*Journal of Experimental Botany*, Vol. 72, No. 4 pp. 1104–1118, 2021
doi:10.1093/jxb/eraa505 Advance Access Publication 1 November 2020

RESEARCH PAPER

# Natural polymorphisms in a pair of NSP2 homoeologs can cause loss of nodulation in peanut

**Ze Peng[1], Huiqiong Chen[2], Lubin Tan[3], Hongmei Shu[1], Rajeev K. Varshney[4], Zhekai Zhou[2], Zifan Zhao[1], Ziliang Luo[1], Annapurna Chitikineni[4], Liping Wang[1], James Maku[5], Yolanda López[1], Maria Gallo[6], Hai Zhou[1,2,*] and Jianping Wang[1,7,*]**

[1] Agronomy Department, University of Florida, Gainesville, FL 32610, USA
[2] State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, College of Life Sciences, South China Agricultural University, Guangzhou 510642, China
[3] Department of Plant Genetics and Breeding, China Agricultural University, Beijing 100193, China
[4] Center of Excellence in Genomics and Systems Biology, International Crops Research Institute for the Semi-Arid Tropics, Patancheru 502324, India
[5] Sciences and Mathematics Department, Glenville State College, Glenville, WV 26351, USA
[6] Delaware Valley University, Doylestown, PA 18901, USA
[7] Genetics Institute, Plant Molecular and Cellular Biology Program, University of Florida, Gainesville, FL 32610, USA

* Correspondence: wangjp@ufl.edu or haizhou@scau.edu.cn

Received 25 June 2020; Editorial decision 16 October 2020; Accepted 27 October 2020

Editor: Miriam Gifford, University of Warwick, UK

## Abstract

**Microbial symbiosis in legumes is achieved through nitrogen-fixing root nodules, and these are important for sustainable agriculture. The molecular mechanisms underlying development of root nodules in polyploid legume crops are largely understudied. Through map-based cloning and QTL-seq approaches, we identified a pair of homoeologous GRAS transcription factor genes, *Nodulation Signaling Pathway 2* (*AhNSP2-B07* or *Nb*) and *AhNSP2-A08* (*Na*), controlling nodulation in cultivated peanut (*Arachis hypogaea* L.), an allotetraploid legume crop, which exhibited non-Mendelian and Mendelian inheritance, respectively. The segregation of nodulation in the progeny of *Nananbnb* genotypes followed a 3:1 Mendelian ratio, in contrast to the 5:3~1:1 non-Mendelian ratio for *nanaNbnb* genotypes. Additionally, a much higher frequency of the *nb* allele (13%) than the *na* allele (4%) exists in the peanut germplasm collection, suggesting that *Nb* is less essential than *Na* in nodule organogenesis. Our findings reveal the genetic basis of naturally occurred non-nodulating peanut plants, which can be potentially used for nitrogen fixation improvement in peanut. Furthermore, the results have implications for and provide insights into the evolution of homoeologous genes in allopolyploid species.**

**Keywords:** Allotetraploid, GRAS transcription factor, nodulation, non-Mendelian inheritance, NSP2, peanut, symbiosis.

## Introduction

Legumes fix nitrogen ($N_2$) biologically through a symbiotic association with $N_2$-fixing rhizobia within root nodules. This symbiosis is impactful as it contributes approximately 21.45 million tons of N to legume crops annually (Herridge *et al.*, 2008), and accounts for 30–60% of the total N accumulated in plants (Peoples *et al.*, 2009). Extensive studies have been carried

out to understand the molecular and genetic mechanisms of legume symbiosis, mostly by studying model legumes, such as *Lotus japonicus* and *Medicago truncatula*, which are diploids. To briefly summarize, legumes release flavonoids that attract rhizobia in the soil, and in response, rhizobia release specific lipochitooligosaccharides, so called nodulation factors or Nod factors (NFs). The NFs are recognized by NF receptors on the membranes of root epidermal cells, Nod Factor Receptor 1 (NFR1) and NFR5 in *L. japonicus* (Madsen *et al.*, 2003; Radutoiu *et al.*, 2003) or LysM-type Receptor-like Kinase 3 (LYK3) and Nod Factor Perception (NFP) in *M. truncatula* (Limpens *et al.*, 2003; Arrighi *et al.*, 2006). Following NF recognition, the signal is transduced in infected root cells to induce nuclear-associated calcium oscillations, which are subsequently decoded by the Calcium- and Calmodulin-dependent Protein Kinase (CCaMK)/Doesn't Make Infections 3 (DMI3) and CYCLOPS/Interacting Protein of DMI3 (IPD3) complex (Levy *et al.*, 2004; Mitra *et al.*, 2004; Messinese *et al.*, 2007; Yano *et al.*, 2008). The signal is further transmitted through several key transcription factors such as Nodulation Signaling Pathway 1 (NSP1), NSP2, Nodule Inception (NIN), and ERF Required for Nodulation (ERN), which coordinate the expression of genes involved in rhizobial infection and nodule morphogenesis (Oldroyd and Long, 2003; Smit *et al.*, 2005; Marsh *et al.*, 2007; Middleton *et al.*, 2007). However, nodulation in many polyploid legume crops, such as cultivated peanut (*Arachis hypogaea*), alfalfa (*Medicago sativa*), and soybean (*Glycine max*; paleopolyploid), has received less attention. It has yet to be determined how legume symbiosis would accommodate polyploidization and what the evolutionary fate would be of multiple copies of symbiosis genes with redundant functions.

Many legume species exhibit whole genome duplication (WGD) or polyploidization. Papilionoid legumes, such as *M. truncatula*, *L. japonicus*, *G. max*, and *A. hypogaea*, share a WGD event around 58 million years ago (MYA) (Schlueter *et al.*, 2004; Cannon *et al.*, 2006; Bertioli *et al.*, 2009), and some lineages underwent more recent duplication events such as *G. max* (5–13 MYA) and *A. hypogaea* (~3.5 MYA and 0.42–0.47 MYA) (Schmutz *et al.*, 2010; Kang *et al.*, 2014; Chen *et al.*, 2016; Zhuang *et al.*, 2019). The hybridization and genome doubling can affect the polyploids. An extensive review on auto- or allopolyploid legumes showed that N fixation was enhanced in polyploids (natural or synthesized) when compared with their diploid ancestors (Forrester and Ashman, 2018). In the allopolyploid *Glycine dolichocarpa*, enhanced nodulation capacity was observed compared with its diploid ancestors (Powell and Doyle, 2016). In *M. truncatula*, several nodulation genes retained after the ancient WGD event underwent sub- or neofunctionalization, which has increased the complexity of the legume–rhizobia symbiosis (Young *et al.*, 2011). Non-additive contributions of genes on sub-genomes to gene expression have been frequently observed in some other non-legume allopolyploids (Bottani *et al.*, 2018). For instance, in cultivated cotton (*Gossypium hirsutum*), which evolved 1–2 MYA, silencing or unequal expression of homoeologs was reported for 10 out of 40 homoeologous gene pairs evaluated (Adams *et al.*, 2003). Some of the silencing events were epigenetically induced (Adams *et al.*, 2003).

