MPMI Vol. 35, No. 10, 2022, pp. 893–905, https://doi.org/10.1094/MPMI-03-22-0054-R

Transcriptomic Identification of a Unique Set of Nodule-Specific Cysteine-Rich Peptides Expressed in the Nitrogen-Fixing Root Nodule of *Astragalus sinicus*

Feng Wei,[1] Yuan Liu,[1] Donglai Zhou,[1] Wenlong Zhao,[1] Zhennan Chen,[1] Dason Chen,[1] Youguo Li,[1,†] and Xue-Xian Zhang[2]

[1] State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, China
[2] School of Natural Sciences, Massey University at Albany, Auckland 0745, New Zealand

Accepted for publication 22 June 2022.

Legumes in the inverted repeat-lacking clade (IRLC) each produce a unique set of nodule-specific cysteine-rich (NCR) peptides, which act in concert to determine the terminal differentiation of nitrogen-fixing bacteroid. IRLC legumes differ greatly in their numbers of NCR and sequence diversity. This raises the significant question how bacteroid differentiation is collectively controlled by the specific NCR repertoire of an IRLC legume. *Astragalus sinicus* is an IRLC legume that forms indeterminate nodules with its microsymbiont *Mesorhizobium huakuii* 7653R. Here, we performed transcriptome analysis of root and nodule samples at 3, 7, 14, 28 days postinoculation with *M. huakuii* 7653R and its isogenic Δ*bacA* mutant. BacA is a broad-specificity peptide transporter required for the host-derived NCRs to target rhizobial cells. A total of 167 NCRs were identified in the RNA transcripts. Comparative sequence and electrochemical analysis revealed that *A. sinicus* NCRs (AsNCRs) are dominated by a unique cationic group (termed subgroup C), whose mature portion is relatively long (>60 amino acids) and phylogenetically distinct and possessing six highly conserved cysteine residues. Subsequent functional characterization showed that a 7653R variant harboring AsNCR083 (a representative of subgroup C AsNCR) displayed significant growth inhibition in laboratory media and formed ineffective white nodules on *A. sinicus* with irregular symbiosomes. Finally, bacterial two-hybrid analysis led to the identification of GroEL1 and GroEL3 as the molecular targets of AsNCR067 and AsNCR076. Together, our data contribute to a systematic understanding of the NCR repertoire associated with the *A. sinicus* and *M. huakuii* symbiosis.

†Corresponding author: Y. Li; youguoli@mail.hzau.edu.cn

Funding: This research was financially supported by the funds from the National Key Research and Development Program of China (grant number 2019YFA0904704) and the National Natural Science Foundation of China (grant number 31772401, 31970267). Research in X. X. Zhang's lab was supported by the New Zealand MBIE Catalyst Fund (project number 92846082).

***e*-Xtra:** Supplementary material is available online.

The author(s) declare no conflict of interest.

Keywords: *Astragalus sinicus*, bacteroid differentiation, indeterminate nodule, IRLC legumes, NCR peptides, nitrogen fixation, *Rhizobium*

Nitrogen is often a limiting factor for plant growth, but only microorganisms (including rhizobia) are capable of transforming atmospheric nitrogen into bioavailable ammonia. Unlike free-living nitrogen-fixing bacteria, rhizobia specifically infect leguminous plants, leading to the development of a specialized organ, i.e., root nodule (Ledermann et al. 2021). Within the nodule, rhizobial cells differentiate into bacteroids for active nitrogen fixation. There are two major types of root nodules differing significantly in the extent of bacteroid differentiation (Downie and Kondorosi 2021). Determinate nodules grow by expansion, and their nitrogen-fixing bacteroid cells remain viable. In contrast, indeterminate nodules possess a persistent meristem at their apex, with a characteristic elongated oval shape. Bacteroids within an indeterminate nodule have variable shape and permanently lose the ability to divide. The terminally differentiated bacteroid cells have no reproductive future but conduct nitrogen fixation, which benefits the whole rhizobial population in planta (Alunni and Gourion 2016; Haag and Mergaert 2019; Zhang and Rainey 2010). Thus, they are considered a typical example of altruistic behavior in microorganisms (Masson-Boivin and Sachs 2018; West et al. 2006). The underlying molecular mechanisms have been subjected to continuous investigation.

The legume-rhizobial symbiosis is initiated by perception of plant-derived flavonoids and subsequent production of nodulation factors by rhizobial cells residing in the rhizosphere. However, the later stage of the symbiosis (or the fate of rhizobia once entering plant cells) are predominantly controlled by the host plant (Pan and Wang 2017). Legumes in the inverted repeat-lacking clade (IRLC), like genera *Medicago* and *Pisum*, form indeterminate nodules. Recent progress shows that IRLC legumes mediate bacteroid differentiation through the production of a unique set of antimicrobial peptides, collectively known as nodule-specific cysteine-rich (NCR) peptides (Alunni et al. 2007; Mergaert et al. 2003; Van de Velde et al. 2010). NCR peptides have a highly conserved N-terminal signal domain, which is cleaved in the endoplasmic reticulum before reaching the symbiosome (Alunni and Gourion 2016; Roy et al. 2020). After cleavage of the signal segments, mature NCRs are 35 to 55 residues in length, targeting many bacteroid differentiation processes, such as cell division and gene transcription (Farkas et al. 2014).

Rhizobial genes are involved in the mediation and antagonization of the antimicrobial activities provoked by the plant-derived NCRs (Nicoud et al. 2021). These include *bacA* (or *bclA*), a gene encoding a broad-specificity peptide transporter system belonging to the ATP-binding cassette family (Barriere et al. 2017). Inactivation of *bacA* in *Sinorhizobium meliloti* results in an NCR-hypersensitive phenotype with *Medicago truncatula*, and the mutant dies rapidly after releasing into the plant symbiotic cell (Haag et al. 2011; Nicoud et al. 2021). BacA provides resistance by re-directing NCRs away from their target sites on bacterial cell surfaces and, thereby, reducing damages to membrane permeability. The imported peptides can potentially be destroyed by intracellular peptidases. However, some NCRs may target enzymes in the cytoplasm, and thus, it is also possible that BacA may act as an exporter to reduce their cytoplasmic toxicity (Barriere et al. 2017; Haag et al. 2011).

Regarding the specific NCR molecular targets, a previous pull-down experiment with tagged NCR247 peptide from *Medicago truncatula* identified at least nine rhizobial protein complexes, including the chaperonin GroEL, ribosomal proteins, FtsZ, pyruvate dehydrogenase, transaldolase, RNA polymerase, elongation factors, a Maf-like protein, and nitrogenase (Farkas et al. 2014). GroEL is of particular interest, as it is functionally required by hundreds of proteins for their proper folding (Kumar et al. 2015; Roy et al. 2020). Among the five *groEL* homologues encoded in the *S. meliloti* genome, one gene was shown to be essential for symbiosis, and the corresponding *groEL1* mutant formed an ineffective nodule in *Medicago sativa* (Bittner et al. 2007). This finding led to a hypothesis that GroEL may act as a chaperone protein involved in the correct folding of NCRs and its interaction with NCR247 potentially affects the GroEL-dependent functions.

An intriguing feature of NCR peptides is their abundance and diversity among IRLC legumes. While over 600 NCRs were identified by sequencing of RNAs in *Medicago truncatula* root nodules, only seven were found in *Glycyrrhiza uralensis* (Chinese licorice), 63 for *Cicer arietinum* (chickpea), 353 for *Pisum sativum* (pea), and 469 for *Medicago sativa* (alfalfa) (Downie and Kondorosi 2021). NCRs can be roughly categorized into two subgroups with either four or six conserved cysteine residues. The number of cysteine residues is a functionally relevant feature, as they are responsible for the formation of two or three potential disulphide bridges (Roy et al. 2020). Moreover, IRLC legumes differ in the composition of anionic and cationic NCRs. Physiochemical properties are crucial for NCRs to mediate their antimicrobial effects. It has been proposed that anionic NCRs may act to antagonize the antimicrobial activities of cationic NCRs (Montiel et al. 2017; Roy et al. 2020). Therefore, the unique set of NCRs produced by an IRLC legume in a nodule tissue–specific manner holds the key to understanding the molecular basis of nitrogen fixation specificity (Kang et al. 2020; Montiel et al. 2017; Pan and Wang 2017; Wang et al. 2017; Yang et al. 2017).

Our work has been focused on *Astragalus sinicus* (Chinese milkvetch), an IRLC legume forming indeterminate nodules with *Mesorhizobium huakuii* (Turner et al. 2002; Zhang et al. 2000). Several NCRs have been identified in our previous studies (Chou et al. 2006; Peng et al. 2014), but a systematic understanding is lacking.

Here, we report a plant transcriptome experiment designed to identify NCRs that are specifically expressed in the root nodule of *A. sinicus*. RNA-seq was performed with six plant tissues whereby RNA samples were prepared from the root and the nodule at 3, 7, 14, and 28 days postinoculation (dpi) with wild-type (WT) 7653R and its derived Δ*bacA* mutant. The plant transcriptome data led to the identification of 167 AsNCRs in total. Next, representative AsNCRs were selected for genetic characterization to demonstrate their anti-rhizobial activities in vitro, with purified peptides, and in vivo, by expressing AsNCRs within rhizobial cells. A bacterial two-hybrid assay was employed to identify AsNCR target proteins in *M. huakuii* 7653R. Our results showed that AsNCR067 and AsNCR076 are capable of binding with the GroEL chaperones (GroEL1 and GroEL3). Data enables us to discuss how the AsNCR repertoire (particularly those in subgroup C) contributes to bacteroid differentiation and their potential role in determining the nitrogen fixation specificity in the *A. sinicus–M. huakuii* symbiotic system.

RESULTS

Transcriptomic identification of AsNCR genes encoded in the genome of *A. sinicus*.

RNA-seq was performed with a total of six treatments (Supplementary Table S1), designated according to the tissue type followed by dpi (3, 7, 14, or 28) with WT 7653R or its derived Δ*bacA* mutant. For example, treatment R_7WT refers to the root samples at 7 dpi with WT strain, whereas N_14Δ*bacA* is the nodule sample at 14 dpi with the Δ*bacA* mutant. High-quality clean reads were first mapped to the *M. huakuii* 7653R genome to eliminate the rhizobial sequences and were then subjected to *de novo* assembly of the *A. sinicus* transcriptome (Supplementary Table S2). Genes with $|\log_2 (\text{fold change [FC]})| \geq 1$ and false discovery rate (FDR) (corrected *P* value) <0.05 were designated differentially expressed genes (DEGs). Cluster analysis showed that the expression patterns were first separated by rhizobial symbiosis (roots versus nodules) and then the microsymbiont (WT versus Δ*bacA* mutant) (Fig. 1). Relatively small differences were noted between different timepoints (i.e., roots between 3 and 7 dpi, nodules between 14 and 28 dpi).

