The chromosome-level genome assembly of *Astragalus sinicus* and comparative genomic analyses provide new resources and insights for understanding legume-rhizobial interactions

Danna Chang¹,², Songjuan Gao³, Guopeng Zhou¹, Shuhan Deng⁴, Jizeng Jia⁵,*, Ertao Wang⁶,* and Weidong Cao¹,³,*

¹Key Laboratory of Plant Nutrition and Fertilizer, Ministry of Agriculture and Rural Affairs, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China
²Graduate School, Chinese Academy of Agricultural Sciences, Beijing 100081, China
³College of Resources and Environmental Sciences, Nanjing Agricultural University, Nanjing 210095, China
⁴Glitzia Biological Science and Technology, Co, Ltd, Beijing, China
⁵Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China
⁶National Key Laboratory of Plant Molecular Genetics, Chinese Academy of Sciences Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

*Correspondence: Jizeng Jia (jiajizeng@caas.cn), Ertao Wang (etwang@cemps.ac.cn), Weidong Cao (caoweidong@caas.cn)

https://doi.org/10.1016/j.xplc.2021.100263

ABSTRACT

The legume species *Astragalus sinicus* (Chinese milk vetch [CMV]) has been widely cultivated for centuries in southern China as one of the most important green manures/cover crops for improving rice productivity and preventing soil degeneration. In this study, we generated the first chromosome-scale reference genome of CMV by combining PacBio and Illumina sequencing with high-throughput chromatin conformation capture (Hi-C) technology. The CMV genome was 595.52 Mb in length, with a contig N50 size of 1.50 Mb. Long terminal repeats (LTRs) had been amplified and contributed to genome size expansion in CMV. CMV has undergone two whole-genome duplication (WGD) events, and the genes retained after the WGD shared by Papilionoideae species shaped the rhizobial symbiosis and the hormonal regulation of nodulation. The chalcone synthase (CHS) gene family was expanded and was expressed primarily in the roots of CMV. Intriguingly, we found that resistance genes were more highly expressed in roots than in nodules of legume species, suggesting that their expression may be increased to bolster plant immunity in roots to cope with pathogen infection in legumes. Our work sheds light on the genetic basis of nodulation and symbiosis in CMV and provides a benchmark for accelerating genetic research and molecular breeding in the future.

Key words: *Astragalus sinicus*, genome, chalcone synthase (CHS) gene, R genes

Chang D., Gao S., Zhou G., Deng S., Jia J., Wang E., and Cao W. (2022). The chromosome-level genome assembly of *Astragalus sinicus* and comparative genomic analyses provide new resources and insights for understanding legume-rhizobial interactions. Plant Comm. 3, 100263.

INTRODUCTION

The legume species *Astragalus sinicus* (Chinese milk vetch [CMV], 2n = 2x = 16), a winter-growing green manure/cover crop that originates in China, has been widely cultivated in South Asian countries like China, Korea, and Japan for its high nitrogen-fixing ability (Li et al., 2008). Growing CMV in rice soil during the winter fallow season can substitute for 40% of mineral nitrogen fertilizer, maintain high rice yields, improve soil fertility (Xie et al., 2016; Yang et al., 2019; Gao et al., 2020), and take full advantage of natural resources like light, heat, and water (Crews and Peoples, 2004; Voisin et al., 2013). Long-term excessive use of inorganic nitrogen fertilizer has had serious negative impacts, such as decreased soil bacterial and archaeal diversity (Ding et al., 2016) and has caused severe soil acidification of Chinese croplands.
CMV has strict host specificity and forms only indeterminate-type nitrogen-fixing nodules with its rhizobia (Li et al., 2008). *Mesorhizobium huakuii* was the first rhizobial species isolated from CMV and was sequenced in 2014 (Wang et al., 2014). With its smaller size, shorter generation time, and strict symbiosis specificity with *M. huakuii*, CMV can be used as a model plant to study the molecular mechanisms that regulate symbiosis specificity and indeterminate nodule formation. The sequenced genomes of legume species have advanced our understanding of the evolution of symbiotic nitrogen fixation (SNF) (Young et al., 2011; Schmutz et al., 2010, 2014; Pecrix et al., 2018; Kamal et al., 2020; Li et al., 2020). To date, genetic and genome studies in CMV have lagged behind those in other legume species owing to the lack of a reference genome. Limited knowledge on the role of nodulation and CMV genetics motivated us to sequence and annotate the CMV genome.

Rhizobia can actively suppress or escape the plant immune system to avoid identification as pathogens by their compatible hosts, but this occurs at the cost of pathogenic infection (Gourion et al., 2015; to avoid identification as pathogens by their compatible hosts, but Rhizobia can actively suppress or escape the plant immune system to determine symbiosis specificity (Wang et al., 2017; Yang et al., 2017). The host can recognize rhizobia-derived effector proteins via R genes, which activate effector-triggered immunity and determine symbiosis specificity (Wang et al., 2018). *Rj2* and *Rfg1* are allelic genes encoding a typical Toll interleukin-1 receptor (TIR)-nucleotide-binding site (NBS)-leucine-rich repeat (LRR) resistance protein that restricts soybean nodulation with specific strains of *Bradyrhizobium japonicum* and *Sinorhizobium fredii* (Yang et al., 2019). *GmNNL1* encodes a newly discovered R protein that interacts with the nodulation outer protein P effector from *Bradyrhizobium* USDA110 to trigger immunity and inhibit nodulation through root hair infection (Zhang et al., 2021). Nodule cysteine-rich (NCR) peptides, which resemble defensin-like antimicrobial peptides in plant immunity, also determine symbiosis specificity in *Medicago truncatula* (Guerrachi et al., 2014; Liu et al., 2014; Wang et al., 2017; Yang et al., 2017). However, the R genes and NCR peptides in the CMV genome have not been systematically identified. Thus, the symbiosis mechanism and defense processes of CMV remain unclear.

In the present study, the genome of CMV (*Figure 1*) was assembled de novo through a combination of sequencing strategies, including PacBio sequencing, Illumina sequencing, and Hi-C technology. Genome comparisons with other sequenced legume species and transcriptome analyses were performed to obtain fundamental insights into the genetic basis of legume nodulation and rhizobial symbiosis. We aimed to provide a benchmark for studying the genetics and genomics of CMV and exploring the evolution of SNF.

**RESULTS**

**Genome sequencing, de novo assembly, and annotation**

Yijiangzi, a landrace accession of CMV, was used for sequencing and genome assembly (*Figure 2*). Based on 17 k-mer analysis (Supplemental Note 1), the genome size of CMV was estimated to be 625.15 Mb, with a heterozygosity rate of 1.21%, a repetitive sequence ratio of 68.32%, and a Guanine Cytosine (GC) content of 35.91% (Supplemental Figure 1; and Supplemental Table 1). A combination of Illumina short-read sequencing, PacBio long-read sequencing, and Hi-C technology was used to assemble the CMV genome.