Cultivated peanut is an allotetraploid ($2n=4x=40$, AABB) crop species, in which rhizobial infection takes the crack-entry form in contrast to the root hair entry form (with infection thread formation) of model legumes (Sprent and James, 2007; Bertioli *et al.*, 2016; Maku *et al.*, 2018). Currently, few studies on the molecular mechanisms of symbiosis have been reported in species with crack-entry. The crack-entry mechanism is promising and more realistic for engineering biological nitrogen fixation into non-legumes, as a relatively small set of genetic adaptations from bacteria and the host would be involved (Ibáñez *et al.*, 2017). Although not reported in the germplasm, non-nodulating (Nod−) plants have been discovered in cultivated peanut during breeding processes through hybridizations. The Nod− peanut lines were first reported in the 1970s in an $F_3$ population from the cross between two nodulating (Nod+) genotypes, 487A-4-1-2 and PI 262090 (Gorbet and Burton, 1979). Although limited polymorphisms have been reported in peanut, natural variability in nodulation is evidenced during this breeding work. Subsequently, another non-nodulation line was discovered through hybridization of two normally Nod+ germplasm, and a two-gene model was proposed (Dutta and Reddy, 1988). Further investigation of the Nod− inheritance in the cross between 487A-4-1-2 and PI 262090 generated a three-gene model explaining the genetics of peanut nodulation (Gorbet and Burton, 1979; Gallo-Meagher *et al.*, 2001). The Nod− plants are precious genetic material for the study of nodulation in peanut. However, so far, no nodulation genes have been mapped, cloned, or characterized in peanut through a forward genetics approach. Recently, two transcriptome studies revealed numerous genes involved in regulation of nodulation in peanut, including genes specific to peanut, which share no homology with those of model legumes, and putative orthologs of symbiosis genes having a different expression pattern from those of model legumes, indicating a unique gene regulation network in the peanut symbiosis (Peng *et al.*, 2017; Karmakar *et al.*, 2019). Through a reverse genetics approach, several genes involved in nodulation were characterized in peanut including those for CCaMK, HK1, and CYCLOPS. However, their sequence structures and functions in peanut were quite conserved compared with model legumes (Sinharoy and DasGupta, 2009; Kundu and DasGupta, 2018; Das *et al.*, 2019). How homoeologous genes in sub-genomes interact to convey nodulation in this allopolyploid species has not been addressed yet, and this is a fascinating question in regard to genome evolution and polyploidization.

In this study, we mapped and cloned a pair of homoeologous genes controlling nodulation in cultivated peanut through a forward genetics approach. Functional analysis and characterization of this pair of homoeologous genes revealed divergent inheritance modes and evolutionary fate of this pair of homoeologous genes and different gene expression patterns involving nodule organogenesis from model legume species.

## Materials and methods

### *Plant materials*

The four peanut lines PI 262090 (Nod+), UF 487A (Nod+), E6 (Nod+), and E7 (Nod−) were used in this study, which were previously described (Peng *et al.*, 2017). Briefly, E6 and E7 are sister recombinant inbred lines (RILs) that were derived from the original cross between PI 262090 and UF 487A. For primary genetic mapping of nodulation genes, a small proportion of an $F_2$ population, comprising 90 lines, derived from a cross between E6 and E7 was planted in a greenhouse in 2013. For fine mapping and gene cloning, the rest of the samples in the $F_2$ population, comprising 1223 lines, were planted in the field at the University of Florida Plant Science Research and Education Unit (UF-PSREU) in 2014. The nodulation phenotype of the plants was recorded after digging and visual assessment in the field. Plants with nodules were recorded as Nod+, and those with no nodules were recorded as Nod−. Each plant was independently recorded twice. In addition, E6, E7, and $F_3$ families were planted at the UF-PSREU in 2018 for chlorophyll content (E6, E7) and pod yield (E6, E7, and $F_3$ plants) measurement, and segregation analysis. Chlorophyll content, indicated by soil plant analysis development (SPAD) values, of E6 and E7 was measured following the same method as previously described (Peng *et al.*, 2018) at 2.5 months (77 d) after planting. Upon harvest, 10 plants from E6, E7, Nod+, or Nod− $F_3$ plants were randomly collected for measuring pod yield after drying. Parental lines and hairy root transgenic lines as described below were planted in a growth chamber (16 h light at 27 °C; 8 h dark at 25 °C; light intensity 130–150 μmol $m^{-2}$ $s^{-1}$). A commercial liquid inoculum of rhizobia, Optimize® liquid for peanut (Monsanto, St Louis, MO, USA), was applied at planting in the field, and a strain of *Bradyrhizobium* Lb8 (Peng *et al.*, 2017) was applied under growth chamber conditions.

### *QTL mapping*

Primary genetic mapping was carried out using the small $F_2$ population of 90 lines. A total of 31 polymorphic simple sequence repeat (SSR) markers, identified between E6 and E7 and located on 12 chromosomes (Peng *et al.*, 2018), were utilized. QTL analysis was performed using QTL IciMapping software (Meng *et al.*, 2015). Fine mapping of the nodulation genes was carried out using Nod− plants obtained from the larger $F_2$ population. The primers for candidate genes (including the candidate gene from QTL-seq analysis as below) were designed using Primer 3 (http://bioinfo.ut.ee/primer3/) and synthesized by Thermo Fisher Scientific. PCR reactions as well as the separation of PCR products using PAGE were performed as previously described (Tseng *et al.*, 2016). The candidate genes among the four parental lines PI 262090, UF 487A, E6, and E7 were sequenced by the Sanger method at the Interdisciplinary Center for Biotechnology Research (ICBR) of the University of Florida.

### *QTL-seq analysis*

A total of 20 Nod+ and 20 Nod− $F_2$ samples were included for this experiment. The DNA of 20 Nod+ (Nod+ pool) and 20 Nod− samples (Nod− pool) were pooled in equimolar amounts. The pooled DNA samples as well as the DNA samples of the two parental lines, E6 and E7, were sequenced using the Illumina HiSeq 2500 platform (125 bp paired-end reads). Raw sequencing reads were trimmed by Trimmomatic (Bolger *et al.*, 2014). Trimmed reads were aligned to the two diploid reference genomes of peanut, *Arachis duranensis* and *Arachis ipaensis* (downloaded from PeanutBase, https://peanutbase.org), using BWA-mem (Li and Durbin, 2009). For subsequent single nucleotide polymorphism (SNP) calling, only uniquely mapped reads were retrieved by filtering off reads with a mapping quality of zero and 'XA:Z' tag. SNP calling was performed using Samtools 1.3.1 (Li *et al.*, 2009). SNPs were filtered based on depth for subsequent analysis as follows. All SNP loci were included for downstream analysis if E6 showed a homozygous genotype (read depth ≥4) and both Nod+ and Nod− pools had at least 10 covered reads. For both the Nod+ and Nod− pools, the E6 allele was treated as the 'reference allele', while the non-E6 allele was treated as the 'alternative allele'. The SNP index value for a SNP locus was calculated as the percentage of alternative allele. ΔSNP index=SNP $index_{Nod-}$−SNP $index_{Nod+}$. The SNP information was formatted following the requirements of input data frame of QTL-seqr (Mansfeld and Grumet, 2018). The 'runQTLseqAnalysis' function of QTL-seqr was used (windowSize=2 Mb, popStruc='F2', bulkSize=20, replications=10 000, intervals=c(95,99)).

### *Vector construction and plant transformation*

To prepare the overexpression construct of *AhNSP2-B07*, a 1783 bp fragment containing the coding region of *AhNSP2-B07* was amplified (primers AhNSP2-B07OE3-F/R) from the E6 line and inserted into the *Bam*HI/*Spe*I sites of the pCmHU vector (Wang *et al.*, 2006) driven by the maize ubiquitin (pUbi) promoter. For *AhNSP2-A08*, a 6365 bp fragment containing the entire gene region and 2606 bp promoter region of *AhNSP2-A08* was amplified from the E6 line (primers AhNSP2-A08C2-F/R), which was inserted into the *Hin*dIII site of vector pCAMBIA1300 (modified to contain green fluorescent protein (GFP)) to make the complementary construct. These constructs as well as the empty constructs were transformed into *Agrobacterium rhizogenes* strain K599, which was subsequently used for peanut hairy root transformation as previously described (Kereszt *et al.*, 2007). For AhNSP2-B07OE lines, the positive transformants were selected based on PCR using primers flanking the hygromycin resistance marker gene in the construct. For AhNSP2-A08C lines the positive transformants were selected based on both GFP fluorescence and PCR using primers flanking the hygromycin and GFP genes in the construct. Genotypes of AhNSP2-B07OE and AhNSP2-A08C lines were finally confirmed using Sanger sequencing.