The gene ontology (GO) enrichment analysis identified 20 GO terms for a total of 867 DEGs between root and nodule samples with WT 7653R (Supplementary Table S3). These include genes with predicted functions in nitrogen fixation, nitrogen metabolisms, and cell-shape determination. Intriguingly, copper-binding proteins are predominantly found in 60 of the 867 DEGs, suggesting a previously unrecognized role of copper in the maintenance of a balanced cytosolic redox status in the root nodule. The Δ*bacA* mutant formed small and white nodules defective in nitrogen fixation (Nod^+, Fix^-) (Tan et al. 2009). A total of 4,109 DEGs were identified in 14-dpi nodules formed by WT and the Δ*bacA* mutant (Supplementary Table S4). More than half of the enriched genes (67.9%) are associated with amino acids and carboxylic acid metabolism, metal cluster binding, and nitrogen fixation (Supplementary Table S3).

The specific up- and downregulated genes associated with *bacA* inactivation are listed in Supplementary Table S5 (N_14Δ*bacA* vs. N_14WT). Genes whose expression was enhanced in Δ*bacA* are mostly intracellular enzymes (36%) such as gibberellin 2-beta-dioxygenase 2, cysteine proteinase, and purple acid phosphatase. It is interesting to note that genes encoding the KDEL-tailed cysteine endopeptidases are involved in the regulation of programmed cell death (Howing et al. 2018). Thus, expression of the cysteine proteinase at elevated levels (FC = 16.9) may explain the observed early senescence of the Δ*bacA* mutant nodule. Many downregulated genes have known functions in nodule development, including 10 leghemoglobin genes responsible for delivering oxygen in root nodules (Ott et al. 2005) and key nitrogen metabolic enzymes, such as glutamine synthetase (Silva and Carvalho 2013) and asparagine synthetase (Garcia-Calderon et al. 2017). However, the most prominent group of downregulated genes are those encoding for nodule-specific proteins or late nodulin proteins, 36 in total, and many of which show sequence similarity with the functionally characterized NCR peptides (Supplementary Table S5).

***In silico* analysis of NCR genes in the genome of *A. sinicus*.**

Next, we performed two rounds of tBLASTn searches for putative NCRs in the assembled *A. sinicus* transcriptomes. This analysis led to identification of 167 NCR-encoding sequences, 11 of which were previously characterized as cysteine clustering proteins, using the method of suppressive substrative hybridization in combination with reverse transcription (RT)-PCR (Chou et al. 2006). These NCRs are designated AsNCR001 to AsNCR69, with the prefix As for *A. sinucus* (of note, AsNCR052 and AsNCR068 were late eliminated due to duplication). They all possess an N-terminal signal peptide as predicted by the SignalP 4.0 server, except for nine AsNCRs (Supplementary Dataset S1). A signal peptide sequence was missing for AsNCRs numbered 004, 018, 021, 047, 049, 084, 131, 132, 169.

The mature portion of AsNCRs showed greater variation in terms of both length and sequence similarity. Sequence alignment revealed four or six highly conserved cysteine residues (Supplementary Fig. S1). The current available data led us to place AsNCRs into four different categories, i.e., subgroups A, B, C, and others (Fig. 2A). Subgroups A and B largely correspond to the previously identified MtNCRs in *Medicago truncatula*, whose mature portion contains the characteristic four versus six conserved cysteine residues, respectively. Like subgroup B, subgroup C AsNCRs possess six conserved cysteine residues, but their sequence structures are more like that of the antimicrobial plant defensins (Fig. 2D). A detailed comparison between AsNCR subgroup C peptides and representative plant defensins from *Medicago truncatula* and *Glycine max* is shown in Supplementary Figure S2. Phylogenetic analysis of the mature portions clearly showed that AsNCR in subgroup C formed a distinct cluster (Fig. 2B). However, there are a few exceptions. For example, AsNCR135 was placed into the subgroup C cluster, but it clearly belongs to subgroup A, as it has only four conserved cysteine residues (Supplementary Fig. S1). When compared with subgroups A and B, peptides of subgroup C are generally longer and most of them (66.1%) possess an additional sequence (>10 amino acids) in the linker region between signal peptide and the first cysteine (Fig. 2A and E). Additionally, AsNCRs in subgroup C have a relatively shorter signal peptide (Fig. 2A).

Finally, we determined the phylogenetic relationship between AsNCRs and NCRs from each of nine other members of ILRC legumes within the subfamily *Faboideae* (or *Papilionaceae*) (Supplementary Fig. S3). The data revealed a general separation of NCRs by plant species. More importantly, however, there are overlaps at multiple phylogenetic levels for NCRs from the two plant species in comparison. The data thus suggest that NCRs within a plant species may have distinct evolutionary histories. It is interesting to note that the AsNCR clusters in subgroup C (marked in red in the outer ring) have close homologues in other plant species (Supplementary Fig. S3).

AsNCRs in subgroup C are mostly cationic peptides.

To understand the electrochemical properties of AsNCRs, theoretical values of net charge and isoelectric point (pI) were

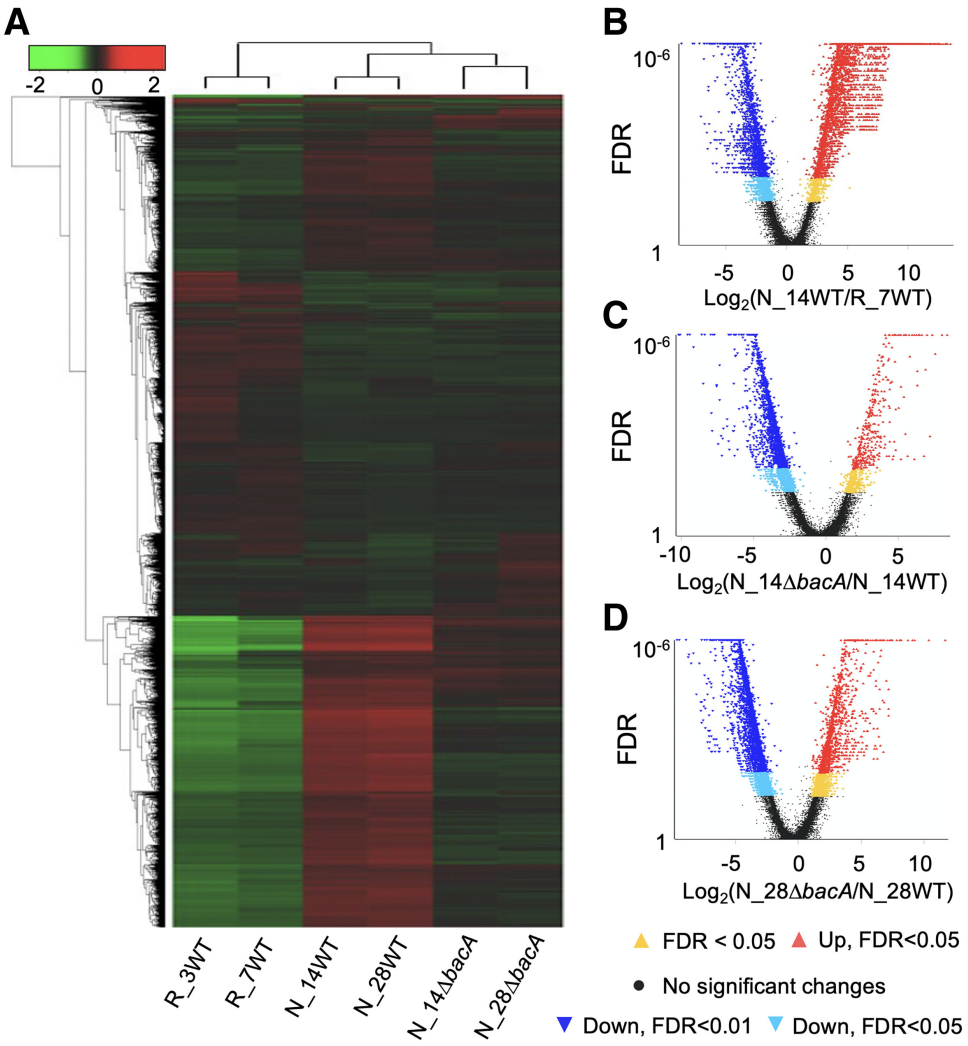

Fig. 1. Comparative transcriptome analysis of *Astragalus sinicus* genes associated with rhizobial symbiosis. **A,** Cluster analysis of differentially expressed genes between two different plant tissues (R = root; N = nodule), on four different days postinoculation (dpi) (3, 7, 14, and 28) with either wild-type (WT) *Mesorhizobium huakuii* 7653R or its derived mutant (Δ*bacA*). **B,** Volcano plot displaying the differentially expressed genes in nodule versus root at 14 dpi. **C,** Volcano plot for nodules at 14 dpi with WT versus the Δ*bacA* mutant. **D,** Volcano plot for nodules at 28 dpi with WT versus the Δ*bacA* mutant. FDR = false discovery rate or *P* value.

calculated on the basis of their deduced amino acid sequences (Fig. 2C and F). We unexpectedly found that AsNCRs in subgroup C are mostly cationic (93.2%, 55 of 59). In contrast, anionic and cationic NCRs are equally distributed in AsNCRs subgroups A and B (Fig. 2C). Most subgroup C AsNCRs have relatively higher pI values (>8.0) (Fig. 2F) and higher net charges (Supplementary Dataset S1). Twenty AsNCRs were found to have a net charge ≥8.0, and all belong to subgroup C, except AsNCR102 from the "others" subgroup. Next, we compared the electrochemical properties of AsNCRs and NCRs from 10 other IRLC leguminous plants. Data indicates that AsNCRs generally have the highest proportion of cationic NCRs (Fig. 3A), NCRs with pI > 8.0 (Fig. 3B), and large NCRs of > 60 amino acid in length (Fig. 3C). However, these unique features are largely attributable to members in subgroup C. Together, the RNA-seq and the computational analysis data revealed the presence of a unique group of large cationic NCRs (subgroup C) in addition to the previously described NCRs in subgroups A and B. The finding led us to ask how they collectively determine the nodule development and the differentiation of the nitrogen-fixing bacteroids.

Re-analysis of the transcriptome data for AsNCR expression.