PacBio long reads (144.11 Gb, ∼230x coverage; Supplemental Table 2) were obtained using the PacBio Sequel II platform and pre-corrected with Wtdbg2 before use in primary genome assembly (Supplemental Note 1). Three rounds of polishing were performed based on the PacBio long reads using Quiver (Supplemental Note 1), and the assembled sequence was further polished with Pilon using 74.58 Gb of Illumina short reads (Supplemental Table 2 and Supplemental Note 1). This assembled sequence comprised 595.52 Mb, accounting for 95.26% of the estimated genome size, and it had a contig N50 size of 1.50 Mb (*Table 1*). To anchor the scaffolds to chromosomes, 64.62 Gb paired-end Hi-C reads were generated (Supplemental Table 2), with a scaffold N50 of 78.42 Mb (*Table 1*). The LACHESIS package (Supplemental Note 1) was used to anchor 575.78 Mb of scaffold sequences into eight pseudochromosomes (2n = 16), accounting for 96.66% of the assembled genome (Supplemental Table 3). The base accuracy and completeness of the assembled CMV genome were validated using short-read sequence alignment, Benchmarking Universal Single-Copy Orthologs (BUSCO), and Core Eukaryotic Genes Mapping Approach (CEGMA) (Supplemental Note 1). The mapping rate was 97.33%, and the genome coverage was 98.37% (Supplemental Table 4). BUSCO showed that 91.10% of the 1440 single-copy plant orthologs were complete (Supplemental Table 5), and CEGMA showed that the assembled genome completely covered 238 (96.77%) of the 248 core eukaryotic genes (CEGs) (Supplemental Table 6). A Hi-C heat map showed that interactions within chromosomes were more frequent than interactions between chromosomes, suggesting that the Hi-C assembly was of high quality (Supplemental Figure 2). Collectively, these data indicated that our genome assembly had high quality and coverage.

A combination of de novo prediction, homology-based searches, and RNA sequencing (RNA-seq) data was used for genome annotation (Supplemental Note 2). The CMV genome contained 34 253 protein-coding genes, with an average sequence length of 2944.85 bp (Supplemental Table 7). Approximately 91.50% of the protein-coding genes were functionally annotated by searching the NR, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and SwissProt databases (Supplemental Table 8). We also identified noncoding RNA (ncRNA) genes, including 8051 small nuclear RNA (snRNA), 3684 microRNA (miRNA), 1156 rRNA, and 699 tRNA genes (Supplemental Note 2 and Supplemental Table 9).

**LTR insertion and genome size expansion**

Repetitive sequences were identified using a combination of de novo and homology-based approaches with reference to the existing Repbase library (Supplemental Note 2). In total, 59.84% of the CMV genome consisted of repetitive sequences (*Table 1*), and 96.97% of the repetitive sequences were transposable elements.
(TEs), which occupied 58.03% of the genome (Supplemental Table 10). Most TEs were LTR retrotransposons, which made up 45.52% of the genome, and the two most common types were Copia and Gypsy elements, which accounted for 29.73% and 9.02% of the genome, respectively (Supplemental Table 10). We found a strong correlation between genome size and total LTR length in several legume species (Spearman correlation coefficient $r = 0.83$) (Figure 3A), indicating that LTRs played a role in the genome size expansion of these legume species. The numbers and total lengths of LTRs (Supplemental Note 2) were greater in CMV than those in Glycyrrhiza uralensis, Cicer arietinum, M. truncatula, and Medicago sativa, and these LTRs had experienced a dramatic burst 0.5 million years ago (MYA) (Figure 3B), which may have led to the genome size expansion.

**Genome evolution and whole-genome duplications shared by legume species**

To investigate the evolutionary history of CMV, we analyzed the genomes of 19 sequenced legume species and an outgroup species,
Arapadopsis thaliana (Kim et al., 2010; Varshney et al., 2011, 2013; Kang et al., 2014; Schmutz et al., 2014; Yang et al., 2015; Bertioli et al., 2016; Hirakawa et al., 2016; Griesmann et al., 2018; Pecrix et al., 2018; Shen et al., 2018; Kreplak et al., 2019; Lonardi et al., 2019; Chen et al., 2020; Li et al., 2020). The most recent common ancestor (MRCA) of the 20 species contained 22,164 gene families and 154 single-copy genes (Supplemental Figure 3). A maximum-likelihood phylogenetic tree was built using the 154 single-copy genes (see section “materials and methods”). CMV and its close relatives in the inverted repeat-lacking clade (IRLC) (G. uralensis, C. arietinum, Pisum sativum, Trifolium pratense, M. truncatula, and M. sativa) were clustered into one monophyletic group. Species divergence times were estimated using MCMCTree, and time correction was performed using divergence times between four known legume species (see section “materials and methods”). We found that CMV diverged from C. arietinum.
Based on 4-fold synonymous third-codon transversion (4DTv) rates of the duplicated gene pairs, CMV had two peaks, one around 0.6, corresponding to the γ polyploidization event shared by eudicot species, and the second between 0.2 and 0.4, consistent with the whole-genome duplication (WGD) event shared by Papilionoideae species (Li et al., 2013; Kim et al., 2013) (Figure 4B). After this WGD, CMV successively diverged from M. truncatula, Glycine max, and Lupinus albus, consistent with the phylogenetic tree. Following the WGD shared by Papilionoideae species, 4562 genes were retained. To gain functional insights into these genes, we performed KEGG enrichment analysis on the retained genes and found that 2171 genes could be annotated using the KEGG database. A total of 110 and 172 genes were significantly enriched in the terms “plant-pathogen interaction” and “plant hormone signal transduction,” respectively (p < 0.05) (Supplemental Table 11). Retained genes related to plant-pathogen interaction encoded calcium-binding protein family members (CaMKs), calcium-dependent kinases (CaMKs), and cyclic nucleotide-gated ion channels (CNGCs) involved in the early symbiotic signaling pathway (Charpentier et al., 2016; Levy et al., 2004) (Supplemental Table 12). The retained genes related to plant hormone signal transduction were those involved in auxin, cytokinin, abscisic acid, ethylene, gibberellin, and brassinosteroid transport (Supplemental Table 12). These hormones can regulate nodulation alone or through interactions with other hormones (Ding et al., 2008; Ng et al., 2015; Kohlen et al., 2018; McGuiness et al., 2019). Specifically, the DELLla protein, a crucial factor for early symbiotic signaling, was retained in the plant hormone signal transduction pathway following the WGD event. Collectively, these results suggested that the WGD shared by Papilionoideae species contributed to the hormonal regulation of nodulation and to symbiosis with rhizobia in CMV.