### *Quantitative RT-PCR*

E6 and E7 lines were inoculated at 6 d after germination using a *Bradyrhizobium* strain, Lb8. The roots of E6 and E7 lines were collected at 0, 0.5, 1, 2, 3, 5, 7 and 9 d after inoculation (DAI) with three biological replicates at each time point. To investigate the influence of phytohormones on the expression of *AhNSP2*, the wild type was treated with $GA_3$, 6-benzylaminopurine (6-BA), and indole-3-acetic acid (IAA) ($10^{-6}$ M) at 10 d after germination with water as a control. Roots were collected at 0, 0.5, 1, 3, 12, and 24 h after treatment with three biological replicates at each time point. Total RNA from roots was extracted using TRIzol (Thermo Fisher Scientific) in accordance with the product instructions. For E6 line, total RNAs from roots, stems, leaves, and flowers were also extracted. cDNA synthesis and quantitative PCR with three technical replicates for each cDNA sample were performed as previously described (Peng *et al.*, 2017). The peanut orthologs to several characterized genes in model legumes were identified using Blast (E-value 1e−5) and OrthoMCL (Inflation value 1.5; Fischer *et al.*, 2011).

### Subcellular localization

The open reading frames of both *AhNSP2-A08* and *AhNSP2-B07* were amplified and fused to the N terminus of GFP in the pUC18 vector (Clontech/Takara, USA) driven by the CaMV35S promoter. A ClonexPress II One Step Cloning Kit (Vazyme, China) was used. The protoplast transformation system in rice using protoplasts from leaf sheaths was utilized as previously described (Zhou *et al.*, 2014). An Olympus BX51 fluorescence microscope was used for GFP signal observation.

### Yeast two-hybrid assay and bimolecular fluorescence complementation assay

The full-length coding regions of *AhNSP1-A09*, *AhNSP1-B09*, *AhNSP2-A08*, and *AhNSP2-B07* were amplified and cloned into vectors pGBKT7 (BD) and pACT2 (AD). The bait vector was transformed into yeast strain Y187. The prey vector was transformed into yeast strain AH109. The mixed transformed yeast cells were grown on synthetic dropout medium lacking Trp, Leu, His, and adenine (−WLHA) for 3–5 d at 30 °C. To make bimolecular fluorescence complementation (BiFC) constructs, the coding regions of *AhNSP2-A08* and *AhNSP2-B07* were cloned into pCAMBIA1300-35S-nYFP-C (NY) vector, named NY-NSP2A and NY-NSP2B, respectively. The coding regions of *AhNSP1-A09*, *AhNSP1-B09*, and *AhNSP2-B07* were cloned into pCAMBIA1300-35S-cYFP-C (CY) vector, named CY-NSP1A, CY-NSP1B, and CY-NSP2B, respectively. NY-NSP2A and CY-NSP1A, NY-NSP2A and CY-NSP1B, NY-NSP2A and CY-NSP2B, NY-NSP2B and CY-NSP1A, and NY-NSP2B and CY-NSP1B were respectively co-transformed into leaves of tobacco (*Nicotiana benthamiana*) through *Agrobacterium* and examined by confocal microscopy as previously described (Lin *et al.*, 2019; Chen *et al.*, 2020). The combinations NY-NSP2A and CY, NY and CY-NSP1B, NY and CY-NSP2B, NY and CY-NSP1A, and NY-NSP2B and CY, respectively, were used as controls.

### Phylogenetic analysis

A phylogenetic tree was constructed for *AhNSP2* and *AhNSP2-like* from *A. hypogaea*, and the orthologs from *G. max*, *P.vulgaris*, *L. japonicus*, *M.truncatula*, and Arabidopsis. The protein sequences from all species were aligned using Mafft (Katoh *et al.*, 2002). The best protein model 'PROTGAMMADAYHOFF' was selected using a perl script, ProteinModelSelection.pl (https://cme.h-its.org/exelixis/web/software/raxml/). RAxML was used for phylogenetic tree construction with 1000 bootstraps and the non-legume Arabidopsis as an outgroup (Stamatakis, 2014). The best-scoring tree is presented. The non-synonymous substitutions ($K_a$) and synonymous substitutions ($K_s$) of the *AhNSP2*, *AhNSP2-like*, and *AhNSP1* genes were calculated by comparing the two genes from the two sub-genomes, pair-wisely, using the codeml method in the PAML package (Yang, 2007). A total of 3458 genome-wide peanut homoeologs were obtained using the Blast and OrthoMCL method as described above and were also included for $K_a$ and $K_s$ calculation. Annotated gene models from the tetraploid Tifrunner reference genome were used for analysis (Bertioli *et al.*, 2019).

### Investigation of AhNSP2 mutations in the US peanut mini core collection

DNA samples were extracted from a total of 107 accessions in the US peanut mini core (Holbrook and Dong, 2005). An Indel marker was developed for *AhNSP2-B07*, which was used to genotype these accessions at the Indel locus. The first round of PCR was carried out using primers AhNSP2-B07Indel-1F/R. The PCR products were diluted 2500 times, after which the second round of PCR was carried out using primers AhNSP2-B07Indel-2F/R. Final PCR products were resolved on 15% polyacrylamide gel. The genotyping of the *AhNSP2-A08* mutation was carried out using PCR amplification and Sanger sequencing. Primers AhNSP2-A08-1C1-F/R were used for amplification and AhNSP2-A08seq-F was the sequencing primer. All primers in this study were provided in Supplementary Table S1.

### Statistical analysis

The SPAD measurements were compared between E6 and E7, while the pod yield measurements were compared between E6 and E7, and between Nod+ and Nod− $F_3$ progeny using Student's *t*-test in Microsoft Excel. The $\chi^2$ goodness-of-fit test was performed using an online calculator with GraphPad (https://www.graphpad.com/quickcalcs/chisquared1.cfm).

## Results

### Phenotypic characterization of the non-nodulating line E7

To clone nodulation genes, four peanut parental lines were used in this study, including PI 262090, UF 487A (both Virginia botanical types), E6, and E7 (see Supplementary Fig. S1A–D). E6 and E7 were a pair of RILs, which were derived from the cross between PI 262090 and UF 487A, followed by self-pollination for more than six generations (Peng *et al.*, 2018). The three lines PI 262090, UF 487A, and E6 form root nodules normally and exhibit a dark green leaf color with application of rhizobia inoculum and no fertilizers (Supplementary Fig. S1A–D). However, under the same condition, E7 exhibited a yellowish leaf color, a smaller plant size, and no single nodule was observed on the root system (Supplementary Fig. S1C, D). Under field conditions, both the leaf chlorophyll content and pod yield of Nod− E7 were significantly reduced when compared with the Nod+ E6 (Supplementary Fig. S1E, F). Similarly, the pod yield from Nod− $F_3$ plants was significantly reduced when compared with Nod+ $F_3$ plants (Supplementary Fig. S1F). The results illustrated that N fixation in nodules was critical for peanut plant growth, development, and productivity. Since the two original parental lines, PI 262090 and UF 487A, were both Nod+ lines, the production of Nod− E7 implied the existence of different natural recessive mutations of nodulation genes in PI 262090 and UF 487A, which could recombine and lead to the Nod− phenotype.

### Mapping and cloning of AhNSP2-B07 in E7

To genetically map the genes resulting in the Nod− phenotype of E7, one $F_2$ population segregating for nodulation was constructed by crossing the Nod+ E6 and Nod− E7 lines. Firstly, a small proportion of the $F_2$ population of 90 lines was planted in the greenhouse with rhizobial inoculation. Of the 90 lines, 71 were Nod+ and 19 were Nod−. Through quantitative trait locus (QTL) analysis, a QTL with phenotypic variation explained (PVE) of 23% on chromosome B07 flanked by two

markers, GM1713 (logarithm of the odds=5.03) and GM1842, was identified (see Supplementary Fig. S2A, B). We named the gene controlling peanut nodulation in the interval on B07 as $N_b$. To further fine map $N_b$, the rest of the $F_2$ lines, numbering 1223, were planted in the field. Of the 1223 lines, 1088 were Nod+ and 135 were Nod− (Supplementary Table S2). The segregation ratio did not fit a 3 (Nod+): 1 (Nod−) ratio but instead fit 7 (Nod+): 1 (Nod−) or 28 (Nod+): 4 (Nod−). This implied that not a single gene was segregating in this population according to the nodulation phenotype and that a recessive allele controlling Nod− should exist. Subsequently, we utilized 118 Nod− $F_2$ lines (with available good quality DNA) for further fine mapping the $N_b$ gene locus to an interval between markers AhTE1007 and TC14B08 (Fig. 1A). By assuming $n_b$ as the recessive allele, and considering Nod− $N_bn_b$ plants as recombinants, there were 31 recombinants between AhTE1007 and $N_b$ gene locus, and 19 recombinants between TC14B08 and $N_b$ gene locus. Candidate genes were further searched within this 8.4 Mb interval between marker TC3B05 and TC14B08. The genes orthologous to nodulation-related genes identified in model legume species and the differentially expressed genes upon infection of rhizobia from previous transcriptome studies in peanut (Peng *et al.*, 2017; Karmakar *et al.*, 2019) were all considered as potential candidate genes. Among a total of 494 annotated gene models located within this interval, only one ortholog of *Nodulation Signaling Pathway 2* (*NSP2*), a nodulation gene characterized in *L. japonicus* and *M. truncatula* (Kaló *et al.*, 2005; Murakami *et al.*, 2006), was identified. It was up-regulated and among the three differentially expressed genes upon rhizobial infection (Supplementary Table S3). Thus, this *NSP2* ortholog, named as *AhNSP2-B07*, was considered as a priority candidate gene at the $N_b$ locus controlling nodulation in this segregating population.

