutoff: 1e-5; inflation: 2; scheme = 4). Among the finally identified orthologous groups, there are 2 orthologous groups with genes present in all non-Leguminosae species and absent in all 16 species of Leguminosae and 5567 conserved orthologous groups with genes existing in at least 47 species.

Species tree construction

To construct the species tree, only orthologous groups with conserved and single/low-copy genes were used. The selection criteria were that all 48 species had genes in the orthologous group, and the genes were single copies in at least 43 species. After selection, a total of 309 orthologous groups were chosen for subsequent species tree construction. If there were two or more copies in one orthologous group in a species, only the copy having the highest similarity with the gene from *Amborella trichopoda* was retained for further analyses. Then we performed phylogenetic analysis for each of these orthologous groups. Forty-two orthologous groups failed to construct trees using coding DNA sequence (CDS) sequences. Therefore, 267 orthologous groups were retained to construct a species tree. Then, alignments of the amino acid sequences from each orthologous group were performed using MAFFT (v7.490) with the default parameters. The generated alignments were adjusted using trimAl (v1.4.rev15) [87] by trimming the low-quality sequences and the poorly aligned regions and then were transformed into nucleotide matrices by PAL2NAL [88]. After trimming and filtering, the nucleotide matrices from each orthologous group were concatenated and used for species tree construction, which was performed with RAxML (v8.2.12) [89] by the maximum likelihood method based on the model GTRGAMMAIX model, which was chosen by ModelTest-NG [90]. MEGA (v11.0.13) [91] and FigTree (v1.4.3) (http://tree.bio.ed.ac.uk/software/figtree/) was used for editing and displaying the species tree.

Gene family phylogenetic analyses

The GRAS members of 7 Leguminosae species were identified by HMMER (v3.1b2) (www.hmmer.org) search by GRAS domain (PF03514) in the Pfam database (v35.0). The other GRAS members of 26 species were extracted from the PlantTFDB database [92]. The subfamily classification is listed in Additional file 2: Table S3. To build the phylogenetic trees, the protein sequences from each gene family were aligned using MAFFT (v7.490) and then subjected to automated trimming by trimAl (v1.4.rev15). The trimmed

alignments were used for phylogenetic tree construction. The phylogenetic trees of GRAS and SDI2 families were constructed by the IQ-TREE (v2.2.0.3) [93]; the model used was chosen by ModelFinder [94]. The phylogenetic tree of the CTSH orthologous group was constructed by RAxML (v8.2.12) [89] or PhyloBayes MPI version 2 [95]. The model was chosen by ModelTest-NG [90]. The phylogenetic tree was edited and visualized by using the online application iTOL v6.6 (Interactive Tree of Life, http://itol.embl.de/) [96].

Validation of the absence of *Lateral suppressor* gene and *SDI2* on the genome of Leguminosae species

To check the absence of Leguminosae-loss genes in Leguminosae, we used the protein sequences of the genes in Ls or SDI2 orthogroups to blast the genome of the 48 species using tblastn and predicated the open reading frame of the best-hit sequences using genewise in wise (v2.4.1) (https://www.ebi.ac.uk/~birney/wise2/). The score value calculated by genewise was used to build a heatmap.

Synteny analysis

We used JCVI utility libraries (v1.2.4) [97] to identify syntenic genes of the *Lateral suppressor* gene or *SDI2* in representative Leguminosae and non-Leguminosae species based on CDS. The syntenic genes of each pair of compared genomes were concatenated together, and the synteny blocks with *Lateral suppressor* gene or *SDI2* genes were selected for plotting.

RNA-seq analysis

We exploited previously reported transcriptome datasets [23, 74] (Additional file 2: Table S5) of *Datisca glomerata* and *Parasponia andersonii*. To analyze transcriptional profiles, we trimmed the sequence reads using trimmomatic (V0.39) [98] and aligned the trimmed reads to the reference genome of *Datisca glomerata* (C. Presl) Baill. (GCA_003255025.1) [22] or *Parasponia andersonii* WU1 (GCA_002914805.1) [23] using HISAT2 (v2.0.0-beta) [99]. Gene expression level inferred by fragments per kilobase per million base pairs sequenced (FPKM) was estimated using StringTie (v1.2.4) [100].

Soybean hairy root transformation and transgenic root analysis

For *35S:SDI2*, *35S:Os7*, *35S:MOC1*, *35S:Ls*, *35S:GmHAM*, *35S:GmSCL3a*, *35S:GmLISCLa*, *35S:GmNSP1a*, *35S:GmNSP1b*, *35S:GmNSP2a*, *35S:GmNSP2b*, *GmNINs*-RNAi, and *GmCTSH*-RNAi constructs, the sequences of candidate genes were amplified from cDNA or DNA templates using KOD Fx Neo polymerase (TOYOBO KFX-201). The appropriate sequences were first cloned and inserted into the Gateway donor vector pDONOR221 using BP ClonaseTM II enzyme mix (Invitrogen 11789–020) and then were recombined into Gateway destination vectors using LR CLONASE II ENZYME MIX (Invitrogen 11791–020). pB7WG2D.1 and pK7GWIWGIIRR destination vectors were used for overexpression and RNA silencing experiments, respectively. The primers used in this study are listed in Additional file 2: Table S8.

For hairy root transformation in soybean, the seedling plants were transformed with *A. rhizogenes* strain K599 carrying the appropriate binary vector based on a previously

described protocol [101, 102]. To inoculate the hairy roots, *B. diazoefficiens* was grown in TY liquid medium containing tryptone 5 g/L, yeast extract 3 g/L, and $CaCl_2 \cdot 2H_2O$ 0.7 g/L at 28 °C for 4–5 days in the presence of the spectinomycin antibiotic to OD_{600} at 0.8 to 1. Bacteria were pelleted and diluted using ddH_2O to an OD_{600} of 0.08 to 0.1. Then, each plant was inoculated with a 30 mL suspension of *B. diazoefficiens*. Nodule numbers, root length, and root fresh weight were scored 21 days after inoculation. To check the gene expression levels, total RNA was extracted using Quick RNA Isolation Kit (Huayueyang 0416–50) according to the manufacturer's instructions.