**Gene family analysis**

A further comparison of Lotus japonicus, G. uralensis, CMV, C. arietinum, and P. sativum revealed 11 868 gene clusters shared by these five close relatives. In addition, 420 gene families were unique to CMV compared with the other four species (Figure 4C). These CMV-specific gene families were significantly enriched in pathways related to cell cycle control, such as nucleotide excision repair, mismatch repair, homologous recombination, and DNA replication (p < 0.05) (Supplemental Table 13). Rhizobium Nod factors can reactivate the cell cycle during infection and nodule primordium formation (Foucher and Kondorosi, 2000; Yang et al., 1994). CMV forms indeterminate-type nodules, which undergo continuous and dynamic development, and these unique gene families may be associated with the indeterminate nodule organogenesis in CMV. An analysis using CAFÉ (Bie et al., 2006) identified 456 expanded and 164 contracted gene families in CMV compared with the common ancestor of CMV, C. arietinum, P. sativum, T. pratense, M. truncatula, and M. sativa (Figure 4A). The families expanded in CMV were enriched in pathways related to the cell cycle (p < 0.05) (Supplemental Table 14), which may contribute to active bacterial differentiation in the nodule. The expanded gene families were also enriched in starch and sucrose metabolism and nitrogen compound metabolic pathways (p < 0.05) (Supplemental Table 14), which confirmed the genetic basis for CMV as a forage grass (Huang et al., 2020).

### Expansion of CHS genes in legumes

According to the KEGG pathway analysis, the gene families expanded in CMV were enriched in flavonoid biosynthesis (p < 0.05). Flavonoids play multiple roles during legume nodulation: initiating signal exchange between symbiotic partners, inhibiting auxin transport in response to rhizobial inoculation, and acting as Nod signal inducers inside plant roots (Subramanian et al., 2007). We analyzed gene families involved in flavonoid biosynthesis (Figure 5A) in CMV, other legume species, and non-legume species. Key genes involved in flavonoid biosynthesis showed greater expansion in CMV and other legumes than in non-legumes (Supplemental Table 15). Their expression was enriched in specific tissues, primarily the roots and then the stem, whereas only four genes were expressed abundantly in the flower and one in the seed (Figure 5B). The spatial expression patterns of flavonoid synthesis-related genes were consistent with the role of flavonoids in recruiting rhizobia to the roots of legume species (Liu and Murray, 2016).

CHS is the first rate-limiting enzyme in flavonoid biosynthesis (Liu and Murray, 2016), and the CHS gene family is expanded in M. truncatula (Young et al., 2011). Notably, CHS was the most dramatically expanded gene family involved in the flavonoid biosynthesis pathway in CMV (15 copies in CMV but only one or two in non-legume species) (Supplemental Table 15). To gain more insight into CHS gene duplication and legume species evolution, CHS genes were identified in legume and non-legume species. We found a consistently higher copy number of CHS genes in both nodulating and non-nodulating legume species compared with non-legumes species, suggesting that CHS duplication is a common feature of legumes (Supplemental Table 16). It has been proposed that non-nodulating legume species have lost the essential genes for nodulation, such as NODULE INCEPTION (NIN), during evolution (Griesmann et al., 2018). Collectively, these data suggest that CHS expansion was required but not sufficient for nodulation in legumes.

### Table 1. The assembly statistics of the CMV genome.

| Metric                          | Value          |
|--------------------------------|----------------|
| Total assembly size (Mb)        | 595.52         |
| Total contig number             | 1829           |
| Maximum contig length (Mb)      | 11.11          |
| Contig N50 length (Mb)          | 1.50           |
| Contig N90 length (kb)          | 133.56         |
| Total scaffolds number          | 242            |
| Scaffold N50 (Mb)               | 78.42          |
| Scaffold N90 (Mb)               | 62.58          |
| GC content (%)                  | 35.91          |
| Gene number                     | 34 253         |
| Repeat content (%)              | 59.84          |
expressed in the root (Figure 5C). The largest phylogenetic group contained eight genes; six were closely adjacent on pseudochromosome 1, one was located on pseudochromosome 4, and another was found on pseudochromosome 6 (Supplemental Table 17). We further analyzed the homologs of the six CHS genes in *M. truncatula* and found that five CHS genes had homologous genes and showed strong synteny (Figure 5D), suggesting that tandem repeats caused the CHS gene expansion in legume species. One of the 15 CHS genes was most highly expressed in flowers, and two were expressed abundantly in the leaf and stem. Most CHS genes were more highly expressed in the root than in other tissues (Supplemental Table 18). The spatial expression patterns of the CHS genes indicated their functional divergence, and this divergence was consistent with their phylogenetic relationships, suggesting that the expansion of CHS homologs was driven by different requirements for flavonoids across the whole organism. The enrichment of CHS and other flavonoid synthesis genes in the root could provide a high concentration of flavonoids for nodulation signaling.

Identification and spatial expression analysis of R genes

Plant immunity plays a crucial role in establishing and maintaining symbiosis. Plant defense is activated by legume R genes when rhizobia-derived effector proteins are recognized. Thus, the suppression of legume defenses is a prerequisite for successful symbiosis (Gourion et al., 2015). We identified R genes in CMV, its related species, and *A. thaliana* (see section, “materials and methods”). These R genes belong to four major groups: nucleotide-binding site (NBS), NBS-(leucine-rich repeat) LRR, (Toll interleukin-1 receptor) TIR-NBS, and TIR-NBS-LRR (Supplemental Table 19). The number of R genes was much higher in *M. truncatula* (687) than in CMV (178). Because both LRR and non-LRR regions of R genes play a role in determining host specificity (Luck et al., 2000; Dodds et al., 2001), we analyzed the phylogenetic relationships among full-length R-gene sequences from CMV, *M. truncatula*, and *A. thaliana* (Figure 6A). Although a similar number of R genes was detected in CMV (178) and *A. thaliana* (170), R gene sequences in CMV and *M. truncatula* were much more divergent than in *A. thaliana*, suggesting that coevolution with rhizobia has greatly shaped R gene diversity in legumes. Because R genes in legumes coevolved with rhizobia, we speculated that differences in R gene number between CMV and *M. truncatula* could have originated from their specificity for distinct rhizobia.

Identification and spatial expression analysis of R genes

We next examined the expression patterns of R genes across CMV tissues (Figure 6B). Most R genes were expressed primarily in the root and then in the nodule, whereas only a few R genes were enriched in stem, leaf, flower, and seed tissues (Figure 6B), suggesting that R genes in CMV mediate symbiosis and defense processes belowground. We analyzed R gene expression in the roots and nodules of *M. truncatula* and *G. max* using publicly available data. Consistently, more R genes were highly expressed in the root than in the nodule (Supplemental Figure 4). Plant immunity is suppressed to set up symbiosis in the nodule (Gourion et al., 2015; Yu et al., 2019), and R gene expression in root tissue suggested a potential enhancement of R-gene-mediated plant defense to counteract the immunity-suppressive effect caused by the rhizobia.