The amplicon sequences of this gene were then compared between E6 and E7. Results revealed that the 119th base 'C' in the coding frame of $N_b$ gene was deleted in E7, which resulted in a shift of reading frame (Fig. 1A) and subsequently a premature stop codon at 340th–342nd base. The peptide length would be reduced from 516 to 113 amino acids in E7. We subsequently genotyped the 118 Nod− $F_2$ plants. Surprisingly, among these Nod− individuals, 53 were heterozygous ($N_bn_b$) at this Indel locus, while the remaining 65 individuals were homozygous having the same genotype as E7 ($n_bn_b$). None of them showed the homozygous E6 genotype ($N_bN_b$). Thus, we speculated the loss of function of *AhNSP2-B07* should be one of the reasons for the Nod− phenotype of E7. However, not only can a homozygous dysfunctional genotype ($n_bn_b$) at this locus lead to Nod− phenotype, a portion of heterozygous genotypes ($N_bn_b$) can also be Nod−. This locus may have caused the segregation ratio in the $F_2$ population to not fit single-gene, two-gene, or three-gene Mendelian ratios. Further Sanger sequencing revealed that the genotype of PI 262090 was the same as Nod+ E6 ($N_bN_b$) while the genotype of UF 487A was the same as Nod− E7 ($n_bn_b$) (Fig. 1A). Since the two original parental lines, PI 262090 and UF 487A, were both Nod+, the mutated $n_b$ allele inherited from UF 487A alone would not be sufficient to lead to the Nod− phenotype. We speculated that another mutation at a different locus must exist in PI 262090, which was then inherited by the Nod− E7 in addition to the mutation of $n_b$ inherited from UF 487A.

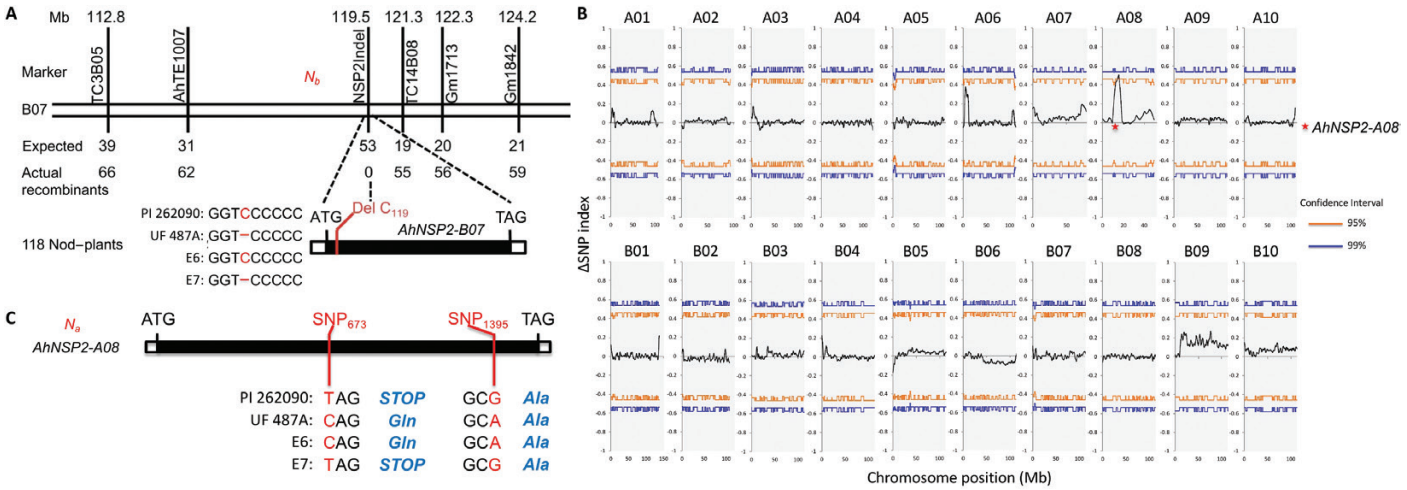



**Fig. 1.** Map-based cloning and quantitative trait locus-sequencing analysis for identification of *AhNSP2-B07* and *AhNSP2-A08*. (A) Fine mapping *AhNSP2-B07* and sequence analysis in parental lines. (B) The tricube-smoothed ΔSNP index for 20 chromosomes of peanut. *y*-Axis shows the smoothed ΔSNP index. *x*-Axis shows chromosome positions in Mb. The two-sided confidence intervals are shown in orange (95% or $P<0.05$) and blue (99%, or $P<0.01$). (C) Locations and effects of the two SNPs identified within *AhNSP2-A08*, and their genotypes in parental lines. $SNP_{673}$ leads to a premature stop codon in E7. $SNP_{1395}$ does not change the amino acid sequence in E7.

### *QTL-seq analysis and identification of a nonsense mutation in* AhNSP2-A08

Since additional polymorphic markers were hardly identified between E6 and E7, we applied the QTL-seq approach with the $F_2$ segregating population to map the other locus of natural recessive mutation. Through whole genome re-sequencing of E6, E7, and the bulked DNA samples of 20 Nod+ and 20 Nod− $F_2$ lines, respectively, an average of 212 100 670 read pairs (125 bp) per sample was obtained, which provided a 19.64× coverage of the peanut genome. After filtering, a total of 130 441 SNP loci (~6500 SNPs per chromosome) had a homozygous genotype for E6 (depth ≥4) and at least 10 reads for both Nod− and Nod+ bulks, which were subsequently analysed using QTL-seq. The smoothed ΔSNP index values were plotted for all 20 chromosomes of peanut (Fig. 1B). Under $P<0.05$, only one significant QTL was identified on chromosome A08 (from ~9 Mb to ~20 Mb). We named the gene controlling peanut nodulation in this QTL interval on A08 as $N_a$. Among a total of 562 annotated gene models located within this 11 Mb interval, only the homoeologous copy of *AhNSP2-B07* in the A subgenome (referred to as *AhNSP2-A08*) was found to be located within this region (at ~11.3 Mb of chromosome A08). It was up-regulated and among the five differentially expressed genes upon rhizobial infection (see Supplementary Table S3). Therefore, *AhNSP2-A08* as the homoeolog of *AhNSP2-B07* was considered as a priority candidate gene at the $N_b$ locus.