M. truncatula hairy root transformation and nodulation phenotype observation

For the *MtCTSH*-RNAi construct in *M. truncatula* Jemalong A17, a 200 bp CDS of *Medtr4g125300.1* (*MtCTSH*) was amplified from the cDNA library using Phanta® Max Super-Fidelity DNA Polymerase (Vazyme Biotech) according to the manufacturer's instructions. The PCR products were inserted into pDONR207 using BP ClonaseTM II enzyme mix and then combined with pUB-GWS-GFP to generate the construct using LR CLONASE II ENZYME MIX. The constructs in the destination vector were introduced into *A. rhizogenes* AR1193 by electroporation and then transformed into *M. truncatula* A17 by hairy root transformation. The transgenic plants were transferred to a vermiculite and perlite mixture and inoculated with *S. meliloti* 1021 carrying the pXLGD4 (*lacZ*) plasmid after 1 week. The infection events, nodule numbers, and root length were analyzed at 21 DAI. The infection events were scored using light microscopy (ECLIPSE Ni, Nikon) after histochemical staining of the roots with X-Gal. Digital images were captured using a digital camera mounted on the light microscope. The primers used in this study are listed in Additional file 2: Table S8.

Gene expression assay

A total of 2 μg of RNA was used for first-strand cDNA synthesis using TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TRANS, AT311-02) following the manufacturer's instructions. Real-time PCR was performed on a Roche LightCycler 480 system with LightCycler 480 SYBR Green I Master Mix (Roche, 04887352001) or TransStart Top Green qPCR SuperMix Top Green (TRANS, AQ131-04). Gene expression was normalized to the expression of *Actin* or *Ubiquitin* and the relative expression level was calculated using the $2^{-\Delta\Delta CT}$ method. The primers used in this study are listed in Additional file 2: Table S8.

Yeast two-hybrid assays

The yeast two-hybrid assay was performed using the Matchmaker GAL4 Two Hybrid System 3 (Clontech) according to the yeast protocols handbook (Clontech). The CDSs of *MOC1*, *Os7*, *Ls*, *OsNSP2*, *MtNSP2*, *GmNSP2a*, *GmNSP2b*, *GmNSP1a*, *GmNSP1b*, and *GmCTSH* were fused with the GAL4 activation domain (AD) in pGADT7 (Takara Bio Inc.) or the GAL4 binding domain (BD) in pGBKT7 (Takara Bio Inc.). The relevant primer sequences are given in Additional file 2: Table S8. Different combinations of plasmids were then co-transformed into the yeast (*Saccharomyces cerevisiae*) strain AH109. The cotransformed yeast was screened on the SD-Trp/-Leu and interactions were screened on SD-Trp/-Leu/-His/-Ade at 30 °C.

The yeast two-hybrid mating system was performed according to the Matchmaker® Gold Yeast Two-Hybrid System (Clontech). The full-length of *Os7*, *MOC1*, and *Ls* CDSs were amplified and subcloned and inserted into pGBKT7, and then transformed into yeast strain AH109. Each Os7, MOC1, and Ls protein was used as bait to screen a cDNA library prepared from poly(A)-containing RNA sampled from root and root nodules. Experimental procedures for screening and plasmid identification were performed according to the manufacturer's user guide.

Split firefly luciferase complementation (SFLC) assays

The CDSs of *MOC1*, *Os7*, *Ls*, *OsNSP2*, *MtNSP2*, *GmNSP2a*, *GmNSP2b*, *CYCLOPS*, *GmCTSH*, and *MtCTSH* were cloned and inserted into pCAMBIA-split_cLUC or pCAMBIA-split_nLUC to generate the needed fusion transgenes. The primers used for vector construction are listed in Additional file 2: Table S8. Different combinations of *A. tumefaciens* strain GV3101 cells containing the above plasmids were transformed into *N. benthamiana* leaves, according to a previous study [103]. After 2 days, the *N. benthamiana* leaves were smeared with 1 mM luciferin to measure luciferase intensity in a NightOWL II LB 983 imaging apparatus [104].

Coimmunoprecipitation (Co-IP) assays

The CDSs of *MOC1*, *Os7*, *Ls*, *OsNSP2*, *MtNSP2*, *CYCLOPS*, *GmNSP2a*, *GmNSP2b*, *GmCTSH*, and *MtCTSH* were amplified and then inserted into either the PUC-35S-HA or the PUC-35S-FLAG vector. The primers used for vector construction are listed in Additional file 2: Table S8. Each plasmid combination was introduced into *Arabidopsis* protoplasts and incubated at 22 °C in the dark for 16 h. Total protein was extracted from harvested protoplasts by treatment with extraction buffer (50 mM HEPES, 1 mM EDTA, 150 mM KCl, 0.3% Trion X-100, 1 mM DTT, 1 mM PMSF, and 1 × complete protease inhibitor cocktail (Roche)). Protoplasts were broken down by vortexing and were then incubated for 10 min on ice. The samples were centrifuged at 15000 rcf for 20 min at 4 °C. Partial supernatant (40 μL) was retained for input samples to ensure that proteins were normally expressed in all samples. The remaining supernatant was incubated with FLAG magnetic beads (MBL) for 2 h at 4 °C. The beads were washed six times with a wash buffer. The immunoprecipitates were separated by electrophoresis in 12% SDS–polyacrylamide gels and transferred to a nitrocellulose membrane (GE Healthcare). Proteins were detected by treating the membranes with anti-HA-tagged mAb-HRP-DirectT (1:5000, MBL, M180-7) or anti-DDDDK-tagged mAb-HRP-DirectT antibodies (1:2000, MBL, M185-7).