Rapid evolution of R genes compared with other protein-coding genes

R genes have had rapid birth and death rates in plants as they evolved and interacted with pathogens (Zheng et al., 2016). We compared the evolutionary rate of R genes with that of genome-wide single-copy genes. The backbone of the tree inferred from single-copy genes (Figure 6C) was similar to that of the tree inferred from R genes shared by the five plant species (Figure 6D). However, the branch lengths of the R gene phylogenetic tree were much longer than those of the genome-wide single-copy gene tree, with more substitutions per site, indicating that R genes evolved much faster than the single-copy genes. Later, we examined whether R genes experienced different selection pressure in legumes than in non-legumes by analyzing the Ka/Ks ratios of R genes. However, R-gene Ka/Ks ratios were not significantly different between legume and non-legume species (Supplemental Figure 5), suggesting that the legume-rhizobia interaction did not accelerate the evolutionary rate of R genes.

Identification of NCR peptides

The antimicrobial NCR peptides are essential for bacterial differentiation and intracellular survival in nodule cells (Czemic et al., 2015; Pan and Wang, 2017). These NCR peptides promote indeterminate nodule development and determine symbiosis specificity in *M. truncatula* (Wang et al., 2017). We identified 107 and 787 NCR peptides in CMV and *M. truncatula*, respectively. A strong positive correlation has been reported between the degree of bacteroid elongation and the number of NCR peptides (Montiel et al., 2017). CMV may form elongated (E)-type bacteroids, whereas *M. truncatula* contains remarkably elongated and elongated-branched(EB)-type bacteroids. We analyzed the phylogeny of NCR peptides in CMV and *M. truncatula* and found that the diversity of NCR genes was...
similar between the two species (Supplemental Figure 6). The difference in NCR peptides between CMV and *M. truncatula* suggests that some key NCR peptides may have coevolved with rhizobia to adapt to environmental changes.

Characterization of SNF genes in CMV

We obtained 199 SNF gene sequences and identified their homologs in the CMV genome by tBLASTn using the criteria of e value <1e\(^{-5}\) and identity >60% (Roy et al., 2020). One hundred and...
ninety SNF homologs were identified in CMV, including the common symbiosis (SYM) signaling pathway genes NENA, DMI1, NUP85, NUP133, SYMRK, CCaMK, DELLA, NIN, ERN, NSP1, and NSP2 (Supplemental Table 20). Of the 190 SNF genes, 34, 31, and 54 genes were involved in early nodulation signaling, rhizobial infection, and nodule organogenesis, respectively. Seventeen genes were involved in the autoregulation of nodule number, and 38 genes were related to nodule nitrogen fixation functions such as symbiosome formation (9), bacterial maturation (7), and symbiotic metabolism and transport (22). Ten and four genes were involved in nodule senescence and plant defense, respectively. Two GmRj4 (Tang et al., 2015) and GmRj2 homologs in CMV may be involved in host range restriction. No CMV homolog of NFR1/LYK3 was identified in the tBLASTn search, as these genes are essential for Nod factor perception and are unlikely to be lost in legumes. We next performed a BLASTP search together with synteny analysis (Supplemental Figures 7 and 8), and two tandem CMV Nod factor receptor homolog genes with sequence similarity ranging from 51% to 67% were identified in the CMV genome. In a Ka/Ks ratio analysis, we found that the potential CMV NFR homologs were under purifying selection (Ka/Ks ratio ~0.2; Supplemental Table 21), suggesting potential functional divergence of the Nod factor receptor in CMV.

CMV homologs of 10 SNF genes were not found in our analysis. Four were NCR peptides in M. truncatula, including MtNFS1 and MtNFS2, which determine the symbiosis specificity in M. truncatula, and MtDNF4 and MtDNF7, which control bacterial differentiation. These lost NCR peptides in CMV may affect its bacterial differentiation and symbiosis specificity. GmmiR167, GmmiR172, and LjCLE-RS1, which control nodule number in G. max and L. japonicus, respectively, were also absent in CMV, suggesting an independent nodulation mechanism across the legume species.

DISCUSSION

Plant Communications A. Astragalus sinicus genome sheds light on legume-rhizobial interactions

The high-quality CMV genome provides a benchmark for functional genomics and molecular breeding

The use of CMV is considered to be an effective management practice for establishing a sustainable green manure/cover.
crop–rice rotation in southern China, replacing a very large portion of chemical fertilizer and improving the productivity of paddy fields (Gao et al., 2020). However, unlike other legume species, CMV has not previously been the subject of genetic and genomic research. More than 30 legume genomes have been sequenced, enabling advances in functional genomics research on economically important traits (Varshney et al., 2013; Li et al., 2014; Chen et al., 2019) and SNF evolution (Griesmann et al., 2018; Pecrix et al., 2018). Here, we have successfully assembled the first high-quality, chromosome-scale reference genome of CMV using state-of-the-art sequencing technologies and algorithms. The assembled CMV genome has a contig N50 size of 1.50 Mb, and 96.66% of the assembled sequences are anchored to eight pseudochromosomes. BUSCO showed that 91.10% of the plant orthologs were complete in the assembled genome, and CEGMA showed that the assembled genome completely covered 96.77% of the core eukaryotic genes. These results indicated a higher quality of completeness and continuity than that of recently sequenced genomes of legumes such as G. uralensis (Mochida et al., 2017) and P. sativum (Kreplak et al., 2019). This high-quality CMV genome will serve as an important resource for molecular breeding, genomics, and evolutionary studies of CMV.