Through comparing the amplicon sequences of this gene between E6 and E7, two SNPs ($SNP_{673}$ and $SNP_{1395}$), both located within the coding sequence (CDS) region of *AhNSP2-A08*, were identified (Fig. 1C). $SNP_{673}$ caused a nonsense mutation, changing the codon encoding Gln into a premature stop codon (Fig. 1C). It was located in the middle of the CDS region (1542 bp in length), thus likely causing peptide length reduction from 513 to 224 amino acids. The $SNP_{1395}$ caused no change of amino acid, thus being a silent mutation (Fig. 1C). Therefore, *AhNSP2-A08* was considered as a top candidate gene ($N_a$) beside $N_b$ controlling nodulation in peanut. Thus, we randomly selected 111 $F_2$ lines, including 52 Nod− and 59 Nod+ $F_2$ lines, which were subjected to genotyping at the $SNP_{673}$ of $N_a$ and at the $Indel_{119}$ of $N_b$ using Sanger sequencing of the amplicons across the mutation sites. The results showed that all of the 52 Nod− $F_2$ lines were $n_an_a$ genotype, 27 Nod− lines were $n_bn_b$, while the other 25 Nod− plants were $N_bn_b$ genotype (see Supplementary Table S4). Out of the 59 Nod+ $F_2$ plants, for $N_a$ gene locus, 48 were either $N_aN_a$ or $N_an_a$, and 11 were $n_an_a$. Out of the 11 $n_an_a$ Nod+ lines, three were $N_bN_b$ and eight were $N_bn_b$ at the $N_b$ locus. Thus, for these 11 Nod+ lines, though $N_a$ was dysfunctional, $N_b$ could complement the function leading to Nod+ phenotype. The genotyping data further supported the hypothesis that *AhNSP2-B07* and *AhNSP2-A08* are the candidate genes controlling peanut nodulation and segregating in the $F_2$ population.

### *Genetic inheritance of* AhNSP2-A08 *and* AhNSP2-B07

The segregation patterns of these two genes were different, as a Nod− plant could be either $n_an_an_bn_b$ or $n_an_aN_bn_b$. To further confirm the segregation at these two loci, we planted the seeds of $F_2$ plants, including genotypes $n_an_aN_bn_b$ (Nod+ or Nod−), $N_an_an_bn_b$ (Nod+), $N_an_aN_bn_b$ (Nod+), and $n_an_an_bn_b$ (Nod−). As expected, the selfing populations of $n_an_an_bn_b$ plants were uniformly Nod− without segregation (see Supplementary Table S5). The segregation in the $F_2$-derived $F_3$ families of $N_an_an_bn_b$ plants followed a 3:1 (Nod+: Nod−) ratio, a typical Mendelian ratio. However, segregation in the $F_3$ families of $n_an_aN_bn_b$ plants (no matter whether Nod+ or Nod−) did not fit a 3:1 (Nod+: Nod−) ratio, thus exhibiting non-Mendelian inheritance at the $N_b$ locus. The observed segregation ratio (Nod+: Nod−) was in a range from 5:3 to 1:1 (37.5~50% Nod−). In other words, a portion (approximately 25~50%) of $n_an_aN_bn_b$ plants were Nod− instead of all being Nod+. The segregation in the selfing populations of $N_an_aN_bn_b$ plants did not fit a 15:1 ratio. Instead, the observed ratio (Nod+: Nod−) in those populations was in a range from 28:4 to 29:3, which was consistent with independent segregation at these two loci with one following a Mendelian ratio of 3:1 and the other in a range of 5:3 to 1:1 (Supplementary Fig. S3; Supplementary Tables S2, S5). The above results confirmed that *AhNSP2-A08* followed Mendelian inheritance, while *AhNSP2-B07* followed non-Mendelian inheritance.

### *Functional validation of* AhNSP2-B07 *and* AhNSP2-A08

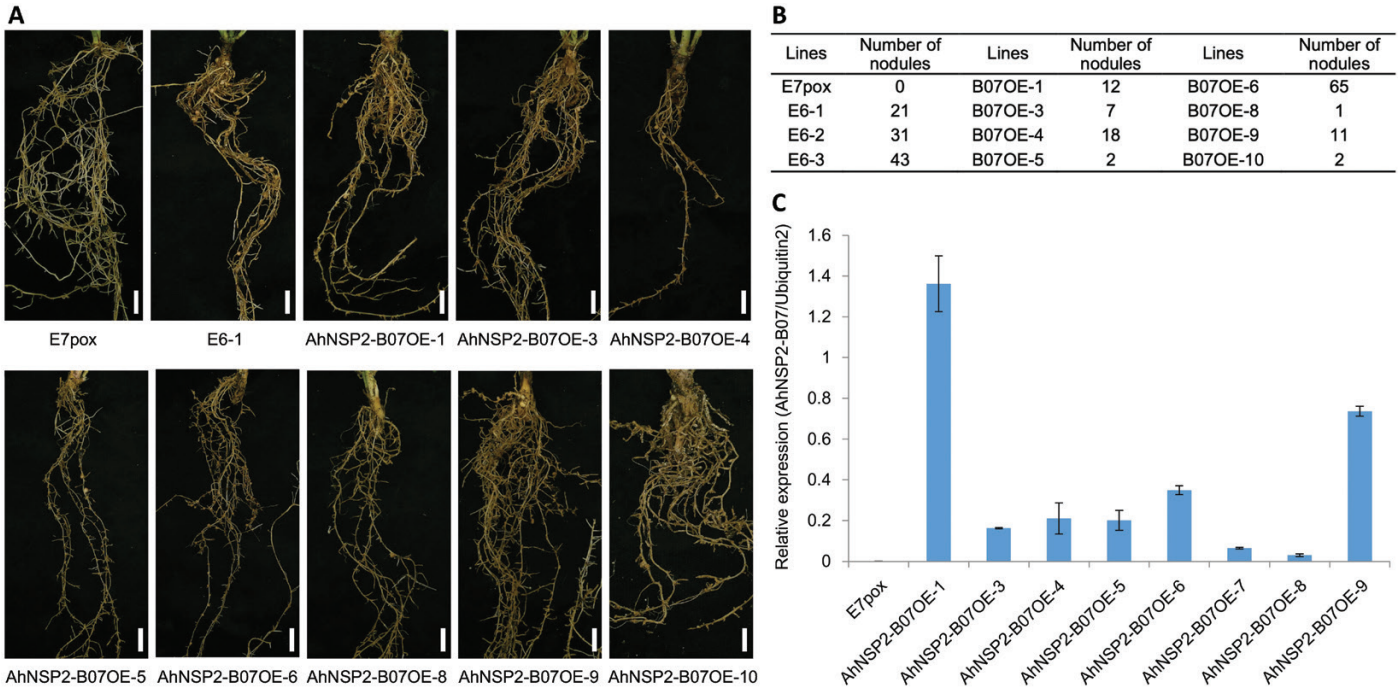
Considering that a portion of the Nod− $F_2$ lines were heterozygous at the AhNSP2-B07 locus, transforming a single copy of wild type allele of *AhNSP2-B07* into E7 may not restore nodulation in E7. Therefore, to validate the function of *AhNSP2-B07* in peanut nodulation, we made an overexpressing construct of the wild type *AhNSP2-B07* and transformed it into E7 through a hairy root transformation system. A total of 23 putative transgenic plants (with hairy roots) named as 'AhNSP2-B07OE' lines were obtained. After inoculation with a *Bradyrhizobium* strain, Lb8, a total of eight AhNSP2-B07OE lines produced various numbers of nodules, and all the E7 lines transformed with empty vector produced no nodules (Fig. 2A, B). Further, the PCR for amplification of the hygromycin resistance marker gene using hairy roots DNA as templates showed that all these eight plants were transgenic positive (see Supplementary Fig. S4A). Sanger sequencing confirmed that these eight lines were homozygous recessive ($n_bn_b$) at the AhNSP2-B07 $Indel_{119}$ locus based on leaf DNA samples and were heterozygous ($N_bn_b$) at the locus based on hairy root DNA samples. Compared with the expression of *AhNSP2-B07* in E7 plants transformed with the empty vector, these eight AhNSP2-B07OE lines showed increased expression of *AhNSP2-B07* with different levels based on qRT-PCR

(Fig. 2C). The overexpression experiment confirmed that the loss of function of *AhNSP2-B07* was one of the causes for E7 losing nodulation function.