We first sought to determine the expression patterns of AsNCR, while noting that 36 NCRs were picked up in the

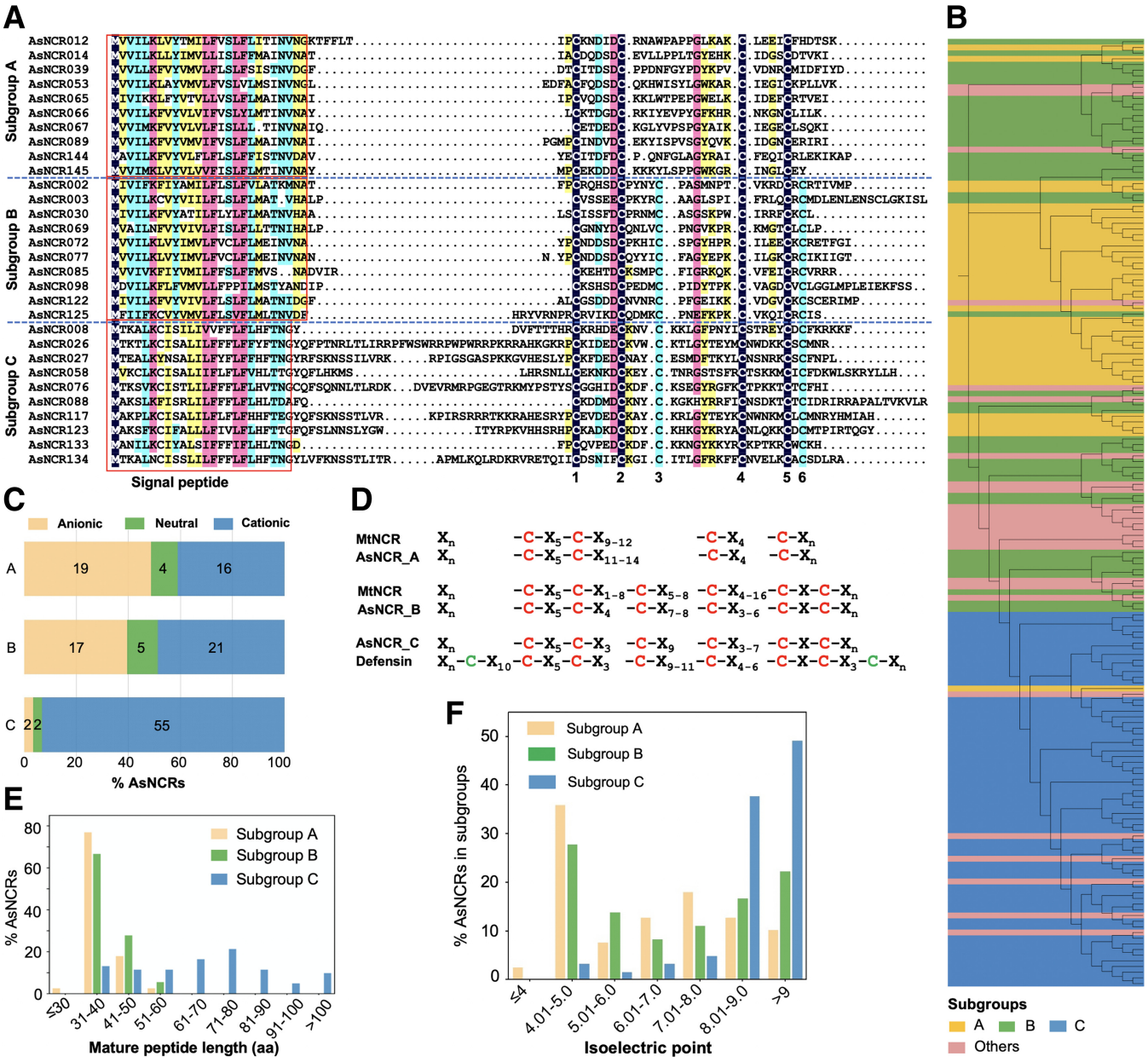

Fig. 2. *In silico* analysis of *Astragalus sinicus* nodule-specific cysteine-rich peptides (AsNCRs). **A,** Alignment of amino acid sequences of AsNCRs. The signal peptide regions are shown in red boxes, and the conserved cysteine residues are numbered 1 to 6. Sequence identity is marked by different colors as follows: dark blue (100%), purple (>80%), aqua (>60%), yellow (>40%). **B,** Phylogenetic tree of AsNCRs based on sequences of the mature peptides. **C,** Relative abundance of anionic, neutral, and cationic peptides within the three AsNCR subgroups. **D,** Backbone structures of the mature AsNCR subgroup A, B, and C peptides in comparison with the corresponding *Medicago truncatula* NCRs (MtNCRs) and plant defensin. **E,** Length distribution of the AsNCR subgroup A, B, and C peptides. **F,** Distribution of AsNCR peptides in subgroup A, B and C according to their isoelectric point (pI) values. The classification is primarily based on NCR sequence structure (particularly the conserved cysteine residues). Subgroups A and B have four and six highly conserved cysteines, respectively. Subgroup C NCRs possess the same six cysteines as subgroup B, but they contain extra sequences (and thus are longer) forming a separate phylogenetic cluster.

abovementioned DEG analysis (Supplementary Table S5). To this end, FPKM values were extracted from the transcriptome data (Supplementary Table S6). Little or no expression was detected in the root samples at 3 dpi, and some were expressed at moderate levels in roots at 7 dpi (Fig. 4). This finding suggests potential roles of these AsNCRs in root development. Overall, majority of the AsNCRs showed elevated levels of expression in the nodules formed by WT 7653R when compared with the roots (Fig. 4). Notably, a set of AsNCRs were expressed at highest levels in both young (14 dpi) and mature nodules (28 dpi) with a fragments per kilobase per million reads (FPKM) value >1,000. However, the highly expressed AsNCRs in the N_14WT samples almost solely belong to subgroups A and B (Fig. 4).

Inactivation of the rhizobial *bacA* gene resulted in downregulation of AsNCRs whose expression was induced in the nodule (Fig. 4). Notably, most subgroup A AsNCRs (92.3%) were downregulated (FC $\geq$ 2) in the Fix$^-$ Δ*bacA* mutant nodule. About 8.5% of the subgroup C AsNCRs were highly downregulated with FC >10, whereas relatively larger proportions of subgroup A (38.5%) and subgroup B (25.6%) AsNCRs were highly affected by *bacA* inactivation (Supplementary Table S7). Together, the RNA-seq data implicate different roles of these AsNCRs in the late stage of *A. sinicus*–mesorhizobial symbiosis.

Finally, we sought to confirm the nodule-specific expression of the subgroup C AsNCRs. To this end, *A. sinincus* seedlings were inoculated with WT 7653R and total RNAs were prepared at 14 dpi from four types of *A. sinincus* tissues, namely, root, root nodule, stem, and leaf. Results of RT-PCR showed that the 18S ribosomal RNA gene was universally expressed in the four tissues, whereas the leghemoglobin gene was expressed in nodule tissue only (Supplementary Fig. S4). Nodule-specific expression was confirmed for all 59 AsNCRs, albeit the expression levels were very low for AsNCR019, AsNCR087, AsNCR116, and AsNCR124.

Testing the effects of purified AsNCR peptides on rhizobial growth inhibition.

In an attempt to functionally characterize AsNCRs in vitro, we purified three glutathione-*S*-transferase (GST)-tagged AsNCRs (numbers 105, 100, and 076), representing subgroups A, B, and C, respectively. A GST-tagged and a custom-synthesized AsNCR067 (subgroup A) were additionally included in this work. Isopropyl β-D-thiogalactopyranoside (IPTG)-induced expression of the four AsNCRs in *Escherichia coli* was confirmed by the quantitative (q)RT-PCR method (Supplementary Fig. S5), and purity of the purified peptides was checked by sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis (Supplementary Fig. S6). We first measured the effects on cellular membrane permeability when the peptides were added at varying concentrations up to 20 μM (Fig. 5A). Significant concentration-dependent inhibition was only detected with AsNCR100 (Fig. 5A). Consistently,

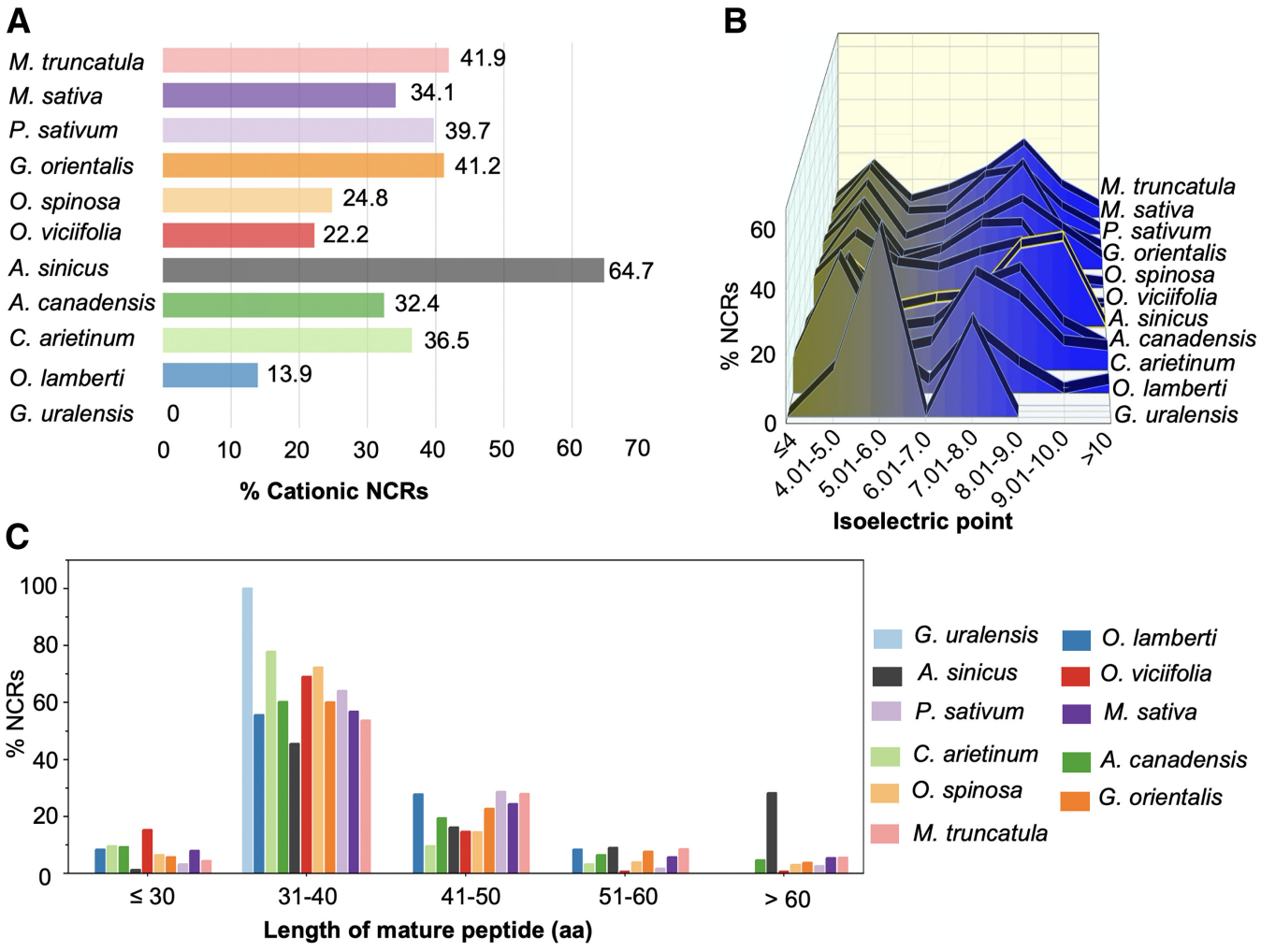

Fig. 3. A comparison of nodule-specific cysteine-rich peptides (NCRs) from eleven leguminous plants. **A,** Percentage of cationic NCRs encoded in the genome. **B,** Distribution of NCRs in terms of their isoelectric points. **C,** Variation in the distribution of length for NCRs encoded by different leguminous species.

cell-growth effects were observed in terms of colony-forming units (CFU) (Fig. 5B). The purified GST tag was used as control in these in-vitro assays. A minor but significant effect on cell membrane permeability was detected for AsNCR076 (Fig. 5A), but addition of purified AsNCR076 produced no significant differences in CFU (Fig. 5B).