### Comparative genomic analyses shed light on the genetic basis of nodulation and symbiosis

Legume-rhizobial interaction begins with signal exchange, which is initiated by flavonoid compounds exuded from the legume roots (Subramanian et al., 2007; Liu et al., 2016). We identified genes related to flavonoid biosynthesis in legume species (including nodulating and non-nodulating legumes) and non-legume species and found that the CHS gene family was larger in legume species. Genes encoding enzymes involved in flavonoid biosynthesis were duplicated and retained after the polyploidy event, which facilitated flavonoid synthesis (Li et al., 2013). Specifically, CHS
plays an indispensable role in flavonoid synthesis because it is the first committed step of the flavonoid pathway. No flavonoid compounds were detected in CHS RNAi roots, and unlike control roots of M. truncatula, flavonoid-deficient roots were unable to initiate nodules (Wasson et al., 2006; Zhang et al., 2009). The number of CHS genes was higher in CMV, and most CHS genes (10 out of 15) had higher expression levels in roots than in other tissues (Supplemental Table 18). The expanded CHS genes also showed higher root expression in soybean, but not in non-legume species (Anguraj Vadivel et al., 2018). These results suggest an essential role for CHS genes in promoting nodulation in legume roots. However, CHS genes have played a role in nodulating and non-nodulating legumes, suggesting that CHS coordinates with other essential nodulation genes to support the emergence of a nodulation phenotype. For example, non-nodulating legume species, such as Nissolia schottii and Cercis canadensis, have lost the NIN gene, which is indispensable for nodulation (Griesmann et al., 2018). WGD events have occurred throughout plant evolution and are important drivers of specialization and of the emergence of novel traits and functions (Jiao et al., 2011). CMV underwent two WGD events, the γ event shared by eudicot species and the WGD event shared by the Papilionoideae. The 58-MYA WGD event shared by most papilionoid legumes is thought to have played a role in the evolution of rhizobial nodulation in M. truncatula and its relatives (Young et al., 2011; Kim et al., 2013; Li et al., 2013). The retained duplicates are frequently involved in processes crucial for root nodule symbiosis establishment, such as symbiotic signaling, nodule organogenesis, rhizobial infection, and nutrient exchange and transport (Li et al., 2013). In the plant-pathogen interaction pathway, some key early symbiotic signaling components, like CNGC and CCaMK, were retained. CNGC and CCaMK modulate calcium spiking, which is an essential part of the signaling cascade that leads to nodule development (Levy et al., 2004; Charpentier et al., 2016). In the plant hormone signal transduction pathway, some key genes related to hormone signals were also retained. DELLA can interact with NSP2-NSP1 and enhance the expression of Nod-factor-inducible genes (Jin et al., 2016). It may also regulate the metabolism of some hormones such as gibberellin and cytokinin upon nodulation (Fonouni-Farde et al., 2016; Dolgikh et al., 2019). Cytokinin signaling can activate or release certain flavonoids in the root; these can interfere with auxin transport, thereby causing auxin accumulation, which leads to the cell divisions that produce the nodule primordium (Ng et al., 2015). Abscisic acid can suppress Nod factor signal transduction and regulate cytokinin induction during the regulation of nodulation (Ding et al., 2008). These genes retained after the WGD shared by Papilionoideae species appear to have a key role in various aspects of the hormonal regulation of nodulation and rhizobial symbiosis.

The balance between legume defense and symbiosis
Host plants have evolved various strategies to dial back their defense systems in order to invite rhizobial infection, colonization, and differentiation (Faulkner and Robatzek, 2012; Gourion et al., 2015; Yu et al., 2019). Plant immunity is reduced or suppressed within the nodules (Berrabah et al., 2015; Wang et al., 2016; Benezech et al., 2020). Likewise, the arbuscular mycorrhizal (AM) symbiosis is also established at the expense of plant immunity (He et al., 2019; Zhang et al., 2020). How legume species cope with soil pathogens despite the attenuated immunity caused by the rhizobial symbiosis is largely unknown. R genes have evolved much faster than many other protein-coding genes. The number of R genes transiently expanded after the WGD at 58 MYA and then underwent a large-scale contraction in legume species after 20 MYA due to diploidization or artificial selection (Zheng et al., 2016). It has been assumed that R genes are expressed in different tissues to recognize various types of pathogens. Interestingly, we did not find a difference in R gene number nor or selection pressure between legume and non-legume species.

We found that more R genes were actively expressed in the root than in the nodule, suggesting that the legumes suppressed defense responses in the nodule to promote symbiosis but promoted immunity in the root to defend against pathogen infection. Consistent with this assumption, M. truncatula nodules were more susceptible than roots to the pathogenic bacterium Ralstonia solanacearum (Benezech et al., 2020). Interestingly, R gene expression was not enriched in the roots of A. thaliana, which has lost the capacity for symbiosis with AM and rhizobia (Tan et al., 2007; Munch et al., 2018). R genes are expressed in different tissues to recognize various types of pathogens, and the root-specific expression of R genes suggested an increased capacity to recognize microbes in CMV, M. truncatula, and G. max. We speculate that this pattern could promote root defense mechanisms to cope with the immune suppression that facilitates rhizobial infection. Rather than different R gene numbers or selection pressures in legumes and non-legumes, the distinct expression patterns of R genes in the root and nodule may have facilitated symbiosis in the legume species. Thus, R genes expressed in roots and nodules exhibit different functions in mediating interactions with pathogens and nitrogen-fixing bacteria.
Astragalus sinicus genome sheds light on legume-rhizobial interactions

the phylogenetic tree. A probabilistic graphical model (PGM) was introduced to calculate the probability of transitions in gene family size from parent to child nodes in the phylogeny. Using conditional likelihoods as the test statistics, the corresponding p-values in each lineage were calculated, and a p-value of 0.05 was used to identify gene families that were significantly expanded or contracted. The species-specific, expanded, and contracted gene families in CMV were then subjected to KEGG functional annotation.

**Phylogenetic tree construction and divergence time estimation**

A total of 154 shared single-copy gene families were used to construct the phylogenetic tree. A maximum likelihood (ML) phylogenetic tree was constructed with RAxML software using these CDS alignments (Stamatakis, 2006). The MCMCTree program in the PAML package v4.7 (Yang, 2007) was used to estimate divergence times among the 20 species with the following main parameters: burn-in = 100,000, sample-number = 100,000, and sample-frequency = 2. Four calibration points were selected: including 20–30 MYA for *C. cajan* and *G. max* (Varshney et al., 2011), 20–30 MYA for *C. arietinum* and *L. japonicus*, 10–20 MYA for *C. arietinum* and *M. truncatula* (Varshney et al., 2013), and ~54 MYA for galegoid and miltielioid legume species (Kim et al., 2010).

**WGD analysis**

We identified collinearity within the CMV genome and between the CMV genome and the genomes of *L. albus*, *G. max*, *C. arietinum*, and *M. truncatula*. We searched for putative paralogs and orthologs within and between genomes by running a BLASTP alignment (with an e-value ≤ 1e-3) for each genome pair. Fourfold degenerate sites were located, and the 4DTv values were calculated using in-house Perl scripts. The 4DTv range associated with the salticoid duplication was determined by plotting the distribution histogram of 4DTv values (Wang et al., 2010). The synteny blocks were identified and visualized using MCscan (http://chibba.agtec.uga.edu/duplication/mcscan/).

**Identification of CHS, R, and NCR peptide genes**

Reference chalcone synthase (CHS) genes from *A. thaliana*, UniProt, or reference reports were used to search for homologs in legume and non-legume species. Each reference gene was used as a BLAST query against the protein sequences of the selected species, and the aligned genes were predicted based on the Pfam domain. Preliminary candidate genes were selected based on the score values of the domain and blast alignments and then aligned using MUSCLE software (Edgar, 2004) to obtain an FAS file. A phylogenetic tree was constructed with TreeBeST and modified with EvolView (Zhang et al., 2012). The legume species were *Asi* (*A. sinicus*), *Msa* (*M. sativa*), *Mtr* (*M. truncatula*), *Ccc* (*C. cajan*), *Mpu* (*M. pudica*), *Gna* (*G. max*), *Liap* (*L. japonicus*), *Araya* (*Arachis hypogaea*), *Aradu* (*A. duranensis*), *Araip* (*A. ipaensis*), *Vigum* (*V. unguiculata*), *Phavu* (*P. vulgaris*), *Tripr* (*T. pratense*), *Lupan* (*Lupinus angustifolius*), *Viga* (*V. radiata*), and *Vigan* (*V. angularis*). The non-legume species were *Atla* (*A. thaliana*), *Vvin* (*Vitis vinifera*), *Tosc* (*T. cacao*), *Osat* (*O. sativa*), *Cucus* (*Crocus sativus*), *Aper* (*A. persica*), and *Lescu* (*Lycopersicon esculentum*).