Transient dual luciferase reporter assay

The remote upstream *cis*-regulatory conserved region with putative cytokinin response elements (CE region) in combination with the −5 kb region is sufficient to induce wild-type-like nodule organogenesis in *M. truncatula* [105]. Therefore, the sequence of the CE region fused to the −5 kb region (*ProNIN*$_{CE\text{-}5\ kb}$ [105]) of *MtNIN* was amplified from *M. truncatula* A17 genomic DNA. 4831 bp promoter sequence of *LjNIN* was amplified from *pNIN:GUS* [106]. These sequences of the promoters of *MtNIN* and *LjNIN* were inserted into the pGreen II 0800-LUC vector to be used as a transcriptional activity

reporter plasmid (10 μg of plasmid per transfection). CYCLOPS-DD (S50D and S154D) sequence was generated using Fast MultiSite Mutagenesis System kit (TRANS, FM201-01). The full-length CDSs of *MtNSP1*, *MtNSP2*, *CYCLOPS*, *CYCLOPS-DD*, *Os7*, *MOC1*, and *Ls* were cloned and inserted into the PUC-35S-Flag or PUC-35S-HA vector, and these plasmids were used as the effector (12 μg per transfection). The primers used in this study are listed in Additional file 2: Table S8. The reporter plasmid and corresponding effector plasmids were co-transformed into *Arabidopsis* protoplasts as described previously [107]. The activities of firefly luciferase (LUC) and *Renilla* luciferase (REN) were examined using a Dual-Luciferase Reporter Assay System kit (Promega, E1960).

Statistical analysis

All details of the statistics applied are provided alongside the figure and corresponding legends. Statistical analyses were performed in R.

Accession numbers

Sequence data from this article can be found in the Phytozome under the following accession numbers: *SDI2*, *LOC_Os05g43040*; *MOC1*, *LOC_Os06g40780*; *Os7*, *LOC_Os02g10360*; *Ls*, *Solyc07g066250.1*; *GmHAM*, *Glyma.U013800*; *GmSCL3a*, *Glyma.11G150200*; *GmLISCLa*, *Glyma.11G138300*; *GmNSP1a*, *Glyma.07G039400*; *GmNSP1b*, *Glyma.16G008200*; *GmNSP2a*, *Glyma.06G110800*; *GmNSP2b*, *Glyma.04G251900*; *GmCTSH*, *Glyma.15G177800*; *MtCTSH*, *Medtr4g125300*; *MtUbiquitin*, *Medtr4g088485*; *GmActin*, *Glyma.18G290800*; *GmNIN1a*, *Glyma.04G000600*; *GmNIN1b*, *Glyma.06G000400*; *GmNIN2a*, *Glyma.02G311000*; *GmNIN2b*, *Glyma.14G001600*; *OsNSP2*, *LOC_Os03g15680*; *MtNSP2*, *Medtr3g072710*; *CYCLOPS*, *Lj2g0008424*; *MtNSP1*, *Medtr8g020840*; *MtNIN*, *Medtr5g099060*; *LjNIN*, *LjContig00081g0005881*.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-024-03393-6.

Additional file 1: Supplementary material: figures S1-S6. Fig. S1 *SULFUR DEFICIENCY INDUCED 2* was lost in Leguminosae. Fig. S2 The nodulation phenotype of *SDI2* overexpression and *NIN*s knockdown transformation. Fig. S3 Expression of *Lateral suppressor* gene transformation. Fig. S4 The infection phenotype and growth phenotype of *Lateral suppressor* gene transformation. Fig. S5 Lateral suppressor protein interacts with GmNSP2. Fig. S6 Phylogenetic analyses of CTSH orthologous group.

Additional file 2: Supplementary tables: Supplementary Tables S1-S8. Table S1. The 48 species used in this study. Table S2. Information of Leguminosae-loss orthologous groups. Table S3. Subfamily classification of GRAS members. Table S4. Information of conserved orthologous groups. Table S5. The list of RNA-seq samples analyzed in this study. Table S6. FPKM value of *dgLs* among different tissues. Table S7. FPKM value of *paLs* among different tissues. Table S8. Primers used in this study.

Additional file 3. Uncropped blot images related to Figs. 3, 4 and 5 and Additional file 1: Figs. S2, S3 and S5.

Additional file 4. Review history.

Acknowledgements

We thank Dr. Jeremy Murray (CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences) and Ms. Anita K. Snyder for revising the manuscript; Dr. Xia Li (Huazhong Agricultural University) for sharing hairy root transformation method in soybean and providing the strains; Dr. Weicai Yang (Institute of Genetics and Developmental Biology, Chinese Academy of Sciences) for providing pK7GWIWGIIRR destination vector for RNAi silencing experiments; Dr. Cao Xu (Institute of Genetics and Developmental Biology, Chinese Academy of Sciences) for providing the sample of *Solanum lycopersicum* Heinz 1706; Ms. Min Shi for her assistance on the functional study.

Peer review information
Shuangxia Jin 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.

Review history
The review history is available as Additional file 4.

Authors' contributions
Z.T. designed the experiments and managed the project. T.L., W.X., S.C., J.J.D., Y.J., and Z.T. performed the phylogenomic analyses and evolutionary analyses. T.L., Z.L., J.F., Y.Y., H.L., S.X., X.Y., Y.L., S.L., M.Z., F.X., J.L., and Z.T. performed functional validation and data analysis. Z.T., J.J.D., S.C., F.X., T.L., J.F., Y.Y., H.L., and W.X. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Funding
This work was supported by: The National Natural Science Foundation of China (grant nos. 32388201 and 32300512), "Strategic Priority Research Program" of the Chinese Academy of Sciences (grant no. XDA24030501), CAS Project for Young Scientists in Basic Research (YSBR-078), and the Xplorer Prize.