To validate the function of *AhNSP2-A08* in peanut nodulation, a complementation experiment was performed. Since all of the Nod− $F_2$ plants were homozygous recessive ($n_an_a$), transforming a single copy of wild type allele of *AhNSP2-A08* into E7 should restore its nodulation phenotype. After transforming the complementary construct carrying the wild type allele of *AhNSP2-A08* into E7, we obtained a total of 20 putative transgenic plants whose hairy roots were detected with GFP fluorescence, and they were named as 'AhNSP2-A08C' lines. After inoculation with Lb8, 10 lines produced nodules (five lines shown in Fig. 3A, B). All lines transformed with empty vector produced no nodules. Further, the PCR using a primer pair covering the hygromycin resistance marker gene and GFP gene in the vector showed that all these 10 AhNSP2-A08C lines' hairy roots were transgenic positive (see Supplementary Fig. S4B). Sanger sequencing confirmed that all these 10 lines were $n_an_a$ genotype for leaf DNA while heterozygous $N_an_a$ genotype according to hairy root DNA samples. qRT-PCR was further performed using these hairy root samples. Compared to the expression in E7 plants transformed with the empty vector, all the five tested Nod+ AhNSP2-A08C lines showed increased expression of *AhNSP2-A08* with different levels after inoculation with Lb8 (Fig. 3C). Above results led to the conclusion that the loss of function of *AhNSP2-A08* is another cause of E7 losing the nodulation function. Thus, *AhNSP2-A08* also controls nodulation in peanut. However, we did not observe any significant correlation between gene expression levels of *AhNSP2-B07*/*AhNSP2-A08* in transgenic hairy roots with the nodule numbers. Specifically, some samples had higher expression of *AhNSP2-B07*/*AhNSP2-A08* with relatively smaller nodule numbers and vice versa (Figs 2B, C and 3B, C).

## *Both* AhNSP2-B07 *and* AhNSP2-A08 *encode a GRAS transcription factor*

Both *AhNSP2-B07* and *AhNSP2-A08* encode GRAS transcription factors with 516 and 513 amino acids (96% identity), respectively (Fig. 4A, B). These two genes are homoeologous, and located in two different sub-genomes, B and A, respectively (see Supplementary Fig. S5). The different chromosome number, 07 of B sub-genome and 08 of A sub-genome, was caused by the complex rearrangements between A07 and A08 (Bertioli *et al.*, 2016). By searching the two ancestral diploid genomes of peanut, two more genes orthologous to *AhNSP2* (~60% amino acid identity) were found in peanut genomes, with one located on A08 and the other on B07, which were ~1.6–1.7 Mb away from *AhNSP2-A08* and *AhNSP2-B07*, respectively. Thus, they were named as *AhNSP2-like-A08* and *AhNSP2-like-B07*, respectively. A phylogenetic tree containing orthologs of *NSP2* in several legume species and Arabidopsis




| Lines | Number of nodules | Lines | Number of nodules | Lines | Number of nodules |
|---|---|---|---|---|---|
| E7pox | 0 | B07OE-1 | 12 | B07OE-6 | 65 |
| E6-1 | 21 | B07OE-3 | 7 | B07OE-8 | 1 |
| E6-2 | 31 | B07OE-4 | 18 | B07OE-9 | 11 |
| E6-3 | 43 | B07OE-5 | 2 | B07OE-10 | 2 |

**Fig. 2.** Overexpression of *AhNSP2-B07* in E7. (A) Nodulation phenotype of transgenic hairy roots after overexpression of *AhNSP2-B07* in E7. Scale bar: 1 cm. (B) Summary of nodule numbers after overexpression of *AhNSP2-B07* in E7. (C) Gene expression of *AhNSP2-B07* in transgenic hairy roots of E7 AhNSP2-B07OE plants. The error bars represent standard error of three replicates.


B

| Lines | Number of nodules | Lines | Number of nodules |
|---|---|---|---|
| E7-empty | 0 | AhNSP2-A08C-3 | 46 |
| AhNSP2-A08C-1 | 3 | AhNSP2-A08C-4 | 4 |
| AhNSP2-A08C-2 | 31 | AhNSP2-A08C-5 | 5 |

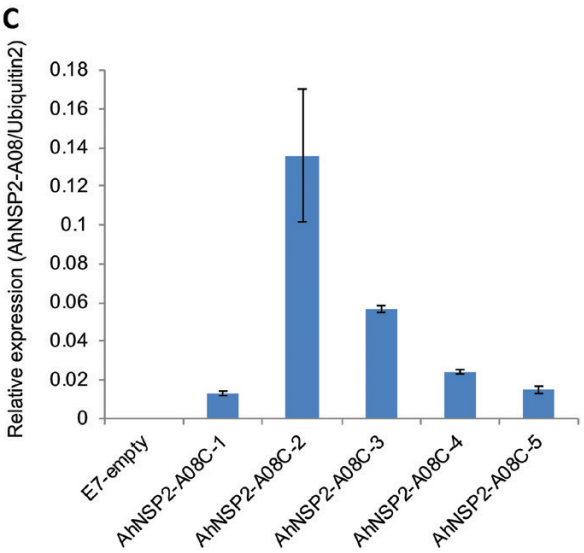



**Fig. 3.** Complementation of *AhNSP2-A08* in E7. (A) Nodulation phenotype of transgenic hairy roots after complementation of *AhNSP2-A08* in E7. The two small figures below correspond to regions labelled with a red dashed circle, which are zoomed in under the microscope with either normal light (left) or GFP light (right). All scale bars: 1 cm. (B) Summary of nodule numbers after complementation of *AhNSP2-A08* in E7. (C) Gene expression of *AhNSP2-A08* in transgenic hairy roots of E7 AhNSP2-A08C plants. The error bars represent standard error of three replicates.

was constructed based on their amino acid sequences (Fig. 4C). As expected, the *AhNSP2* genes clustered with *NSP2* genes from other legumes, while the *AhNSP2-like* genes also clustered with *NSP2-like* genes from other legumes (Fig. 4C). By calculating the non-synonymous substitutions ($K_a$) and synonymous substitutions ($K_s$) of the *AhNSP2*, *AhNSP2-like*, and *AhNSP1* genes in comparison with 3458 genome-wide homoeologous gene pairs, we found that *AhNSP2* had accumulated non-synonymous (percentile 73.2%) and synonymous (percentile 89.6%) mutations at a higher rate than most of the remaining genes in the genome (Supplementary Table S6). Therefore, it is likely that *AhNSP2* may have been diverging and evolved faster than other genes in the peanut genomes.

### *Subcellular localization and interaction with AhNSP1*

To investigate the subcellular localization of the two *AhNSP2* genes' products, each of them was fused to a CaMV35S promoter and GFP and was transformed into a rice protoplast. The transient gene expression of the construct showed that both *AhNSP2* genes' products were functioning in the nucleus (Fig. 5A), which was consistent with their function as a transcription factor (Kaló *et al.*, 2005; Murakami *et al.*, 2006). As the NSP1–NSP2 complex is required for nodulation signaling (Eckardt, 2009), we evaluated the interaction between AhNSP1 and AhNSP2 proteins using the yeast-two hybrid system. The two AhNSP2 proteins were equally capable of interacting with either of the two AhNSP1 proteins encoded from the two sub-genomes and interacting between themselves (Fig. 5C). The interactions between AhNSP1 and AhNSP2 proteins, as well as between the two AhNSP2 proteins were further validated with a bimolecular fluorescence complementation (BiFC) assay (see Supplementary Fig. S6A, B).

### *Tissue specific expression and phytohormone induction*

Further, qRT-PCR was utilized to investigate the tissue specific expression of the two *AhNSP2* genes. *AhNSP2-A08* was expressed mainly in growing root tissues and had very low expression in stem, leaf, or flower, similar to *LjNSP2* and *MtNSP2* (Fig. 5B; see Supplementary Fig. S7A, B). In contrast, *AhNSP2-B07* had relatively high expression in both growing roots and flowers (Fig. 5B). Its high expression in flowers was similar to its solo ortholog in Arabidopsis (Supplementary Fig. S7C). This suggested that the high expression of *AhNSP2-B07* in reproductive organs may influence its functional partition or genetic inheritance in nodulation.