Reference NBS protein sequences were used to search for homologs in the CMV reference genome, and sequences with scores >80% of the best hit were obtained. The upstream and downstream 5-kb sequences were also obtained to predict gene structures. The amino acid sequences from each genome were searched against the HMM profile of the NBS domain (Pfam PF00931) using HMMER3.0 software with default parameter settings. Searches were also performed for PF01582 (TIR), PF00560 (LRR_1), PF12799 (LRR_4), PF13306 (LRR_5), PF13516 (LRR_6), PF13504 (LRR_7), and PF13855 (LRR_8) domains. The identified protein sequences were analyzed with MUSCLE software (Edgar, 2004), and a phylogenetic tree was constructed with TreeBeST and modified with EvolView (Zhang et al., 2012).

Three sequences (mature peptide, domain, and motif; motif refers to Mardti et al., 2015) were used as BLAST queries against all genes in the CMV and *M. truncatula* genomes. Three sequences were extracted from the PSIBLAST alignment result, which was greater than 4 C.

**Transcriptome data analysis**

RNA sequencing data were aligned to the CMV reference genome using HISAT2 (v2.0.4; Kim et al., 2015) with default parameters. Gene-level read counts were calculated using HTSeq (v0.6.1p1; Kim et al., 2015). DESeq2 (v1.18.1) was used for data normalization (Love et al., 2014).

All the sequence data from this article can be found in the NCBI database under BioProject PRJNA681368.

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at Plant Communications Online.

**AUTHOR CONTRIBUTIONS**

W.C., J.J., and E.W. conceived the project and revised the manuscript. D.C. coordinated the project and wrote the manuscript. S.G., G.Z., and S.D. prepared the samples and performed the analyses of the genome sequence.

**FUNDING**

This research was financially supported by China Agricultural Research System of MOF and MARA (CARS-22), Chinese Outstanding Talents Program in Agricultural Science, Agricultural Science and Technology Innovation Program of CAAS, and China National Crop Germplasm Resources Platform for Green Manure (NIGCR-2021-19).

**ACKNOWLEDGMENTS**

The authors declare no competing interests.

Received: August 14, 2021
Revised: November 2, 2021
Accepted: November 5, 2021
Published: November 8, 2021

**REFERENCES**

Anguraj Vadivel, A.K., Krysiak, K., Tian, G., and Dhaubhadel, S. (2018). Genome-wide identification and localization of chalcone synthase family in soybean (*Glycine max*) [L]. Merr. BMC Plant Biol. 18:325.

Benezech, C., Berrabah, F., Jardinaud, M.F., Le Scornet, A., Mihes, M., Jiang, G., George, J., Ratet, P., Vailleau, F., and Gourion, B. (2020). Medicago-sinorhizobium-ralstonia co-infection reveals legume nodules as pathogen confined infection sites developing weak defenses. Curr. Biol. 30:1–8.

Bertioli, D.J., Cannon, S.B., Froenicke, L., Huang, G., Farmer, A.D., Cannon, E.K.S., Liu, X., Gao, D., Clevenge, J., Dash, S., et al. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat. Genet. 48:438–446.

Berrabah, F., Ratet, P., and Gourion, B. (2015). Multiple steps control immunity during the intracellular accommodation of rhizobia. J. Exp. Bot. 66:1977–1985.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics 22:1269–1271.

Cao, Y., Halane, M.K., Gassmann, W., and Stacey, G. (2017). The role of plant innate immunity in the legume-rhizobium symbiosis. Annu. Rev. Plant Biol. 68:535–561.

Charpentier, M., Sun, J., Vaz Martins, T., Radhakrishnan, G.V., Findlay, K., Soumpourou, E., Thouin, J., Very, A.A., Sanders, D., Morris, R.J., et al. (2016). Nuclear-localized cyclic nucleotide-gated channels mediate symbiotic calcium oscillations. Science 352:1102–1105.

Chen, H., Zeng, Y., Yang, Y., Huang, L., Tang, B., Zhang, H., Hao, F., Liu, W., Li, Y., Liu, Y., et al. (2020). Allele-aware chromosome-level
Astragalus sinicus genome sheds light on legume-rhizobial interactions

Chen, X., Lu, Q., Liu, H., Zhang, J., Hong, Y., Lan, H., Li, H., Wang, J., Liu, H., Li, S., et al. (2019). Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and oil improvement. Mol. Plant 12:920–934.

Ding, J., Jiang, X., Ma, M., Zhou, B., Guan, D., Zhao, B., Zhou, J., Cao, F., Li, L., and Li, J. (2016). Effect of 35 years inorganic fertilizer and manure amendment on structure of bacterial and archival communities in black soil of northeast China. Appl. Soil Ecol. 105:187–195.

Dolgikh, A.V., Kirienko, A.N., Tikhonovich, I.A., Foo, E., and Dolgikh, E.A. (2019). The DELLAs proteins influence the expression of cytokinin biosynthesis and response genes during nodulation. Front. Plant Sci. 10:432.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:157.

Faucher, F., and Kondorosi, E. (2000). Cell cycle regulation in the course of nodule organogenesis in Medicago. Plant Mol. Biol. 43:773–786.

Gao, S.J., Zhou, G.P., and Cao, W.D. (2020). Effects of milk vetch (*Astragalus sinicus*) as winter green manure on rice yield and rate of fertilizer application in rice paddies in south China. J. Plant Nutr. Fertil. 26:1–12.

Gourion, B., Berrabah, F., Ratet, P., and Stacey, G. (2015). Rhizobium-legume symbioses: the crucial role of plant immunity. Trends Plant Sci. 20:186–194.

Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M.B., Billault-Penneteau, B., Laureenstegues, D., Keller, J., Imanishi, L., et al. (2018). Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. Science 361:1–11.

Guo, J.H., Liu, X.J., Zhang, Y., Shen, J.L., Han, W.X., Zhang, W.F., Christie, P., Goulding, K.W., Vitousek, P.M., and Zhang, F.S. (2010). Significant acidification in major Chinese croplands. Science 327:1008–1010.

Huang, L., Feng, G., Yan, H., Zhang, Z., Bushman, B.S., Wang, J., Bombarely, A., Li, M., Yang, Z., Nie, G., et al. (2020). Genome assembly provides insights into the genome evolution and flowering regulation of orchardgrass. Plant Biotechnol. J. 18:373–388.