Contrary to our expectation, AsNCR067 significantly enhanced rhizobial growth in a concentration-dependent manner (Fig. 5B), and growth kinetics were shown in Supplementary Figure S7 when the GST-tagged AsNCR067 was added at 20 μM. However, the GST-tagged AsNCR067 did not produce any significant effects on cell membrane permeability (Fig. 5A). The growth enhancement was further confirmed in a separate assay with the addition of 20 μM custom-synthesized AsNCR067 (Fig. 5C). Furthermore, we measured DNA contents of synchronized 7653R cells subjected to treatments by 20 μM AsNCR067 (Fig. 5D). Results showed that the AsNCR067-treated population was dominated by diploid cells as an indication of active cell division. In contrast, about two-third (65.2%) of the control cells treated with phosphate buffer solution (pH, 7.4) were mostly in haploid phase. Finally, we examined the cellular locations of an enhanced green fluorescent protein (eGFP)-AsNCR067 protein fusion expressed in *M. huakuii* 7653R bacteroid cells in the nodule. Unlike the eGFP marker itself, eGFP-AsNCR067 was not evenly distributed in the whole cell (Supplementary Fig. S8). Together, our data implicate a surprising role of AsNCR067 in enhancing rhizobial growth.

Genetic analysis of AsNCR expression in *M. huakuii*.

To gain insights into the biological functions of AsNCRs, we adopted a genetic approach by directly expressing the plant-derived peptides in *M. huakuii* 7653R. To this end, the mature AsNCR coding region was cloned into a plasmid vector pSRKGm (Khan et al. 2008), and its expression was placed under the control of an IPTG-induced P_{lac} promoter. A total of 15 AsNCRs were picked up for genetic characterization, according to their theoretical electric charges, representing the three AsNCR subgroups, four from subgroup A, four from B, seven from C (Supplementary Table S8). Effects on rhizobial growth in tryptone yeast (TY) agar plates are summarized in Figure 5E. The 16 resultant 7653R mutants, including one carrying the empty vector, grew similarly under the un-induced condition. Significant growth inhibition was observed for AsNCR051 and AsNCR083 when the IPTG inducer was added into the medium (Fig. 5E) and their expression was confirmed by qRT-PCR (Supplementary Fig. S9). AsNCR051 belongs to subgroup B with pI = −4.07, whereas AsNCR083 is a representative subgroup C cationic peptide (pI = +9.62).

Next, *AsNCR051* and *AsNCR083* genes were expressed in plasmid vector pMP2463 under the control of a weak constitutive P_{lac} promoter, and the phenotypic consequences were determined in the genetic background of WT *M. huakaii* 7653R during symbiosis with *A. sinicus*. Expression of AsNCR083 in the rhizobial symbiont disrupted the late nodule development and bacteroid differentiation (Fig. 6), resulting in white nodules incapable of fixing nitrogen (Nod^+, Fix^-). Transmission electron microscopy (TEM) examination showed that the ineffective nodules formed irregular and disordered symbiosomes without central vacuoles (Fig. 6F). The symbiosomes showed shrunken peri-bacteroid membranes, suggesting an early senescence of the symbiotic cells (Fig. 6G). However, the *AsNCR051* weak-expression strain formed normal nitrogen-fixing nodules like the WT 7653R (data not shown).

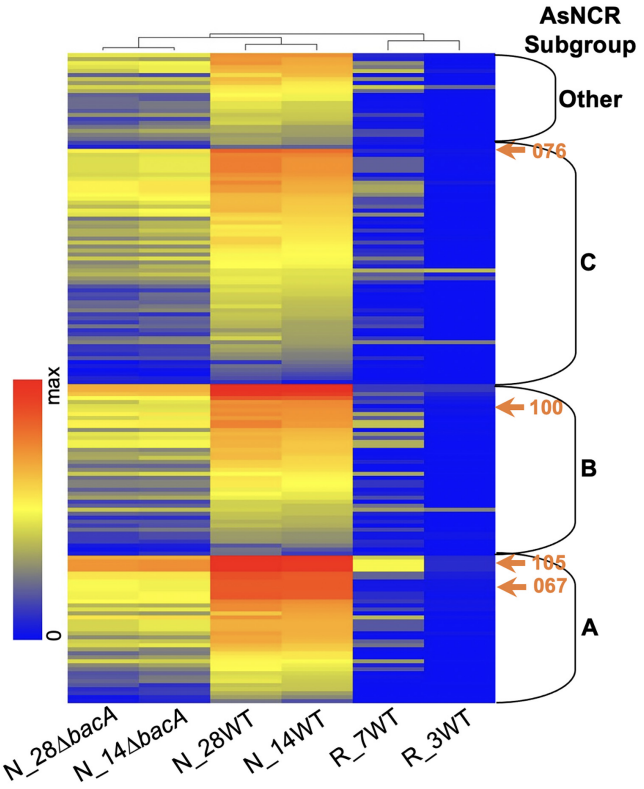

Fig. 4. Symbiosis-induced expression of *Astragalus sinicus* nodule-specific cysteine-rich peptides (AsNCRs). Expression patterns are compared between roots and root nodules formed by wild-type (WT) *Mesorhizobium huakuii* 7653R and its derived Δ*bacA* mutant. AsNCRs in subgroup A and B were generally expressed at higher levels than those in subgroup C (and others). The four AsNCRs selected for protein characterization are indicated by left arrows.

Determination of the AsNCR interactions with GroEL chaperones.

Finally, we sought to determine the molecular targets of AsNCRs in the genome of *M. huakuii* 7653R. To this end, we first performed a bacterial two-hybrid system (B2H) analysis whereby two AsNCRs (AsNCR067 and AsNCR076) were separately cloned in the bait plasmid pBT whereas 7653R genomic DNA was cloned into the target plasmid pTRG. The assay led to the identification of *groEL1* (gene ID MCHK_RS07240) and *groEL3* (gene ID MCHK_RS17605). Both *groEL1* and *groE3* were picked up in the B2H assay with AsNCR067 and AsNCR076 as the baits. The direct protein-protein interactions between the two AsNCRs (067 and 076) and the two GroEL homologues (GroEL1 and GroEL3) were further verified using B2H with clean pBT and pTRG constructs (Supplementary Fig. S10).

The chromosome of *M. huakuii* 7653R contains three GroEL homologues whose expressions differ when rhizobia grew in vitro and in planta (Peng et al. 2014). The *groEL2* gene was included in the above-described bacterial two-hybrid assays, but reliable protein-protein interactions were not detected with AsNCR067 and AsNCR076 (data not shown). Phylogenetic analysis showed that GroEL1 of *M. huakuii* is closely related to the GroEL1/2 cluster from *Sinorhizobium meliloti* 1021, whereas the closest homologues of *M. huakuii* GroEL2 and GroEL3 are *Mesorhizobium loti* BAV47041.1 and BAV46874.1, respectively (Supplementary Fig. S11). This suggests that the multiple GroEL genes of *A. sinicus* likely have different biological functions. In an attempt to construct knockout mutants for each of the three *groEL* genes, we have successfully deleted *groEL3* in WT 7653R. The obtained Δ*groEL3* mutant grew normally in laboratory medium (Supplementary Fig. S12), but it showed severe defects in symbiosis with *A. sinicus*, forming ineffective white nodules (Fig. 7). Finally, we compared the expression levels of

10 representative AsNCRs in nodules formed by WT 7653R and the Δ*groEL3* mutant. qRT-PCR results showed that AsNCR067 and AsNCR089 were expressed at significantly lower levels in Δ*groEL3* nodules (Supplementary Fig. S13). However, expression of AsNCR076 was reduced but the difference was not significant ($P < 0.05$).

DISCUSSION

The legume-rhizobial symbiosis is a paradigmatic example of mutualism between prokaryotes and their eukaryotic hosts (Poole et al. 2018). An important current research question is how a unique species-specific set of NCR peptides produced by an IRLC legume function in concert with each other to mediate efficient nitrogen fixation with specific microsymbionts. We here report the identification of 167 AsNCR peptides expressed in the root nodule of *A. sinicus*. AsNCRs are dominated by long cationic peptides. More specifically, AsNCRs can be placed into three major subgroups based on sequence similarity and electrochemical property. AsNCR subgroups A and B largely correspond to NCRs with four and six conserved cysteine residues that have been identified in the *Medicago truncatula* model system. However, AsNCRs in subgroup C are all cationic peptides with pI values larger than 8.0. They have six conserved cysteines residues and form a distinct phylogenetic cluster, with shorter signal peptides and longer mature peptides (>60 amino acids). Together, our data reveal a species-specific pattern of NCRs for *A. sinicus*. The functional significance of the dominant cationic long AsNCR peptides remains elusive.