We further investigated the responses of expression of the *AhNSP2* genes upon treatment with gibberellin ($GA_3$), cytokinin (6-BA), or auxin (IAA) within a 1 d window. The results

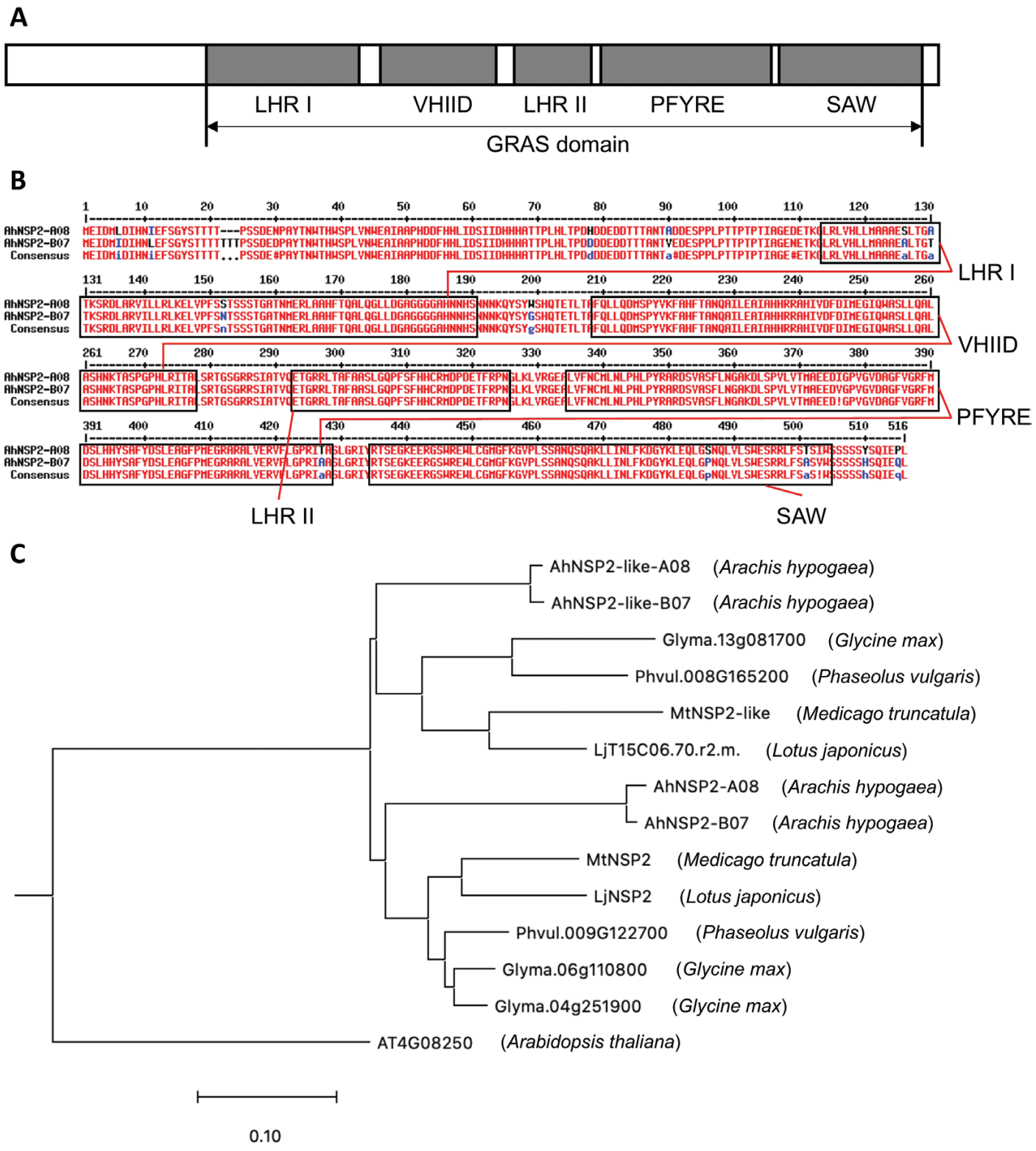



**Fig. 4.** Conserved domain and phylogenetic analysis of *AhNSP2-A08* and *AhNSP2-B07*. (A) Schematic representation of the structure and motifs of AhNSP2 GRAS protein. (B) Conserved motifs labeled in the sequences of the two AhNSP2 proteins. Both AhNSP2 proteins (96% identity) contain GRAS conserved domain, including the leucine heptad repeat I (LHR I), VHIID, LHR II, PFYRE, and SAW subdomains. (C) Phylogenetic analysis for *AhNSP2* from *Arachis hypogaea* and the orthologs from *Glycine max*, *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula*, and *Arabidopsis thaliana*. The scale bar represents genetic distance or substitutions per site.

showed that the expression of both *AhNSP2* genes was regulated by these three hormones, and the two genes exhibited a similar expression trend after treatment (see Supplementary Fig. S8A, B). After treatment with $GA_3$, the expression of both *AhNSP2* genes gradually increased (highest at 3 h), then gradually decreased after 3 h. Similarly, the expression of both *AhNSP2* genes gradually increased over time after treatment with 6-BA, reaching the highest levels at 3 h. However, the expression of *AhNSP2-A08* showed no obvious change after 3 h, while the expression of *AhNSP2-B07* slightly decreased after 3 h. Within the first 12 h after treatment with IAA, the expression of both *AhNSP2* genes gradually increased. At 24 h, the expression of *AhNSP2-A08* slightly decreased, while the expression of *AhNSP2-B07* showed no obvious change compared with 12 h. These results suggested that gibberellin, cytokinin, and auxin all played a role in regulating the expression of *AhNSP2*s in peanut nodulation.

### *Temporal expression patterns of the two* AhNSP2 *genes and other related nodulation genes in peanut infected with Lb8*

To determine the temporal expression patterns of both *AhNSP2* genes during nodule formation, we inoculated the roots of E6 and E7 at 6 d after germination by using the *Bradyrhizobium* strain Lb8, and collected the 5–6 cm-long inoculated root tissues at 0, 0.5, 1, 2, 3, 5, 7, and 9 DAI. At 7 DAI, nodules became visible to the naked eye at the position where rosette root hairs form. Through qRT-PCR, we observed that in Nod− E7, both genes had little expression, with or without rhizobial infection. In E6, both genes responded early to rhizobial inoculation (as early as 12 h), indicating their role in rhizobial infection, and showed the highest expression levels at 3–5 DAI when nodules were initiating (see Supplementary Fig. S9A, B). Sanger sequencing coupled with resequencing did not reveal any sequence variations between E6 and E7 in the promoter and 1 kb sequence downstream of the stop codon of *AhNSP2-B07* and *AhNSP2-A08*. These results indicated that expression of the two genes was likely autoregulated. Alternatively, the differential gene expression between E6 and E7 could be caused by upstream gene mutations or degradation of *AhNSP2* transcripts.

We further investigated the temporal expression patterns of peanut orthologs of four genes involved up- and downstream of the *NSP2* gene based on model legume species during the processes (Oldroyd *et al.*, 2011), including *Cytokinin Response1* (*CRE1*), *Response Regulator4 RR4* (Gonzalez-Rizzo *et al.*, 2006), both upstream of *NSP2* during nodule organogenesis, *NIN* (Schauser *et al.*, 1999) and *HAP2.1* (Combier *et al.*, 2006), downstream of *NSP2* during bacterial infection (same regulation level during nodule organogenesis). qRT-PCR primers for these genes were designed to target both copies in sub-genomes A and B of peanut since they were highly similar in sequence. Results showed that *NIN* was not expressed at early stages of rhizobial infection and gradually expressed more