Availability of data and materials
Previously published soybean RNA-seq datasets analyzed here were from the Genome Sequence Archive [108] in the National Genomics Data Center [109], China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences under the accession number CRA003430 [53]. The published RNA-seq data for the ***Datisca glomerata*** was downloaded from the CNGB Sequence Archive (CNSA) of the China National Gene Bank DataBase (CNGBdb) under CNP0004055 (Additional file 1: Table S5) [74]. The published RNA-seq data for the ***Parasponia andersonii*** was downloaded from the NCBI SRA database under PRJNA272473 (Additional file 1: Table S5) [23]. The datasets of 48 species analyzed here are noted in Additional file 2: Table S1 [22, 23, 80, 81, 110–145]. The protein alignments used to make the species tree (OMIX003873) [146] and GRAS gene family phylogenies (OMIX003876) [147] have been deposited in the OMIX, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/omix) [109]. No customized code was generated in this study. All bioinformatic tools and software used in this study are cited in the text.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Key Laboratory of Seed Innovation, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. [2]Key Laboratory of Plant Carbon Capture, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China. [3]Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. [4]State Key Laboratory of Plant Diversity and Specialty Crops, Institute of Botany, the Chinese Academy of Sciences, Beijing, China. [5]School of Integrative Plant Science, Sections of Plant Biology and Plant Breeding & Genetics, Cornell University, Ithaca, NY, USA. [6]State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. [7]University of Chinese Academy of Sciences, Beijing, China. [8]Department of Molecular Biology, Max Planck Institute for Biology Tübingen, Tübingen, Germany. [9]Hebei Key Laboratory of Crop Genetics and Breeding, Institute of Cereal and Oil Crops, Hebei Academy of Agricultural and Forestry Sciences, Shi-Jiazhuang, China. [10]Yazhouwan National Laboratory, Sanya, Hainan, China.

Received: 17 April 2024 Accepted: 12 September 2024
Published online: 30 September 2024

References

1. Dent D, Cocking E. Establishing symbiotic nitrogen fixation in cereals and other non-legume crops: The Greener Nitrogen Revolution. Agric Food Secur. 2017;6:1–9.
2. Jhu M-Y, Oldroyd GED. Dancing to a different tune, can we switch from chemical to biological nitrogen fixation for sustainable food security? PLoS Biol. 2023;21: e3001982.
3. Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc Natl Acad Sci USA. 1995;92:2647–51.
4. Sprent JI, Ardley J, James EK. Biogeography of nodulated legumes and their nitrogen-fixing symbionts. New Phytol. 2017;215:40–56.

5. Libourel C, Keller J, Brichet L, Cazalé A-C, Carrère S, Vernié T, Couzigou J-M, Callot C, Dufau I, Cauet S, et al. Comparative phylotranscriptomics reveals ancestral and derived root nodule symbiosis programmes. Nat Plants. 2023;9:1067–80.
6. Wang H, Moore MJ, Soltis PS, Bell CD, Brockington SF, Alexandre R, Davis CC, Latvis M, Manchester SR, Soltis DE. Rosid radiation and the rapid rise of angiosperm-dominated forests. Proc Natl Acad Sci USA. 2009;106:3853–8.
7. Kates HR, O'Meara BC, LaFrance R, Stull GW, James EK, Liu S-Y, Tian Q, Yi T-S, Conde D, Kirst M, et al. Shifts in evolutionary lability underlie independent gains and losses of root-nodule symbiosis in a single clade of plants. Nat Commun. 2024;15:4262.
8. Albalat R, Canestro C. Evolution by gene loss. Nat Rev Genet. 2016;17:379–91.
9. Guijarro-Clarke C, Holland PWH, Paps J. Widespread patterns of gene loss in the evolution of the animal kingdom. Nat Ecol Evol. 2020;4:519–23.
10. Merenyi Z, Krizsan K, Sahu N, Liu XB, Balint B, Stajich JE, Spatafora JW, Nagy LG. Genomes of fungi and relatives reveal delayed loss of ancestral gene families and evolution of key fungal traits. Nat Ecol Evol. 2023;7:1221–31.
11. Puttick MN, Morris JL, Williams TA, Cox CJ, Edwards D, Kenrick P, Pressel S, Wellman CH, Schneider H, Pisani D, et al. The interrelationships of land plants and the nature of the ancestral embryophyte. Curr Biol. 2018;28:733–45.
12. Harris BJ, Clark JW, Schrempf D, Szollosi GJ, Donoghue PCJ, Hetherington AM, Williams TA. Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. Nat Ecol Evol. 2022;6:1634–43.
13. Ferrandez-Roldan A, Fabrega-Torrus M, Sanchez-Serna G, Duran-Bello E, Joaquin-Lluis M, Bujosa P, Plana-Carmona M, Garcia-Fernandez J, Albalat R, Canestro C. Cardiopharyngeal deconstruction and ancestral tunicate sessility. Nature. 2021;599:431–5.
14. Cai L, Arnold BJ, Xi Z, Khost DE, Patel N, Hartmann CB, Manickam S, Sasirat S, Nikolov LA, Mathews S, et al. Deeply altered genome architecture in the endoparasitic flowering plant *Sapria himalayana* Griff. (Rafflesiaceae). Curr Biol. 2021;31:1002–11.
15. Timilsena PR, Barrett CF, Pineyro-Nelson A, Wafula EK, Ayyampalayam S, McNeal JR, Yukawa T, Givnish TJ, Graham SW, Pires JC, et al. Phylotranscriptomic analyses of mycoheterotrophic monocots show a continuum of convergent evolutionary changes in expressed nuclear genes from three independent nonphotosynthetic lineages. Genome Biol Evol. 2023;15: evac183.
16. Hottes AK, Freddolino PL, Khare A, Donnell ZN, Liu JC, Tavazoie S. Bacterial adaptation through loss of function. PLoS Genet. 2013;9: e1003617.
17. Schopfer CR, Nasrallah ME, Nasrallah JB. The male determinant of self-incompatibility in *Brassica*. Science. 1999;286:1697–700.
18. Stein JC, Howlett B, Boyes DC, Nasrallah ME, Nasrallah JB. Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. Proc Natl Acad Sci USA. 1991;88:8816–20.
19. Shimizu KK, Shimizu-Inatsugi RIE, Tsuchimatsu T, Purugganan MD. Independent origins of self-compatibility in *Arabidopsis thaliana*. Mol Ecol. 2007;17:704–14.
20. Hoballah ME, Gübitz T, Stuurman J, Broger L, Barone M, Mandel T, Dell'Olivo A, Arnold M, Kuhlemeier C. Single gene-mediated shift in pollinator attraction in *Petunia*. Plant Cell. 2007;19:779–90.
21. Zufall RA, Rausher MD. Genetic changes associated with floral adaptation restrict future evolutionary potential. Nature. 2004;428:847–50.
22. Griesmann M, Chang Y, Liu X, Song Y, Haberer G, Crook MB, Billault-Penneteau B, Lauressergues D, Keller J, Imanishi L, et al. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. Science. 2018; 361:eaat1743.
23. van Velzen R, Holmer R, Bu F, Rutten L, van Zeijl A, Liu W, Santuari L, Cao Q, Sharma T, Shen D, et al. Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. Proc Natl Acad Sci USA. 2018;115:E4700–9.
24. Cathebras C, Gong X, Andrade RE, Vondenhoff K, Hayashi M, Keller J, Delaux PM, Griesmann M, Parniske M. A novel *cis*-element enabled bacterial uptake by plant cells. bioRxiv. 2022. https://doi.org/10.1101/2022.03.28.486070.
25. Legume Phylogeny Working Group (LPWG). A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. Taxon. 2017;66:44–77.
26. Aarabi F, Kusajima M, Tohge T, Konishi T, Gigolashvili T, Takamune M, Sasazaki Y, Watanabe M, Nakashita H, Fernie AR, et al. Sulfur deficiency–induced repressor proteins optimize glucosinolate biosynthesis in plants. Sci Adv. 2016;2: e1601087.
27. Schumacher K, Schmitt T, Rossberg M, Schmitz G, Theres K. The *Lateral suppressor* (*Ls*) gene of tomato encodes a new member of the VHIID protein family. Proc Natl Acad Sci USA. 1999;96:290–5.
28. Greb T, Clarenz O, Schafer E, Muller D, Herrero R, Schmitz G, Theres K. Molecular analysis of the *LATERAL SUPPRESSOR* gene in *Arabidopsis* reveals a conserved control mechanism for axillary meristem formation. Genes Dev. 2003;17:1175–87.
29. Li X, Qian Q, Fu Z, Wang Y, Xiong G, Zeng D, Wang X, Liu X, Teng S, Hiroshi F, et al. Control of tillering in rice. Nature. 2003;422:618–21.
30. Cenci A, Rouard M. Evolutionary analyses of GRAS transcription factors in angiosperms. Front Plant Sci. 2017;8:273.
31. Mizzotti C, Galliani BM, Dreni L, Sommer H, Bombarely A, Masiero S. ERAMOSA controls lateral branching in snapdragon. Sci Rep. 2017;7: 41319.
32. Ponraj U, Theres K. Keep a distance to be different: axillary buds initiating at a distance from the shoot apical meristem are crucial for the perennial lifestyle of *Arabis alpina*. New Phytol. 2020;227:116–31.
33. Feng J, Cheng L, Zhu Z, Yu F, Dai C, Liu Z, Guo W-W, Wu X-M, Kang C. GRAS transcription factor *LOSS OF AXILLARY MERISTEMS* is essential for stamen and runner formation in wild strawberry. Plant Physiol. 2021;186:1970–84.
34. Jiang Y, Zhang A, He W, Li Q, Zhao B, Zhao H, Ke X, Guo Y, Sun P, Yang T, et al. GRAS family member LATERAL SUPPRESSOR regulates the initiation and morphogenesis of watermelon lateral organs. Plant Physiol. 2023;193:2592–604.