Jia, Y., Wickett, N.J., Ayyampalayam, S., Chandlerbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Solitis, P.S., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature 473:97–100.

Kim, Y., Liu, H., Luo, D., Yu, N., Dong, W., Wang, C., Zhang, X., Dai, H., Yang, J., and Wang, E. (2016). DELLA proteins are common components of symbiotic rhizobial and mycorrhizal signalling pathways. Nat. Commun. 7:1243–12514.

Kamal, N., Mun, T., Reid, D., Lin, J.S., Akyl, T.Y., Sandal, N., Asp, T., Hirakawa, H., Stougaard, J., Mayer, K.F.X., et al. (2020). Insights into the evolution of symbiosis gene copy number and distribution from a chromosome-scale Lotus japonicus Gifu genome sequence. DNA Res. 27:1–10.

Kang, Y.-J., Kim, S.K., Kim, M.Y., Lestari, P., Kim, K.H., Ha, B.K., Jun, T.H., Hwang, W.J., Lee, T., Lee, J., et al. (2014). Genome sequence of mungbean and insights into evolution within Vigna species. Nat. Commun. 5:5443.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12:357–360.

Kohlen, W., Ng, J.L.P., Deinum, E.E., and Mathiessen, U. (2018). Auxin transport, metabolism, and signalling during nodule initiation: indeterminate and determinate nodules. J. Exp. Bot. 69:229–244.

Kreplak, J., Madoui, M.A., Capal, P., Novak, P., Labadie, K., Aubert, G., Paly, B.E., Gali, K.K., Syme, R.A., Main, D., et al. (2019). A reference genome for pea provides insight into legume genome evolution. Nat. Genet. 51:1411–1422.

Levy, J., Bres, C., Geurts, R., Chalhoub, B., Kulikova, O., Duc, G., Journet, E.P., Ané, J.M., Lauber, E., Bisseling, T., et al. (2004). A putative Ca2+ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. Science 303:1361–1364.

Li, H., Jiang, F., Wu, P., Wang, K., and Cao, Y. (2020). A high-quality genome sequence of model legume *Lotus japonicus* (MG-20) provides insights into the evolution of root nodule symbiosis. Genes (Basel) 11:1–15.

Li, Q.G., Zhang, L., Li, C., Dunwell, J.M., and Zhang, Y.M. (2013). Comparative genomics suggests that an ancestral polyploidy event leads to enhanced root nodule symbiosis in the papilionoidaeae. Mol. Biol. Evol. 30:2602–2611.
Astragalus sinicus genome sheds light on legume-rhizobial interactions

Li, Y., Zhou, L., Li, Y., Chen, D., Tan, X., Lei, L., and Zhou, J. (2008). A nodule-specific plant cysteine proteinase, AsNODF32, is involved in nodule senescence and nitrogen fixation activity of the green manure legume Astragalus sinicus. New Phytol. 180:185–192.

Li, Y.H., Zhou, G., Ma, J., Li, Y., Jin, L.G., Zhang, Z., Guo, Y., Zhang, J., Sui, Y., Zheng, L., et al. (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat. Biotechnol. 32:1045–1052.

Liu, C.W., and Murray, J.D. (2016). The role of flavonoids in nodulation host-range specificity: an update. Plants 5:1–13.

Liu, J., Yang, S., Zheng, Q., and Zhu, H. (2014). Identification of a dominant gene in Medicago truncatula that restricts nodulation by Sinorhizobium melloti strain Rm14. BMC Plant Biol. 14:167.

Lonardi, S., Muñoz-Amatriain, M., Liang, Q., Shu, S., Wanamaker, S.I., Love, M.I., Huber, W., and Anders, S. (2019). The genome of cowpea (Vigna unguiculata [L.] Walp.). Plant J. 98:767–782.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550–621.

Luck, J.E., Lawrence, G.J., Dodds, P.N., Shepherd, K.W., and Ellis, J.G. (2000). Regions outside of the leucine-rich repeats of flax rust host-range specificity: an update. Plants Agron. 12:265–275.

Li, Y., Zhou, L., Li, Y., Chen, D., Tan, X., Lei, L., and Zhou, J. (2015). Plant cysteine-rich peptides that inhibit pathogen growth and control rhizobial differentiation in legume nodules. Curr. Opin. Plant Biol. 26:57–63.

McGuiness, P.N., Reid, J.B., and Foo, E. (2019). The role of gibberellins and brassinosteroids in nodulation and arbuscular mycorrhizal associations. Front. Plant Sci. 10:289.

Montiel, J., Downie, J.A., and Kondorosi, É. (2015). Plant cysteine-rich peptides in legume nodules. Curr. Opin. Plant Biol. 26:1–13.

Mochida, K., Sakurai, T., Seki, H., Yoshida, T., Takahagi, K., Sawai, S., Uchiyama, H., Muranaka, T., and Saito, K. (2017). Draft genome assembly and annotation of Glycyrhiza uralensis, a medicinal legume. Plant J. 89:181–194.

Munch, D., Gupta, V., Bachmann, A., Busch, W., Kelly, S., Mun, T., and Andersen, S.U. (2018). The Brassicaceae family displays divergent, shoot-skewed NLR resistance gene expression. Plant Physiol. 176:1598–1609.

Ng, J.L., Hassan, S., Truong, T.T., Hocart, C.H., Laffont, C., Frugier, F., and Mathiesius, U. (2015). Flavonoids and auxin transport inhibitors rescue symbiotic nodulation in the Medicago truncatula cytokinin perception mutant cre1. Plant Cell 27:2210–2226.

Pan, H., and Wang, D. (2017). Nodule cysteine-rich peptides maintain a working balance during nitrogen-fixing symbiosis. Nat. Plants 3:17048.

Pecrix, Y., Staton, S.E., Sallet, E., Lelandais-Brière, C., Moreau, S., Carrere, S., Blein, T., Jardinoud, M.F., Latrasse, D., Zouine, M., et al. (2018). Whole-genome landscape of Medicago truncatula symbiotic genes. Nat. Plants 4:1017–1025.

Roy, S., Liu, W., Nandety, R.S., Crook, A., Mysore, K.S., Pislariu, C.I., Frugoli, J., Dickstein, R., and Udvardy, M.K. (2020). Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. Plant Cell 32:45–41.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., et al. (2010). Genome sequence of the palaeopolyploid soybean. Nature 463:178–183.

Schmutz, J., McClean, P.E., Mamidi, S., Wu, G.A., Cannon, S.B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C., et al. (2014). A reference genome for common bean and genome-wide analysis of domestications. Nat. Genet. 46:707–713.

Shen, Y., Liu, J., Geng, H., Zhang, J., Liu, Y., Zhang, H., Xing, S., Du, J., Ma, S., and Tian, Z. (2018). De novo assembly of a Chinese soybean genome. Sci. China Life Sci. 61:871–884.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Subramanian, S., Stacey, G., and Yu, O. (2007). Distinct, crucial roles of flavonoids during legume nodulation. Trends Plant Sci. 12:282–285.