The cationic long subgroup C AsNCRs are phylogenetically positioned between the typical symbiotic NCRs and plant defensins. NCRs are known to share a common ancestor with plant defensins, which represent a large diverse group of cationic cysteine-rich peptides produced by plants for defense against invading pathogens and parasites (Sathoff and Samac 2019). As one part of the plant innate immune system, defensins have a relatively simple role of suppressing pathogen invasion. However, the primary function of NCRs is to regulate the growth and activity of the rhizobial microsymbionts. The overall toxicity of various NCRs must thus be tightly controlled to either inhibit or permit rhizobial growth within the plant cell, depending on the physiological demand for nitrogen (Pan and Wang 2017). Loss of plant control over rhizobial growth in vivo would lead to the breakdown of the mutualistic relationship. Thus, the antimicrobial NCRs are collectively used by legumes as a tool rather than a weapon (Alunni and Gourion 2016). However, this does not exclude the possibility that some symbiotic NCRs have multiple functions and may participate in plant defense against bacterial infections (Maroti et al. 2015). The key point is the potential

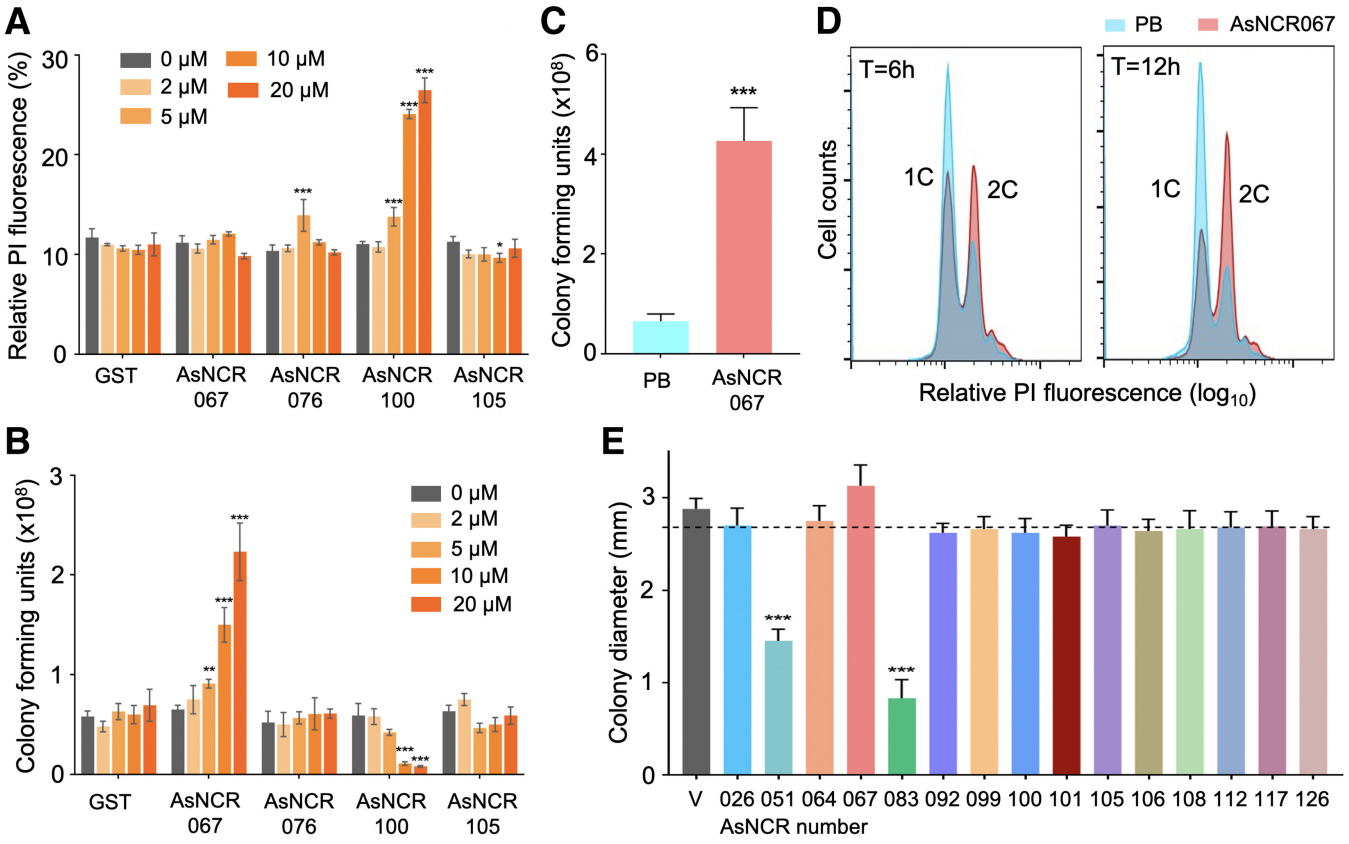

Fig. 5. Effects of *Astragalus sinicus* nodule-specific cysteine-rich peptides (AsNCRs) on rhizobial growth in vitro. **A,** Growth effect was determined in terms of cell membrane permeability and **B,** colony forming units. Glutathione-*S*-transferase (GST)-tagged AsNCRs were added at five different concentrations from 0, 2, 5, 10, and 20 μM. GST tag only was included as the negative control. **C,** A separate experiment showing the effects of AsNCR067 (20 μM) on colony forming units. The custom-synthesized AsNCR067 was dissolved in the phosphate buffered saline buffer. Data are means and standard deviation of at least three biological replicates. **D,** Fluorescence-activated cell sorting profiles of synchronized wild-type 7653R cells after 6- and 12-h treatments of custom-synthesized AsNCR067 (20 μM). Phosphate buffer (PB) was used as control. Cellular contents of DNA stained by propidium iodide (PI) are expressed at the *x* axis. Haploid and diploid cells are labelled as 1C and 2C, respectively. **E,** Effects of AsNCRs on rhizobial growth in agar plates. Colony diameters were measured in tryptone yeast plates containing gentamycin (20 μg/ml) at 28°C for 5 days. Expression of the AsNCR genes cloned in vector pSRKGm was induced by isopropyl β-D-thiogalactopyranoside (4 mM), and the data were compared with that of uninduced cells. The discontinuous line indicates the average level of un-induced cells, and no significant difference ($P < 0.05$) was detected among different strains, including a mutant containing the empty vector. Asterisks (***) denote significance at $P < 0.001$, revealed by the Student's *t* test. Data are means and standard deviations of at least 10 colonies.

antagonistic effects among various NCRs and their interactions with rhizobial factors (Pan and Wang 2017). To this end, our work revealed an interesting finding that the purified AsNCR067 (an anionic peptide) enhanced rhizobial growth, while a growth inhibition was expected. An antagonistic biological function has been hypothesized for certain anionic and neutral NCR peptides (Roy et al. 2020).

The evolutionary force driving bacteroid differentiation sounds very clear. It is to direct valuable resources to the nitrogen-fixing activity while minimizing the rhizobial metabolism. However, it would be technically challenging for plants to use antimicrobial NCRs to control rhizobial metabolism without killing the rhizobial cells. Interestingly, a recent comparative study showed that the number and composition of NCRs are

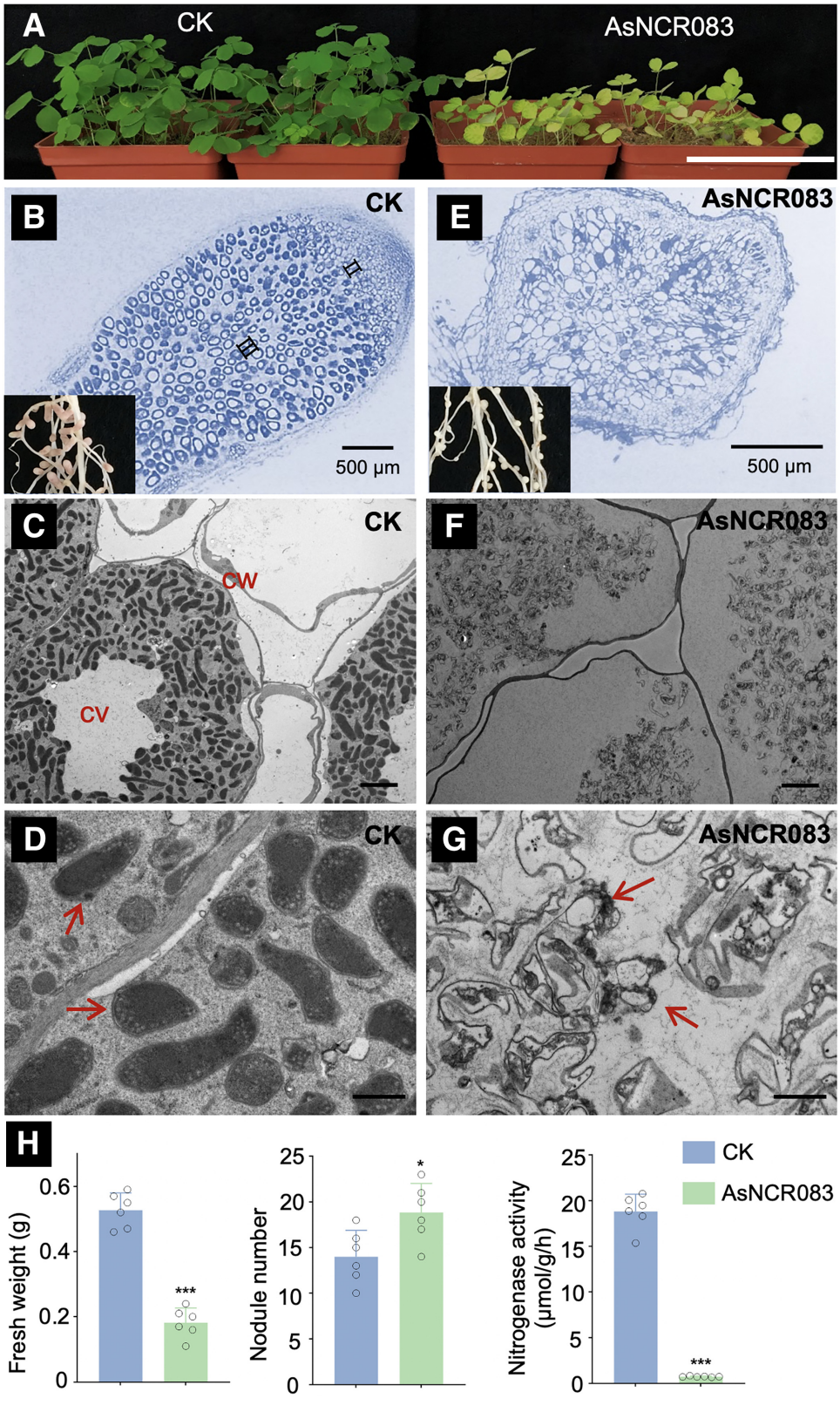

Fig. 6. Symbiotic consequences of AsNCR083 expression in the genetic background of wild-type *Mesorhizobium huakuii* 7653R. **A,** Images of *Astragalus sinicus* plants harvested at 28 days postinoculation with 7653R carrying AsNCR083 and the control strain carrying an empty vector (CK). **B to D,** Representative images of the paraffin-embedded sections of a root nodule (root nodules in inset) and that of the transmission electron microscopic (TEM) analysis are shown for the control strain and **E to G,** for AsNCR-overexpression strains. Representative bacteroid cells are indicated by arrows (cv = central vacuole; cw = cell wall). Infection zone (II) and nitrogen-fixing zone (III) are marked in the nodule formed by the control rhizobia (B). **H,** Symbiotic performance was assessed in terms of fresh shoot weight, nodule number per plant, and nitrogenase activity. Significant levels of Student's *t* test are indicated by one asterisk (*) for $P < 0.05$ and three (***) for $P < 0.001$.

correlated with the extent of bacteroid differentiation (Montiel et al. 2017). This suggests that a more complex set of NCRs is required for an IRLC species to achieve a higher degree of bacteroid differentiation.