35. Aarabi F, Rakpenthai A, Barahimipour R, Gorka M, Alseekh S, Zhang Y, Salem MA, Brückner F, Omranian N, Watanabe M, et al. Sulfur deficiency-induced genes affect seed protein accumulation and composition under sulfate deprivation. Plant Physiol. 2021;187:2419–34.
36. Liu J, Bisseling T. Evolution of NIN and NIN-like genes in relation to nodule symbiosis. Genes. 2020;11: 777.
37. Schauser L, Roussis A, Stiller J, Stougaard J. A plant regulator controlling development of symbiotic root nodules. Nature. 1999;402:191–5.
38. Fu M, Sun J, Li X, Guan Y, Xie F. Asymmetric redundancy of soybean *Nodule Inception* (*NIN*) genes in root nodule symbiosis. Plant Physiol. 2022;188:477–89.
39. Hirsch S, Kim J, Munoz A, Heckmann AB, Downie JA, Oldroyd GED. GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in *Medicago truncatula*. Plant Cell. 2009;21:545–57.
40. He C, Gao H, Wang H, Guo Y, He M, Peng Y, Wang X. GSK3-mediated stress signaling inhibits legume–rhizobium symbiosis by phosphorylating GmNSP1 in soybean. Mol Plant. 2021;14:488–502.
41. Singh S, Katzer K, Lambert J, Cerri M, Parniske M. CYCLOPS, a DNA-binding transcriptional activator, orchestrates symbiotic root nodule development. Cell Host Microbe. 2014;15:139–52.
42. Rojas Echenique JI, Kryazhimskiy S, Nguyen Ba AN, Desai MM. Modular epistasis and the compensatory evolution of gene deletion mutants. PLoS Genet. 2019;15: e1007958.
43. Szamecz B, Boross G, Kalapis D, Kovács K, Fekete G, Farkas Z, Lázár V, Hrtyan M, Kemmeren P, Koerkamp MJAG, et al. The genomic landscape of compensatory evolution. PLoS Biol. 2014;12: e1001935.
44. Helsen J, Voordeckers K, Vanderwaeren L, Santermans T, Tsontaki M, Verstrepen KJ, Jelier R. Gene loss predictably drives evolutionary adaptation. Mol Biol Evol. 2020;37:2989–3002.
45. Ocana-Pallares E, Williams TA, Lopez-Escardo D, Arroyo AS, Pathmanathan JS, Bapteste E, Tikhonenkov DV, Keeling PJ, Szollosi GJ, Ruiz-Trillo I. Divergent genomic trajectories predate the origin of animals and fungi. Nature. 2022;609:747–53.
46. Su L, Xu C, Zeng S, Su L, Joshi T, Stacey G, Xu D. Large-scale integrative analysis of soybean transcriptome using an unsupervised autoencoder model. Front Plant Sci. 2022;13: 831204.
47. Libault M, Zhang X-C, Govindarajulu M, Qiu J, Ong YT, Brechenmacher L, Berg RH, Hurley-Sommer A, Taylor CG, Stacey G. A member of the highly conserved FWL (tomato FW2.2-like) gene family is essential for soybean nodule organogenesis. Plant J. 2010;62:852–64.
48. Qiao Z, Brechenmacher L, Smith B, Strout GW, Mangin W, Taylor C, Russell SD, Stacey G, Libault M. The Gm*FWL1* (*FW2-2-like*) nodulation gene encodes a plasma membrane microdomain-associated protein. Plant Cell Environ. 2017;40:1442–55.
49. van Wyk SG, Du Plessis M, Cullis CA, Kunert KJ, Vorster BJ. Cysteine protease and cystatin expression and activity during soybean nodule development and senescence. BMC Plant Biol. 2014;14:294.
50. Oh CJ, Lee H, Kim HB, An CS. Isolation and characterization of a root nodule-specific cysteine proteinase cDNA from soybean. J Plant Biol. 2004;47:216–20.
51. Vincent JL, Brewin NJ. Immunolocalization of a cysteine protease in vacuoles, vesicles, and symbiosomes of pea nodule cells. Plant Physiol. 2000;123:521–30.
52. Deng J, Zhu F, Liu J, Zhao Y, Wen J, Wang T, Dong J. Transcription factor bHLH2 represses *CYSTEINE PROTEASE77* to negatively regulate nodule senescence. Plant Physiol. 2019;181:1683–703.
53. Liu T, Liu H, Xian W, Liu Z, Yuan Y, Fan J, Xiang S, Yang X, Liu Y, Liu S, et al. Duplication and sub-functionalization of flavonoid biosynthesis genes plays important role in Leguminosae root nodule symbiosis evolution. J Integr Plant Biol. 2024. https://doi.org/10.1111/jipb.13743.
54. Smit P, Raedts J, Portyanko V, Debelle F, Gough C, Bisseling T, Geurts R. NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. Science. 2005;308:1789–91.
55. Kalo P, Gleason C, Edwards A, Marsh J, Mitra RM, Hirsch S, Jakab J, Sims S, Long SR, Rogers J, et al. Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. Science. 2005;308:1786–9.
56. Fonouni-Farde C, Tan S, Baudin M, Brault M, Wen J, Mysore KS, Niebel A, Frugier F, Diet A. DELLA-mediated gibberellin signalling regulates Nod factor signalling and rhizobial infection. Nat Commun. 2016;7: 12636.
57. Jin Y, Liu H, Luo D, Yu N, Dong W, Wang C, Zhang X, Dai H, Yang J, Wang E. DELLA proteins are common components of symbiotic rhizobial and mycorrhizal signalling pathways. Nat Commun. 2016;7: 12433.
58. Ren B, Wang X, Duan J, Ma J. Rhizobial tRNA-derived small RNAs are signal molecules regulating plant nodulation. Science. 2019;365:919–22.
59. Dong W, Zhu Y, Chang H, Wang C, Yang J, Shi J, Gao J, Yang W, Lan L, Wang Y, et al. An SHR-SCR module specifies legume cortical cell fate to enable nodulation. Nature. 2021;589:586–90.
60. Ripodas C, Clua J, Battaglia M, Baudin M, Niebel A, Zanetti ME, Blanco F. Transcriptional regulators of legume-rhizobia symbiosis: nuclear factors Ys and GRAS are two for tango. Plant Signal Behav. 2014;9: e28847.
61. Xu C, Wang Y, Yu Y, Duan J, Liao Z, Xiong G, Meng X, Liu G, Qian Q, Li J. Degradation of MONOCULM 1 by APC/C^{TAD1} regulates rice tillering. Nat Commun. 2012;3:750.
62. Lin Q, Wang D, Dong H, Gu S, Cheng Z, Gong J, Qin R, Jiang L, Li G, Wang JL, et al. Rice APC/C^{TE} controls tillering by mediating the degradation of MONOCULM 1. Nat Commun. 2012;3:752.
63. Lin Q, Zhang Z, Wu F, Feng M, Sun Y, Chen W, Cheng Z, Zhang X, Ren Y, Lei C, et al. The APC/C^{TE} E3 ubiquitin ligase complex mediates the antagonistic regulation of root growth and tillering by ABA and GA. Plant Cell. 2020;32:1973–87.
64. Vinardell JM, Fedorova E, Cebolla A, Kevei Z, Horvath G, Kelemen Z, Tarayre S, Roudier F, Mergaert P, Kondorosi A, et al. Endoreduplication mediated by the anaphase-promoting complex activator CCS52A is required for symbiotic cell differentiation in *Medicago truncatula* nodules. Plant Cell. 2003;15:2093–105.
65. Liao Z, Yu H, Duan J, Yuan K, Yu C, Meng X, Kou L, Chen M, Jing Y, Liu G, et al. SLR1 inhibits MOC1 degradation to coordinate tiller number and plant height in rice. Nat Commun. 2019;10:2738.