Tang, F., Yang, S., Liu, J., and Zhu, H. (2015). Rj4, a gene controlling nodulation specificity in soybeans, encodes a thaumatin-like protein but not the one previously reported. Plant Physiol. 1:1–17.

Tan, X., Meyers, B.C., Kozik, A., West, M.A., Morgante, M., St Clair, D.A., Bent, A.F., and Michelmore, R.W. (2006). Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis. BMC Plant Biol. 7:56.

Varshney, R.K., Chen, W., Li, Y., Bharti, A.K., Saxena, R.K., Schlueter, J.A., Donoghue, M.T., Azam, S., Fan, G., Whaley, A.M., et al. (2011). Draft genome sequence of pigeonpea ( Cajanus cajan), an orphan legume crop of resource-poor farmers. Nat. Biotechnol. 30:83–89.

Varshney, R.K., Song, C., Saxena, R.K., Azam, S., Yu, S., Sharpe, A.G., Cannon, S., Baek, J., Rosen, B.D., Tar’an, B., et al. (2013). Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. Nat. Biotechnol. 31:240–246.

Voisin, A.-S., Guéguen, J., Hugyche, G., Jeuffroy, M.-H., Magrini, M.-B., Meynard, J.-M., Mougel, C., Pellerin, S., and Polzer, E. (2013). Legumes for feed, food, biomaterials and bioenergy in Europe: a review. Agron. Sustain. Dev. 34 (2):361–380.

Wang, Q., Liu, J., and Zhu, H. (2018). Genetic and molecular mechanisms underlying symbiotic specificity in legume-rhizobium interactions. Front. Plant Sci. 9:313.

Wang, Q., Yang, S., Liu, J., Terecskei, K., Abraham, E., Gombar, A., Domonkos, A., Szucs, A., Kormocz, P., Wang, T., et al. (2017). Host-secreted antimicrobial peptide enforces symbiotic selectivity in Medicago truncatula. Proc. Natl. Acad. Sci. U S A. 114:6854–6859.

Wasson, A.P., Pellerone, F.I., and Mathiesius, U. (2006). Silencing the flavonoid pathway in Medicago truncatula inhibits root nodule formation and prevents auxin transport regulation by rhizobia. Plant Cell 18:1617–1629.

Wang, C., Yu, H., Luo, L., Duan, L., Cai, L., He, X., Wen, J., Mysore, K.S., Li, G., Xiao, A., et al. (2016). Nodules with activated defense 1 is required for maintenance of rhizobial endosymbiosis in Medicago truncatula. New Phytol. 212:176–191.

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics 8:77–80.

Wasson, S., Hao, B., Li, J., Gu, H., Peng, J., Xie, F., Zhao, X., Frech, C., Chen, N., Ma, B., et al. (2014). Whole-genome sequencing of Mesorhizobium huakuii 7653R provides molecular insights into host specificity and symbiosis island dynamics. BMC Genomics 15:440–517.

Xie, Z., Tu, S., Shah, F., Xu, C., Chen, J., Han, D., Liu, G., Li, H., Muhammad, I., and Cao, W. (2016). Substitution of fertilizer-N by green manure improves the sustainability of yield in double-rice cropping system in south China. Field Crops Res. 188:142–149.

Yang, K., Tian, Z., Chen, C., Luo, L., Zhao, B., Wang, Z., Yu, L., Li, Y., Sun, Y., Li, W., et al. (2015). Genome sequencing of adzuki bean (Vigna
Astragalus sinicus genome sheds light on legume-rhizobial interactions

Plant Communications

*angularis* provides insight into high starch and low fat accumulation and domestication. Proc. Natl. Acad. Sci. U S A. 112:13213–13218.

Yang, L., Zhou, X., Liao, Y., Lu, Y., Nie, J., and Cao, W. (2019). Co-incorporation of rice straw and green manure benefits rice yield and nutrient uptake. Crop Sci. 59 (2):749–759.

Yang, S., Tang, F., Gao, M., Krishnan, H.B., and Zhu, H. (2010). R gene-controlled host specificity in the legume-rhizobia symbiosis. Proc. Natl. Acad. Sci. U S A. 107:18735–18740.

Yang, S., Wang, Q., Fedorova, E., Liu, J., Qin, Q., Zheng, Q., Price, P.A., Pan, H., Wang, D., Griffiths, J.S., et al. (2017). Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula*. Proc. Natl. Acad. Sci. U S A. 114:6848–6853.

Yang, W.C., de Blank, C., Meskiene, I., Hirt, H., Bakker, J., van Kammen, A., Franssen, H., and Bisseling, T. (1994). Rhizobium nod factors reactivate the cell cycle during infection and nodule primordium formation, but the cycle is only completed in primordium formation. Plant Cell 6:1415–1426.

Yang, Z. (2007). Paml 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Young, N.D., Debelle, F., Oldroyd, G.E., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F., Gouzy, J., Schoof, H., et al. (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature 480:520–524.

Yu, X.J., Zheng, H.K., Wang, J., Wang, W., and Su, B. (2006). Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup. Genomics 88:745–751.

Yu, H., Bao, H., Zhang, Z., and Cao, Y. (2019). Immune signaling pathway during terminal bacteroid differentiation in nodules. Trends Plant Sci. 24:299–302.

Zhang, B., Wang, M., Sun, Y., Zhao, P., Liu, C., Qing, K., Hu, X., Zhong, Z., Cheng, J., Wang, H., et al. (2021). *Glycine max* NNL1 restricts symbiotic compatibility with widely distributed bradyrhizobia via root hair infection. Nat. Plants 7 (1):73–86.

Zhang, H., Gao, S., Lercher, M.J., Hu, S., and Chen, W.H. (2012). EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. Nucleic Acids Res. 40:W569–W572.

Zhang, C., He, J., Dai, et al. (2020). Discriminating symbiosis and immunity signals by receptor competition in rice. Proc. Natl. Acad. Sci. U S A. 118:1–8.

Zhang, J., Subramanian, S., Stacey, G., and Yu, O. (2009). Flavones and flavonols play distinct critical roles during nodulation of *Medicago truncatula* by *Sinorhizobium meliloti*. Plant J. 57:171–183.

Zhang, X., Zhang, R., Gao, J., Wang, X., Fan, F., Ma, X., Yin, H., Zhang, C., Feng, K., and Deng, Y. (2017). Thirty-one years of rice-rice-green manure rotations shape the rhizosphere microbial community and enrich beneficial bacteria. Soil Biol. Biochem. 104:208–217.

Zheng, F., Wu, H., Zhang, R., Li, S., He, W., Wong, F.L., Li, G., Zhao, S., and Lam, H.M. (2016). Molecular phylogeny and dynamic evolution of disease resistance genes in the legume family. BMC Genomics 17:402–413.