The number of NCRs is moderate for *A. sinicus*, as compared with the lowest seven NCRs found in Chinese licorice and the highest 639 NCRs detected in *Medicago truncatula* nodules (Roux et al. 2014). However, the unique subgroup C AsNCRs are more closely related to the plant defensins, and homologues are predominantly found in *A. sinicus* only, thus representing one of the most significant species-specific features of AsNCRs. Interestingly, bacteroids in *A. sinicus* nodules appear to show a moderate degree of differentiation. They normally have a spindle-like shape (Fig. 6), similar to the elongated type described by Montiel et al. (2017). Furthermore, we have obtained empirical evidence showing that expression of a cationic subgroup C peptide (AsNCR083) in *M. huakuii* 7653R caused growth inhibition in vitro (Fig. 5E), and the mutant strain produced inefficient nodules (Nod^+, Fix^-) (Fig. 6). While the data demonstrate the anti-rhizobial activities of AsNCR083, the observed severe symbiotic phenotype is likely a consequence of the host plant losing control of the AsNCR toxicity. AsNCR083 was expressed at low levels (data not shown) and the early nodulation process was not affected. Therefore, the severe effects at the late stage of the symbiosis can be explained by the disruption of the normal NCR interaction network caused by weak expression of extra AsNCR083 peptide.

Finally, it is worth noting that GroEL chaperones were identified as the potential targets of AsNCRs in an initial pull-down assay, and the interactions were further confirmed between GroEL1 or GroEL3 and two AsNCRs (067 and 076). AsNCR067 is anionic (net charge −3.0, subgroup A), whereas AsNCR076 is a cationic peptide (net charge 7.0, subgroup C). This suggests that the GroEL interactions are not restricted to certain types of NCRs. The precise roles of GroELs in NCR-mediated nitrogen fixation specificity are yet unknown. GroEL is responsible for the correct folding of 10 to 15% of the cellular proteins in the well-studied *E. coli* model (Kumar et al. 2015). The plant-derived NCRs may need GroEL to assist their folding in the cytoplasm of rhizobial cells. Moreover, the NCR/GroEL binding may interfere with the normal function of GroEL, producing global effects on cellular metabolisms. Multiple *groEL* genes have been reported and functionally characterized in rhizobia, including three in *Rhizobium leguminosarum* (Rodriguez-Quinones et al. 2005) and five in *S. meliloti* (Bittner et al. 2007; Farkas et al. 2014). However, only one of the five *S. meliloti groEL* genes (*groEL1*) is required for successful nitrogen fixation (Bittner et al. 2007) and GroEL function is crucial during the entire symbiotic process (Farkas et al. 2014). Here, we have managed to delete one of the three *groEL* genes in *M. huakuii* 7653R. Results

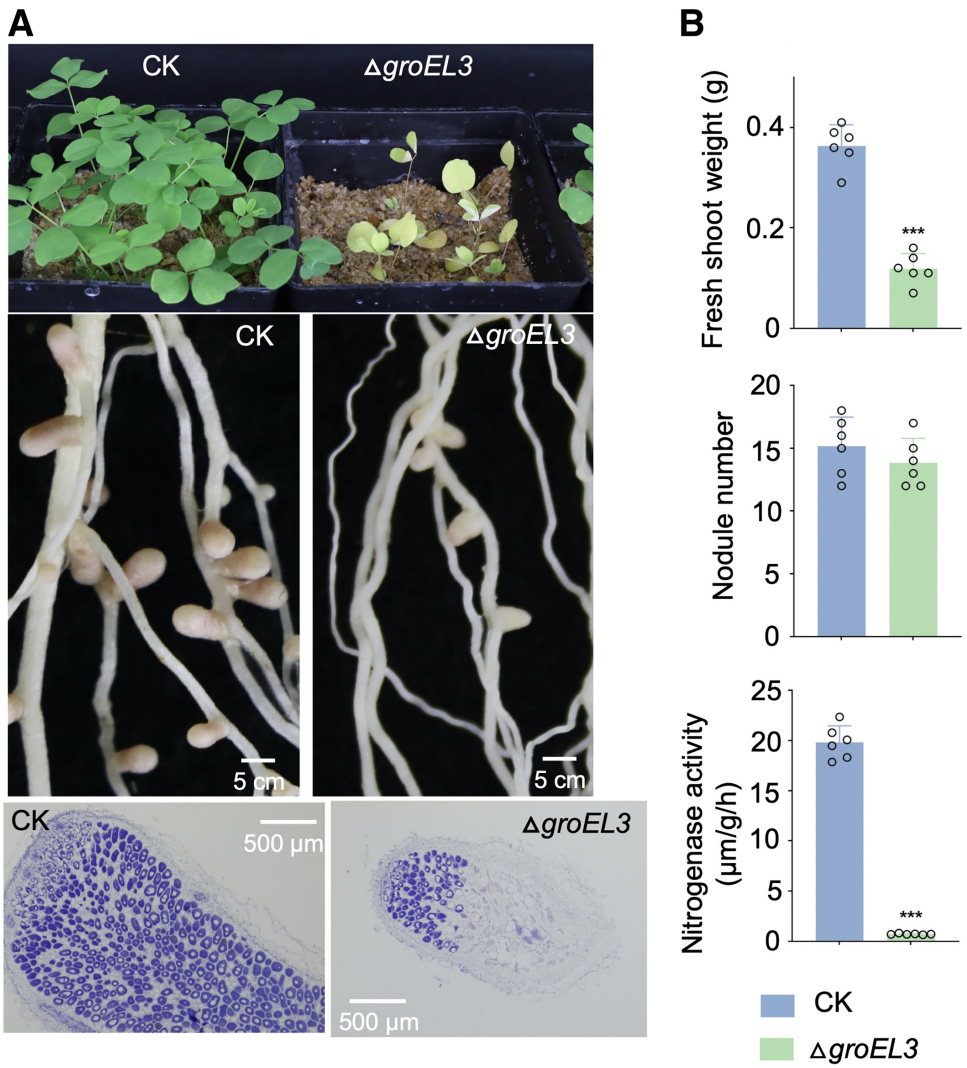

Fig. 7. Symbiotic phenotypes of the *groEL3* gene in *Mesorhizobium huakuii* 7653R. **A,** *Astragalus sinicus* plants and nodules showing the effects of inoculation with wild-type 7653R and its derived △*groEL3* mutant. Representative images of nodule sections stained with toluidine blue are provided at the bottom. **B,** Symbiotic performance was assessed in regard to fresh shoot weight, nodule number, and nitrogenase activity. Asterisks (***) denote significance at $P < 0.001$, revealed by Student's *t* test.

showed that the *groEL3* mutant was defective in nitrogen fixation. Further functional analysis of the three *groEL* genes will help understand how various AsNCRs contribute to the determination of nitrogen-fixing specificity in the *A. sinicus–M. huakuii* 7653R model.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions.

Rhizobial strains and plasmids used in this work are summarized in Table 1. *E. coli* DH5α or S17-1 (Simon et al. 1983) were used for general gene cloning and triparental conjugation into *M. huakuii* strains with the helper plasmid pRK2013. WT 7653R and its derived mutants were routinely grown in TY medium at 28°C and *E. coli* strains in Luria-Bertani medium at 37°C (Zhang et al. 2000). Antibiotics were added at the following concentrations except as otherwise specifically indicated (micrograms per milliliter): ampicillin, 100; streptomycin, 100; gentamicin, 20; and neomycin, 50. Growth dynamics were examined by growing rhizobial strains in 500-ml flasks containing 200 ml of TY broth on a rotary shaker (180 rpm). Optical density was measured at the wavelength of 600 nm.

Strain construction.

Standard protocols were used for plasmid DNA isolation, restriction digestion, ligation, and PCR. Sequences of the oligonucleotide primers are available in Supplementary Table S9. The *groEL3* deletion mutant was constructed with the integrative suicide vector pK19mob (Schäfer et al. 1994), following the same experimental procedure as previously described (Tan et al. 2009). For AsNCR expression in vitro, the coding region was amplified by PCR, and the PCR product was cloned into the broad–host range expression vector pSRKGm (Khan et al. 2008) at the *Nde*I and *Pst*I sites, following instructions of the In-Fusion Cloning kit (Takara Bio Inc.). The resulting recombinant plasmids were then introduced into *M. huakuii* 7653R, using the standard method of triparental mating. Growth inhibition was examined in TY agar plates supplemented with or without IPTG (4 mM). Colony sizes were measured after 5 days incubation at 28°C. To visualize the cellular location of AsNCR067, primers were designed to amplify the DNA region encoding the mature peptide, and the PCR product was cloned into pMP2463 (Stuurman et al. 2000b) at the Acc65I site, using the Trelief SoSoo cloning kit (TsingKe Biotech Ltd). The resultant eGFP-AsNCR067 fusion plasmid was then conjugated into *M. huakuii* 7653R.

Plant experiments and microscopic assays.

A. sinicus seeds were germinated in water agar for about 2 days at 24°C after surface-sterilization with 1% sodium hypochlorite and 75% ethanol, as previously described (Zhao et al. 2021). The germinated seedlings were then cultivated in the nitrogen-free liquid Fahraeus medium in pots filled with sterile sand. Plants in the growth chamber were subjected to a daily 16-h-light and 8-h-dark cycle at the temperature of 22 and 20°C, respectively. Rhizobial inoculation was conducted after cotyledon expansion, and plants with root nodules were normally harvested at 28 dpi. Nitrogenase activity expressed as the rate of acetylene reduction was measured as previously described, using gas chromatograph (Model 163, Hatachi, Tokyo) (Li et al. 2008).