66. Shao G, Lu Z, Xiong J, Wang B, Jing Y, Meng X, Liu G, Ma H, Liang Y, Chen F, et al. Tiller bud formation regulators MOC1 and MOC3 cooperatively promote tiller bud outgrowth by activating *FON1* expression in rice. Mol Plant. 2019;12:1090–102.
67. Krusell L, Madsen LH, Sato S, Aubert G, Genua A, Szczyglowski K, Duc G, Kaneko T, Tabata S, de Bruijn F, et al. Shoot control of root development and nodulation is mediated by a receptor-like kinase. Nature. 2002;420:422–6.
68. Searle IR, Men AE, Laniya TS, Buzas DM, Iturbe-Ormaetxe I, Carroll BJ, Gresshoff PM. Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase. Science. 2003;299:109–12.
69. Schnabel E, Journet EP, de Carvalho-Niebel F, Duc G, Frugoli J. The *Medicago truncatula SUNN* gene encodes a *CLV1*-like leucine-rich repeat receptor kinase that regulates nodule number and root length. Plant Mol Biol. 2005;58:809–22.
70. Tian C, Zhang X, He J, Yu H, Wang Y, Shi B, Han Y, Wang G, Feng X, Zhang C, et al. An organ boundary-enriched gene regulatory network uncovers regulatory hierarchies underlying axillary meristem initiation. Mol Syst Biol. 2014;10:755.
71. Zhang QQ, Wang JG, Wang LY, Wang JF, Wang Q, Yu P, Bai MY, Fan M. Gibberellin repression of axillary bud formation in *Arabidopsis* by modulation of DELLA-SPL9 complex activity. J Integr Plant Biol. 2019;62:421–32.
72. Raman S, Greb T, Peaucelle A, Blein T, Laufs P, Theres K. Interplay of miR164, *CUP-SHAPED COTYLEDON* genes and *LATERAL SUPPRESSOR* controls axillary meristem formation in *Arabidopsis thaliana*. Plant J. 2008;55:65–76.
73. Rossmann S, Kohlen W, Hasson A, Theres K. Lateral suppressor and Goblet act in hierarchical order to regulate ectopic meristem formation at the base of tomato leaflets. Plant J. 2015;81:837–48.
74. Zhang Y, Fu Y, Xian W, Li X, Feng Y, Bu F, Shi Y, Chen S, van Velzen R, Battenberg K, et al. Comparative phylogenomics and phylotranscriptomics provide insights into the genetic complexity of nitrogen-fixing root-nodule symbiosis. Plant Commun. 2024;5: 100671.
75. Schiessl K, Lilley JLS, Lee T, Tamvakis I, Kohlen W, Bailey PC, Thomas A, Luptak J, Ramakrishnan K, Carpenter MD, et al. *NODULE INCEPTION* recruits the lateral root developmental program for symbiotic nodule organogenesis in *Medicago truncatula*. Curr Biol. 2019;29:3657–68.
76. Soyano T, Shimoda Y, Kawaguchi M, Hayashi M. A shared gene drives lateral root development and root nodule symbiosis pathways in *Lotus*. Science. 2019;366:1021–3.
77. Broughton WJ, Dilworth MJ. Control of leghaemoglobin synthesis in snake beans. Biochem J. 1971;125:1075–80.
78. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
79. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421.
80. Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, et al. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat Biotechnol. 2013;31:240–6.
81. Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MTA, Azam S, Fan G, Whaley AM, et al. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol. 2011;30:83–9.
82. Nygaard S, Hu H, Li C, Schiott M, Chen Z, Yang Z, Xie Q, Ma C, Deng Y, Dikow RB, et al. Reciprocal genomic evolution in the ant-fungus agricultural symbiosis. Nat Commun. 2016;7:12233.
83. Yang Y, Smith SA. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. Mol Biol Evol. 2014;31:3081–92.
84. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16:157.
85. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238.
86. Li L, Stoeckert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13:2178–89.
87. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.
88. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;34:W609–12.
89. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
90. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol. 2020;37:291–4.
91. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9.
92. Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, Gao G. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. 2017;45:D1040-45.
93. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.
94. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.
95. Lartillot N, Rodrigue N, Stubbs D, Richer J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst Biol. 2013;62:611–5.
96. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–6.
97. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and collinearity in plant genomes. Science. 2008;320:486–8.
98. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

99. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15.
100. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5.
101. Jian B, Hou W, Wu C, Liu B, Liu W, Song S, Bi Y, Han T. *Agrobacterium rhizogenes*-mediated transformation of *Superroot*-derived *Lotus corniculatus* plants: a valuable tool for functional genomics. BMC Plant Biol. 2009;9: 78.
102. Kereszt A, Li D, Indrasumunar A, Nguyen CD, Nontachaiyapoom S, Kinkema M, Gresshoff PM. *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. Nat Protoc. 2007;2:948–52.
103. Liu S, Fan L, Liu Z, Yang X, Zhang Z, Duan Z, Liang Q, Imran M, Zhang M, Tian Z. A Pd1-Ps-P1 feedback loop controls pubescence density in soybean. Mol Plant. 2020;13:1768–83.
104. Chen H, Zou Y, Shang Y, Lin H, Wang Y, Cai R, Tang X, Zhou J-M. Firefly luciferase complementation imaging assay for protein-protein interactions in plants. Plant Physiol. 2008;146:368–76.
105. Liu J, Rutten L, Limpens E, van der Molen T, van Velzen R, Chen R, Chen Y, Geurts R, Kohlen W, Kulikova O, et al. A remote *cis*-regulatory region is required for *NIN* expression in the pericycle to initiate nodule primordium formation in *Medicago truncatula*. Plant Cell. 2019;31:68–83.
106. Soyano T, Hirakawa H, Sato S, Hayashi M, Kawaguchi M. NODULE INCEPTION creates a long-distance negative feedback loop involved in homeostatic regulation of nodule organ production. Proc Natl Acad Sci USA. 2014;111:14607–12.
107. Yoo S-D, Cho Y-H, Sheen J. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat Protoc. 2007;2:1565–72.
108. Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, Dong L, Zhang Z, Yu C, Sun Y, et al. The genome sequence archive family: toward explosive data growth and diverse data types. Genom Proteom bioinf. 2021;19:578–83.
109. CNCB-NGDC Members and Partners. Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2024. Nucleic Acids Res. 2024;52:D18–32.
110. Amborella Genome Project. The *Amborella* genome and the evolution of flowering plants. Science. 2013;342: 1241089.
111. Mamidi S, Healey A, Huang P, Grimwood J, Jenkins J, Barry K, Sreedasyam A, Shu S, Lovell JT, Feldman M, et al. A genome resource for green millet *Setaria viridis* enables discovery of agronomically valuable loci. Nat Biotechnol. 2020;38:1203–10.
112. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. The B73 maize genome: complexity, diversity, and dynamics. Science. 2009;326:1112–5.
113. McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B, et al. The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. Plant J. 2018;93:338–54.
114. Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, et al. The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Res. 2006;35:D883–7.
115. The International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature. 2010;463:763–8.
116. Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J, et al. A chromosome conformation capture ordered sequence of the barley genome. Nature. 2017;544:427–33.
117. Filiault DL, Ballerini ES, Mandakova T, Akoz G, Derieg NJ, Schmutz J, Jenkins J, Grimwood J, Shu S, Hayes RD, et al. The *Aquilegia* genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history. Elife. 2018;7: e36426.
118. Guo L, Winzer T, Yang X, Li Y, Ning Z, He Z, Teodor R, Lu Y, Bowser TA, Graham IA, et al. The opium poppy genome and morphinan production. Science. 2018;362:343–7.
119. Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Briere C, Owens GL, Carrere S, Mayjonade B, et al. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. Nature. 2017;546:148–52.
120. The Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. Nature. 2011;475:189–95.
121. The Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. Nature. 2012;485:635–41.
122. The French-Italian Public Consortium for Grapevine Genome Characterization. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature. 2007;449:463–7.
123. Wu GA, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F, Perrier X, Ruiz M, Scalabrin S, Terol J, et al. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat Biotechnol. 2014;32:656–62.
124. Motamayor JC, Mockaitis K, Schmutz J, Haiminen N, Livingstone D III, Cornejo O, Findley SD, Zheng P, Utro F, Royaert S, et al. The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. Genome Biol. 2013;14:r53.
125. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 2012;40:D1202–10.
126. Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G, et al. Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol. 2010;28:951–6.
127. Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. and Gray). Science. 2006;313:1596–604.
128. Kang S-H, Pandey RP, Lee C-M, Sim J-S, Jeong J-T, Choi B-S, Jung M, Ginzburg D, Zhao K, Won SY, et al. Genome-enabled discovery of anthraquinone biosynthesis in *Senna tora*. Nat Commun. 2020;11:5875.

129. Hufnagel B, Marques A, Soriano A, Marques L, Divol F, Doumas P, Sallet E, Mancinotti D, Carrere S, Marande W, et al. High-quality genome sequence of white lupin provides insight into soil exploration and seed quality. Nat Commun. 2020;11:492.
130. Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, Liu X, Gao D, Clevenger J, Dash S, et al. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet. 2016;48:438–46.
131. Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, et al. Genome structure of the legume, *Lotus Japonicus*. DNA Res. 2008;15:227–39.
132. Young ND, Debelle F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H, et al. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature. 2011;480:520–4.
133. Hovde BT, Daligault HE, Hanschen ER, Kunde YA, Johnson MB, Starkenburg SR, Johnson SL. Detection of abrin-like and prepropulchellin-like toxin genes and transcripts using whole genome sequencing and full-length transcript sequencing of *Abrus precatorius*. Toxins. 2019;11: 691.
134. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, et al. Genome sequence of the palaeopolyploid soybean. Nature. 2010;463:178–83.
135. Liu Y, Zhang X, Han K, Li R, Xu G, Han Y, Cui F, Fan S, Seim I, Fan G, et al. Insights into amphicarpy from the compact genome of the legume *Amphicarpaea edgeworthii*. Plant Biotechnol J. 2021;19:952–65.
136. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, et al. A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet. 2014;46:707–13.
137. Yang K, Tian Z, Chen C, Luo L, Zhao B, Wang Z, Yu L, Li Y, Sun Y, Li W, et al. Genome sequencing of adzuki bean (*Vigna angularis*) provides insight into high starch and low fat accumulation and domestication. Proc Natl Acad Sci USA. 2015;112:13213–8.
138. Lonardi S, Munoz-Amatriain M, Liang QH, Shu SQ, Wanamaker SI, Lo S, Tanskanen J, Schulman AH, Zhu TT, Luo M-C, et al. The genome of cowpea (*Vigna unguiculata* [L.] Walp.). Plant J. 2019;98:767–82.
139. Ramos AM, Usie A, Barbosa P, Barros PM, Capote T, Chaves I, Simoes F, Abreu I, Carrasquinho I, Faro C, et al. The draft genome sequence of cork oak. Sci Data. 2018;5: 180069.
140. Martinez-Garcia PJ, Crepeau MW, Puiu D, Gonzalez-Ibeas D, Whalen J, Stevens KA, Paul R, Butterfield TS, Britton MT, Reagan RL, et al. The walnut (*Juglans regia*) genome sequence reveals diversity in genes coding for the biosynthesis of non-structural polyphenols. Plant J. 2016;87:507–32.
141. Ye G, Zhang H, Chen B, Nie S, Liu H, Gao W, Wang H, Gao Y, Gu L. De novo genome assembly of the stress tolerant forest species *Casuarina equisetifolia* provides insight into secondary growth. Plant J. 2019;97:779–94.
142. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, et al. The genome of woodland strawberry (*Fragaria vesca*). Nat Genet. 2011;43:109–16.
143. The International Peach Genome Initiative. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet. 2013;45:487–94.
144. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, et al. The genome of the domesticated apple (*Malus* x *domestica* Borkh.). Nat Genet. 2010;42:833–9.
145. Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, Gonzalez VM, Henaff E, Camara F, Cozzuto L, Lowy E, et al. The genome of melon (*Cucumis melo* L.). Proc Natl Acad Sci USA. 2012;109:11872–7.
146. Liu T, Liu Z, Fan J, Yuan Y, Liu H, Xian W, Xiang S, Yang X, Liu Y, Liu S, et al. Loss of *Lateral suppressor* gene is associated with evolution of root nodule symbiosis in Leguminosae. Datasets. OMIX. 2024. https://ngdc.cncb.ac.cn/omix/view/OMIX003873.
147. Liu T, Liu Z, Fan J, Yuan Y, Liu H, Xian W, Xiang S, Yang X, Liu Y, Liu S, et al. Loss of *Lateral suppressor* gene is associated with evolution of root nodule symbiosis in Leguminosae. Datasets. OMIX. 2024. https://ngdc.cncb.ac.cn/omix/view/OMIX003876.

Publisher's Note