Standard light and electron microscopy techniques were used to visualize the cellular structure of root nodule and rhizobial cells. For light microscopy analysis, root nodules (28 dpi) were fixed with formalin-aceto-alcohol buffer, followed by dehydration with absolute ethanol. Nodules embedded in paraffin were cut longitudinally. The resulting nodule slices were stained with toluidine blue and were observed under an Olympus light microscope. Electron microscopy analysis was performed with nodules that were sequentially treated with 2.5% glutaraldehyde for 4 h, postfixing in 1% osmium tetroxide for 3 h, followed by serial ethanol dehydration. The nodules were finally embedded in London resin white, and ultrathin sections were examined by

Table 1. Bacterial strains and plasmids used in this work

Strain or plasmid	Genotype and phenotype	Antibiotic resistance[a]	Source
M. huakuii			
7653R	Wild type, Nod^+, Fix^+	Str^R	Zhang et al. 2000
7653RΔ*bacA*	7653R devoid of *bacA* with an insertion of Tet^R cassette, Nod^+, Fix^-	Str^R, Tet^R	Tan et al. 2009
7653RΔ*groEL3*	7653R devoid of *groEL3*	Str^R, Neo^R	This work
7653R-pMP2463	7653R marked with e*gfp* in plasmid pMP2463	Str^R, Gm^R	Zhao et al. 2021
Plasmid			
pK19mob	Integration suicide vector, Mob^+	Km^R	Schäfer et al. 1994
pRK2013	Helper plasmid, Tra^+	Km^R	Ditta et al. 1980
pK19mob-*groEL3*	pK19mob derivative containing a 1032 bp PCR product internal to the *groEL3* gene	Km^R	This work
pBBR1MCS5	Gene expression vector, Km^R		Kovach et al. 1995
pBBR5-*groEL3*	Wild-type *groEL3* cloned into pBBR1MCS-5		This work
pSRKGm	Gene expression vector with the P_{lac} promoter	Gm^R	Khan et al. 2008
pSRK-*AsNCR051*	pSRKGm carrying AsNCR051 gene	Gm^R	This work
pSRK-*AsNCR083*	pSRKGm carrying AsNCR083 gene	Gm^R	This work
pMP2463	Expression vector for the eGFP fusion	Gm^R	Stuurman et al. 2000a
pMP-*AsNCR051*	pMP2463 carrying *AsNCR051-egfp* fusion	Gm^R	This work
pMP-*AsNCR067*	pMP2463 carrying *AsNCR067-egfp* fusion	Gm^R	This work
pMP-*AsNCR083*	pMP2463 carrying *AsNCR083-egfp* fusion	Gm^R	This work
pTRG	BTH bait expression vector	Tet^R	Stratagene
pTRG-*groEL1*	pTRG with *groEL1* from 7653R	Tet^R	This work
pTRG-*groEL3*	pTRG with *groEL3* from 7653R	Tet^R	This work
pBT	BTH target expression vector	Km^R	Stratagene
pBT-*AsNCR067*	pBT carrying *AsNCR067* gene	Km^R	This work
pBT-*AsNCR076*	pBT carrying *AsNCR076* gene	Km^R	This work
pGEX-6p-1	GST-tagged protein expression vector	Amp^R	GE Healthcare
pGEX-*AsNCR067*	pGEX-6p-1 expressing mature AsNCR067	Amp^R	This work

[a] Str^R, Tet^R, Neo^R, Gm^R, and Km^R indicate resistance to streptomycin, tetracycline, neomycin, gentamicin, and kanamycin, respectively.

TEM. To visualize the GFP-tagged AsNCR067 peptide in bacteroids, nodules on 28 dpi were chopped with a surgical blade and were then directly observed under an Olympus fluorescence microscope.

RNA-seq and *in silico* analysis.

The TRIzol reagent (Aidlab Biotech) was used to prepare total RNA from *A. sinicus* L. roots or nodules formed by WT 7653R or the isogenic Δ*bacA* mutant. Plant materials resuspended in the TRIzol reagent were first frozen in liquid nitrogen and were then subjected to grinding with a mortar and pestle. Residual genomic DNAs were removed by treating the RNA samples with DNaseI (Takara). Libraries were constructed using the Truseq RNA sample prep kit. RNA-seq and analysis of the clean reads were performed on an Illumina HiSeq 4000, using the services of Majorbio Bio-Pharm Technology Co., Ltd. Raw RNA-seq data in fastq format were processed using SeqPrep and Sickle to remove the adapters and sequences of low quality (<20 nt). Rhizobial sequences were removed by mapping to the *M. huakuii* 7653R genome (accession number NZ_CP006581.1), using Bowtie2 (Langmead and Salzberg 2012). Due to the absence of reference genome sequences, the remaining plant sequences are subjected to *de novo* assembly, using Trinity v2.8.5. The resultant clean unigenes were then used in blastx searches for homologues in various databases, including STRING, SWISS-PROT, National Center for Biotechnology Information (NCBI) non-redundant protein, COG/KOG (Cluster of Orthologous Groups of Proteins), and KEGG (Kyoto Encyclopedia of Genes and Genomes). Genes were identified with an E-value 10^{-5} against sequences deposited in the database.

The DEGs were determined with FDR < 0.05 and $\log_2(FC) \geq 1$, using EdgeR (Robinson et al. 2010). GO enrichment analysis of all DEGs was implemented by using blast2GO (Gotz et al. 2008). The Fisher's exact test with adjusted P value < 0.05 was used to identify the significantly enriched GO functions, using Goatools available in GitHub. *Medicago truncatula* NCRs were used as queries to search for putative NCR homologs in the assembled nodule transcriptomes through tBLASTn. Signal peptide was predicted through the online server SignalP 4.0 (Petersen et al. 2011). Molecular weight and pI were calculated in the Swiss Bioinformatics resource portal Expasy. Multisequence alignment of NCRs was performed by MEGA 7.0 (Tamura et al. 2013). For phylogenetic analysis, sequences of mature NCRs were aligned and bootstrapped (1,000 iterations) with ClustalX 2.1 (Larkin et al. 2007). Unrooted trees were displayed and manipulated using Interactive Tree of Life v5 (Letunic and Bork 2021).

Purification of GST-tagged AsNCR peptides.

AsNCR067 was custom-synthesized by GenScript Inc. with greater than 95% (wt/vol) purity. For other AsNCRs, the mature peptide-coding region was amplified by PCR, and the PCR product was cloned into pGEX-6p-1 at the *Eco*RI and *Xho*I site. The GST-AsNCR recombinant protein was purified using affinity chromatography (Zhou et al. 2019). The purity was checked through SDS-PAGE.

Rhizobial sensitivity to AsNCR was examined using the method of Van de Velde et al. (2010), with slight modifications. Briefly, *M. huakuii* 7653R grown in mid-exponential phase in TY broth (optical density at 600 nm [OD_{600}] 0.3 to 0.5) was diluted to $OD_{600} = 0.1$ with 10 mM sodium phosphate buffer (pH 7.0). Next, 200 μl of bacterial suspension was subjected to treatment with peptide at the different concentrations for 6 to 12 h at 28°C. Cell viability was examined by dilute plating onto TY agar plates for 5 to 6 days at 28°C. Effects on membrane permeability were assessed by adding 10 μg of propidium iodide per milliliter into the AsNCR-treated cell suspensions. Fluorescence was measured using a VICTOR Nivo plate reader (PerkinElmer Inc.) with an excitation/emission wavelength of 535/615 nm. To measure DNA content, rhizobial cells were synchronized before they were treated by different concentrations of AsNCR067 for 6 to 12 h (Zhao et al. 2021). Next, cells were subject to dilution and DNA staining with 20 μg of 4′,6-diamidino-2-phenylindole per milliliter for 30 min at room temperature, followed by cell-sorting using the CytoFLEX LX flow cytometer (Beckman Coulter Inc.). Cell synchronization was conducted as described previously (De Nisco et al. 2014), with slight modifications. Briefly, cells in early exponential phase cultures were spun down and re-suspended in minimal medium containing 50 mM 3-(N-morpholino) propane sulfonic acid (MOPS), 1 mM $MgSO_4$, 0.25 9 mM $CaCl_2$, 19 mM glutamic acid, and 0.004 mM biotin (pH 7.0). Cells were cultured at 28°C in a constant-temperature incubator shaker for 12 h to reach synchronization. The synchronized cells were then collected by centrifugation and cultured in TY medium.

B2H assays.

The BacterioMatch II two-hybrid system library construction kit (Stratagene) was used to assess the potential direct protein-protein interactions between AsNCRs and GroEL chaperones. The AsNCR mature peptide region was expressed in the bait plasmid pBT, whereas the GroEL1 or GroEL3 chaperone gene was cloned into the target plasmid pTRG. Next, the recombinant pBT and pTRG plasmid in pairs were transformed into *E. coli* XL-Blue. The interaction was examined by growing the resultant bacteria (24 h at 37°C and, then, 12 h at room temperature) on selective screening medium supplemented with 3-amino-1,2,4-triazole, according to the manufacturer instructions (Zhao et al. 2021). The initial B2H assays were performed similarly by constructing a genomic DNA library of *M. huakuii* 7653R in the plasmid vector pTRG (DNA fragment length 1.5 to 3 Kb), and candidate clones were subjected to the Singer's DNA sequencing.

Data availability.

Raw sequence data were submitted to the NCBI Sequence Reads Archive with a BioProject accession number PRJNA776030.

ACKNOWLEDGMENTS

We thank F. Wang at the State Key Laboratory of Agricultural Microbiology Core Facility for assistance with fluorescence-activated cell sorting analysis.

LITERATURE CITED

Alunni, B., and Gourion, B. 2016. Terminal bacteroid differentiation in the legume-rhizobium symbiosis: Nodule-specific cysteine-rich peptides and beyond. New Phytol. 211:411-417.

Alunni, B., Kevei, Z., Redondo-Nieto, M., Kondorosi, A., Mergaert, P., and Kondorosi, E. 2007. Genomic organization and evolutionary insights on *GRP* and *NCR* genes, two large nodule-specific gene families in *Medicago truncatula*. Mol. Plant-Microbe Interact. 20:1138-1148.

Barriere, Q., Guefrachi, I., Gully, D., Lamouche, F., Pierre, O., Fardoux, J., Chaintreuil, C., Alunni, B., Timchenko, T., Giraud, E., and Mergaert, P. 2017. Integrated roles of BclA and DD-carboxypeptidase 1 in *Bradyrhizobium* differentiation within NCR-producing and NCR-lacking root nodules. Sci. Rep. 7:9063.

Bittner, A. N., Foltz, A., and Oke, V. 2007. Only one of five *groEL* genes is required for viability and successful symbiosis in *Sinorhizobium meliloti*. J. Bacteriol. 189:1884-1889.

Chou, M. X., Wei, X. Y., Chen, D. S., and Zhou, J. C. 2006. Thirteen nodule-specific or nodule-enhanced genes encoding products homologous to cysteine cluster proteins or plant lipid transfer proteins are identified in *Astragalus sinicus* L. by suppressive subtractive hybridization. J. Exp. Bot. 57:2673-2685.

De Nisco, N. J., Abo, R. P., Wu, C. M., Penterman, J., and Walker, G. C. 2014. Global analysis of cell cycle gene expression of the legume symbiont *Sinorhizobium meliloti*. Proc. Natl. Acad. Sci. U.S.A. 111: 3217-3224.

Ditta, G., Stanfield, S., Corbin, D., and Helinski, D. R. 1980. Broad host range DNA cloning system for gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti*. Proc. Natl. Acad. Sci. U.S.A. 77:7347-7351.

Downie, A. J., and Kondorosi, E. 2021. Why should nodule cysteine-rich (NCR) peptides be absent from nodules of some groups of legumes but essential for symbiotic N-fixation in others? Front. Agron. 3:Article 654576.

Farkas, A., Maroti, G., Durgo, H., Gyorgypal, Z., Lima, R. M., Medzihradszky, K. F., Kereszt, A., Mergaert, P., and Kondorosi, E. 2014. *Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. Proc. Natl. Acad. Sci. U.S.A. 111:5183-5188.

Garcia-Calderon, M., Perez-Delgado, C. M., Credali, A., Vega, J. M., Betti, M., and Marquez, A. J. 2017. Genes for asparagine metabolism in *Lotus japonicus*: Differential expression and interconnection with photorespiration. BMC Genomics 18:781.

Gotz, S., Garcia-Gomez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talon, M., Dopazo, J., and Conesa, A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36:3420-3435.

Haag, A. F., Baloban, M., Sani, M., Kerscher, B., Pierre, O., Farkas, A., Longhi, R., Boncompagni, E., Herouart, D., Dall'angelo, S., Kondorosi, E., Zanda, M., Mergaert, P., and Ferguson, G. P. 2011. Protection of *Sinorhizobium* against host cysteine-rich antimicrobial peptides is critical for symbiosis. PLoS Biol. 9:e1001169.

Haag, A. F., and Mergaert, P. 2019. Terminal bacteroid differentiation in the *Medicago–Rhizobium* interaction—A tug of war between plant and bacteria. Pages 600-616 in: The Model Legume *Medicago truncatula*. F. de Bruijn, ed. John Wiley & Sons, New York.

Howing, T., Dann, M., Muller, B., Helm, M., Scholz, S., Schneitz, K., Hammes, U. Z., and Gietl, C. 2018. The role of KDEL-tailed cysteine endopeptidases of *Arabidopsis* (AtCEP2 and AtCEP1) in root development. PLoS One 13:e0209407.

Kang, W., Jiang, Z., Chen, Y., Wu, F., Liu, C., Wang, H., Shi, S., and Zhang, X. X. 2020. Plant transcriptome analysis reveals specific molecular interactions between alfalfa and its rhizobial symbionts below the species level. BMC Plant Biol. 20:293.

Khan, S. R., Gaines, J., Roop, R. M., 2nd, and Farrand, S. K. 2008. Broad-host-range expression vectors with tightly regulated promoters and their use to examine the influence of TraR and TraM expression on Ti plasmid quorum sensing. Appl. Environ. Microbiol. 74:5053-5062.

Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., 2nd, and Peterson, K. M. 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene 166:175-176.

Kumar, C. M., Mande, S. C., and Mahajan, G. 2015. Multiple chaperonins in bacteria - novel functions and non-canonical behaviors. Cell Stress Chaperones 20:555-574.

Langmead, B., and Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357-359.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948.

Ledermann, R., Schulte, C. C. M., and Poole, P. S. 2021. How rhizobia adapt to the nodule environment. J. Bacteriol. 203:e0053920.

Letunic, I., and Bork, P. 2021. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49:W293-W296.

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.

Maroti, G., Downie, J. A., and Kondorosi, E. 2015. Plant cysteine-rich peptides that inhibit pathogen growth and control rhizobial differentiation in legume nodules. Curr. Opin. Plant Biol. 26:57-63.

Masson-Boivin, C., and Sachs, J. L. 2018. Symbiotic nitrogen fixation by rhizobia-the roots of a success story. Curr. Opin. Plant Biol. 44:7-15.

Mergaert, P., Nikovics, K., Kelemen, Z., Maunoury, N., Vaubert, D., Kondorosi, A., and Kondorosi, E. 2003. A novel family in Medicago truncatula consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. Plant Physiol. 132:161-173.

Montiel, J., Downie, J. A., Farkas, A., Bihari, P., Herczeg, R., Balint, B., Mergaert, P., Kereszt, A., and Kondorosi, E. 2017. Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. Proc. Natl. Acad. Sci. U.S.A. 114:5041-5046.

Nicoud, Q., Barriere, Q., Busset, N., Dendene, S., Travin, D., Bourge, M., Le Bars, R., Boulogne, C., Lecroel, M., Jenei, S., Kereszt, A., Kondorosi, E., Biondi, E. G., Timchenko, T., Alunni, B., and Mergaert, P. 2021. *Sinorhizobium meliloti* functions required for resistance to antimicrobial NCR peptides and bacteroid differentiation. mBio 12:e0089521.

Ott, T., van Dongen, J. T., Gunther, C., Krusell, L., Desbrosses, G., Vigeolas, H., Bock, V., Czechowski, T., Geigenberger, P., and Udvardi, M. K. 2005. Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. Curr. Biol. 15:531-535.

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

Peng, J., Hao, B., Liu, L., Wang, S., Ma, B., Yang, Y., Xie, F., and Li, Y. 2014. RNA-seq and microarrays analyses reveal global differential transcriptomes of *Mesorhizobium huakuii* 7653R between bacteroids and free-living cells. PLoS One 9:e93626.

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat. Methods 8:785-786.

Poole, P., Ramachandran, V., and Terpolilli, J. 2018. Rhizobia: From saprophytes to endosymbionts. Nat. Rev. Microbiol. 16:291-303.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. 2010. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140.

Rodriguez-Quinones, F., Maguire, M., Wallington, E. J., Gould, P. S., Yerko, V., Downie, J. A., and Lund, P. A. 2005. Two of the three *groEL* homologues in *Rhizobium leguminosarum* are dispensable for normal growth. Arch. Microbiol. 183:253-265.

Roux, B., Rodde, N., Jardinaud, M. F., Timmers, T., Sauviac, L., Cottret, L., Carrere, S., Sallet, E., Courcelle, E., Moreau, S., Debelle, F., Capela, D., de Carvalho-Niebel, F., Gouzy, J., Bruand, C., and Gamas, P. 2014. An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. Plant J. 77:817-837.

Roy, P., Achom, M., Wilkinson, H., Lagunas, B., and Gifford, M. L. 2020. Symbiotic outcome modified by the diversification from 7 to over 700 nodule-specific cysteine-rich peptides. Genes (Basel) 11:348.

Sathoff, A. E., and Samac, D. A. 2019. Antibacterial activity of plant defensins. Mol. Plant-Microbe Interact. 32:507-514.

Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G., and Pühler, A. 1994. Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: Selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. Gene 145: 69-73.

Silva, L., and Carvalho, H. 2013. Possible role of glutamine synthetase in the NO signaling response in root nodules by contributing to the antioxidant defenses. Front. Plant Sci. 4:372.

Simon, R., Priefer, U., and Pühler, A. 1983. A broad host range mobilization system for *in vivo* genetic engineering: Transposon mutagenesis in gram negative bacteria. Bio/Technology 1:784-791.

Stuurman, N., Bras, C. P., Schlaman, H. R. M., Wijfjes, A. H. M., Bloemberg, G., and Spaink, H. P. 2000a. Use of green fluorescent protein color variants expressed on stable broad-host-range vectors to visualize rhizobia interacting with plants. Mol. Plant-Microbe Interact. 13: 1163-1169.

Stuurman, N., Pacios Bras, C., Schlaman, H. R., Wijfjes, A. H., Bloemberg, G., and Spaink, H. P. 2000b. Use of green fluorescent protein color variants expressed on stable broad-host-range vectors to visualize rhizobia interacting with plants. Mol. Plant-Microbe Interact. 13: 1163-1169.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30:2725-2729.

Tan, X. J., Cheng, Y., Li, Y. X., Li, Y. G., and Zhou, J. C. 2009. BacA is indispensable for successful *Mesorhizobium-Astragalus* symbiosis. Appl. Microbiol. Biotechnol. 84:519-526.

Turner, S. L., Zhang, X. X., Li, F. D., and Young, J. P. W. 2002. What does a bacterial genome sequence represent? Mis-assignment of MAFF 303099 to the genospecies *Mesorhizobium loti*. Microbiology (Reading) 148:3330-3331.

Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., Farkas, A., Mikulass, K., Nagy, A., Tiricz, H., Satiat-Jeunemaitre, B., Alunni, B., Bourge, M., Kucho, K., Abe, M., Kereszt, A., Maroti, G., Uchiumi, T., Kondorosi, E., and Mergaert, P.

2010. Plant peptides govern terminal differentiation of bacteria in symbiosis. Science 327:1122-1126.

Wang, Q., Yang, S., Liu, J., Terecskei, K., Abraham, E., Gombar, A., Domonkos, A., Szucs, A., Kormoczi, P., Wang, T., Fodor, L., Mao, L., Fei, Z., Kondorosi, E., Kalo, P., Kereszt, A., and Zhu, H. 2017. Host-secreted antimicrobial peptide enforces symbiotic selectivity in *Medicago truncatula*. Proc. Natl. Acad. Sci. U.S.A. 114:6854-6859.

West, S. A., Griffin, A. S., Gardner, A., and Diggle, S. P. 2006. Social evolution theory for microorganisms. Nat. Rev. Microbiol. 4:597-607.

Yang, S., Wang, Q., Fedorova, E., Liu, J., Qin, Q., Zheng, Q., Price, P. A., Pan, H., Wang, D., Griffitts, J. S., Bisseling, T., and Zhu, H. 2017. Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula*. Proc. Natl. Acad. Sci. U.S.A. 114:6848-6853.

Zhang, X. X., and Rainey, P. B. 2010. Bet hedging in the underworld. Genome Biol. 11:137.

Zhang, X. X., Turner, S. L., Guo, X. W., Yang, H. J., Debelle, F., Yang, G. P., Denarie, J., Young, J. P., and Li, F. D. 2000. The common nodulation genes of *Astragalus sinicus* rhizobia are conserved despite chromosomal diversity. Appl. Environ. Microbiol. 66:2988-2995.

Zhao, W., Zhu, H., Wei, F., Zhou, D., Li, Y., and Zhang, X. X. 2021. Investigating the involvement of cytoskeletal proteins MreB and FtsZ in the origin of legume-rhizobial symbiosis. Mol. Plant-Microbe Interact. 34:547-559.

Zhou, D., Li, Y., Wang, X., Xie, F., Chen, D., Ma, B., and Li, Y. 2019. *Mesorhizobium huakuii* HtpG interaction with nsLTP AsE246 is required for symbiotic nitrogen fixation. Plant Physiol. 180:509-528.