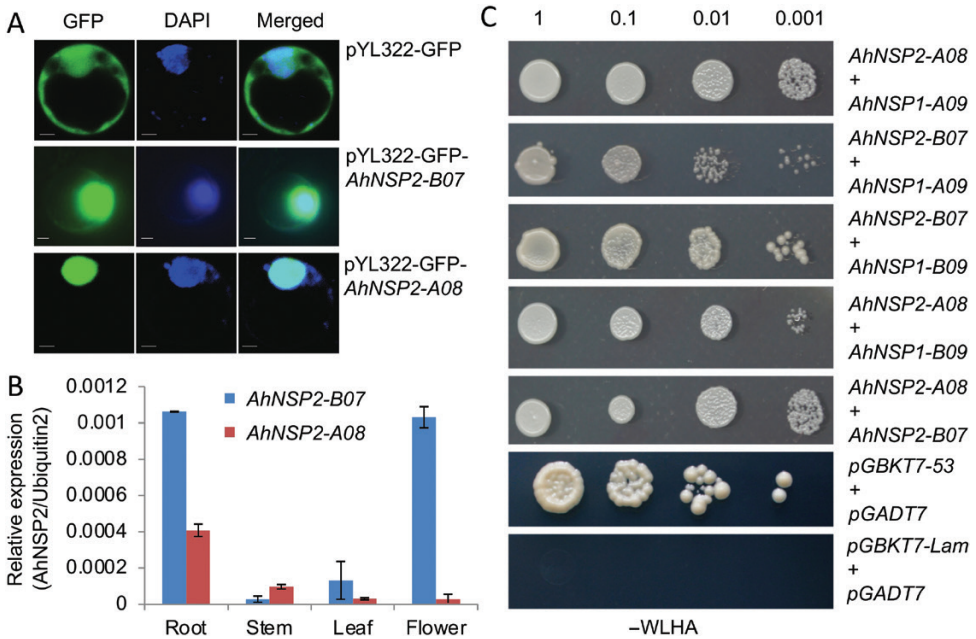



**Fig. 5.** Subcellular localization and tissue specific expression of *AhNSP2s* and their interactions with *AhNSP1s*. (A) Subcellular localization of two *AhNSP2s*. 'pYL322-GFP' indicates the vector carrying CaMV35S promoter and GFP, which was fused to AhNSP2-A08 and AhNSP2-B07. The scale bar indicates 2 μm. DAPI, 4′,6-diamidino-2-phenylindole. (B) Expression of the two *AhNSP2* in root, stem, leaf, and flower. The error bars represent standard error of three replicates. (C) Yeast two-hybrid assays between AhNSP2-A08 (BD) and AhNSP1-A09 (AD); AhNSP2-B07 (BD) and AhNSP1-A09 (AD); AhNSP2-B07 (BD) and AhNSP1-B09 (AD); AhNSP2-A08 (BD) and AhNSP1-B09 (AD); and AhNSP2-A08 (AD) and AhNSP2-B07 (BD). '–WLHA' refers to synthetic dropout medium lacking Trp, Leu, His, and Ade.

highly after inoculation with Lb8 in E6 (see Supplementary Fig. S9C). However, its expression was extremely low and there was no significant change for the other three conditions, implying that *NIN* could be regulated by *NSP2*.The expression of *CRE1* and *RR4* seemed not to be influenced by rhizobial infection in both Nod+ E6 and Nod− E7 (Supplementary Fig. S9D, E), indicating these two genes functioned upstream of *NSP2*. However, in model legumes, the expression of *RR4* was strongly up-regulated at early stages of nodule organogenesis (4 DAI) in *M. truncatula* (Gonzalez-Rizzo *et al.*, 2006).The expression of CCAAT binding factor (*HAP2.1*) was high at later infection stages in both E6 and E7, implying it may not be regulated by *NSP2* in peanut. Instead, it might be related to the root development (Supplementary Fig. S9F). The results suggested that the cytokinin signaling involving *RR4* and other gene regulation during nodule organogenesis in model legumes may involve different mechanisms in cultivated peanut.

### AhNSP2 *mutations in the US peanut mini core collection*

To evaluate the prevalence of the mutated alleles of *AhNSP2-B07* and *AhNSP2-A08* in natural populations of peanut, the US peanut mini core collection (Holbrook and Dong, 2005), representative of the genetic diversities of peanut germplasms, was genotyped at the two loci, $SNP_{673}$ of *AhNSP2-A08* and $Indel_{119}$ of *AhNSP2-B07*. Out of 107 peanut mini core accessions, which are all Nod+, 14 accessions were $N_aN_an_bn_b$ genotype carrying the *AhNSP2-B07* $Indel_{119}$ deletion and only four accessions were $n_an_aN_bN_b$ genotype carrying the *AhNSP2-A08* $SNP_{673}$ mutation (see Supplementary Table S7). This result showed that approximately 13% of the peanut accessions had a loss of function of *AhNSP2-B07*, while a much lower proportion (~4%) had a loss of function of *AhNSP2-A08* naturally, and none of the accessions contained both mutations. By plotting the countries of origin of the mini core accessions, it was shown that South America, the origin of cultivated peanut, and Africa contained accessions carrying either of the two mutations (Supplementary Fig. S10A). For accessions with an origin in Asian countries, only the *AhNSP2-B07* Indel mutation was observed.According to the population structure of mini core germplasm collection, we observed that the accessions carrying the $n_a$ allele was only present in group 2 while the accessions carrying the $n_b$ allele were distributed in all four different groups of the mini core collection with a much wider genetic background (Supplementary Fig. S10B). The wider distribution of $n_b$ than $n_a$ allele suggested that $n_b$ allele was either associated with certain peanut adaptive traits or more likely was mutated or evolved earlier than $n_a$ allele.

## Discussion

In this study, we have identified and cloned two homoeologous nodulation genes, *AhNSP2-A08* and *AhNSP2-B07*, located on two sub-genomes of cultivated peanut, an allotetraploid crop species with the crack-entry infection path. This pair of homoeologous genes exhibited Mendelian and non-Mendelian inheritance and different mutation rates in natural peanut populations. Our study firstly explored the genetic basis of non-nodulating plants derived from recombination of homoeologous alleles in crosses of cultivated peanuts and revealed the diverged evolutionary paths of a pair of homoeologous genes in allotetraploid.

The identification of symbiosis genes through a natural variation approach is often a challenging but elegant approach, which led to the discovery of some other symbiosis genes in legumes, such as SYM2/LYK3 (Geurts *et al.*, 1997; Limpens *et al.*, 2003) and some NCRs (NCR169, Horváth *et al.*, 2015; NCR211, Kim *et al.*, 2015). However, in peanut, a forward genetics approach has not been widely used for gene identification largely due to its features of polyploidy and very closely related sub-genomes. Using a Nod− RIL E7, we were able to apply a forward genetics approach to construct a mapping population by crossing E7 with its Nod+ sister RIL E6. The QTL-seq approach was powerful and allowed us to successfully identify *AhNSP2-A08*, but not to identify the signal at the region harboring *AhNSP2-B07* in the same experiment (Fig. 1B). This is because the $F_2$ lines of heterozygous genotype at this locus ($n_an_aN_bn_b$) could be either Nod+ or Nod−, which would remarkably reduce the power of identifying the signals that are significantly different between the Nod+ and Nod− bulks in QTL-seq analysis. Interestingly, $N_b$ was originally assumed to follow Mendelian inheritance during the mapping process. $n_b$ was treated as a recessive allele, and Nod− $N_bn_b$ plants were considered as recombinants.The region containing $n_b$ was identified based on the theory that markers closer to $n_b$ will have smaller recombination rates. However, it actually follows non-Mendelian inheritance, because when *AhNSP2-A08* was dysfunctional ($n_an_a$), some $n_an_aN_bn_b$ genotypes produced nodules and others did not. Among the Nod− $F_2$ individuals, there were slightly more $n_an_an_bn_b$ plants (about 4/7) than $n_an_aN_bn_b$ plants (about 3/7). Moreover, for the $n_an_aN_bn_b$ Nod− plants, recombined markers could have an E7 (Nod−) genotype. Thus, *AhNSP2-B07* was coincidentally mapped to the QTL region.

One possible explanation for the non-Mendelian inheritance pattern of $N_b$ is a parental effect, which was pointed out in a previous report on the inheritance of non-nodulation in peanut (Gallo-Meagher *et al.*, 2001). In that study, PI 262090, UF 487A, and a Nod− line, M4-2, derived from the cross between PI 262090 and UF 487A were used to make crosses. An $n_an_aN_bn_b$ plant could be Nod+ if it was derived from a combination of $n_aN_b$ male gametes and $n_an_b$ female gametes, and it could be Nod− if it was derived from the combination of $n_an_b$ male gametes and $n_aN_b$ female gametes (see Supplementary Table S2). The Mendelian inheritance of $N_a$ and non-Mendelian inheritance of $N_b$ could be confirmed in the $F_2$ populations derived from the crosses between UF 487A and M4-2, and between PI

262090 and M4-2, in which approximately 37.5~50% of the progeny of $n_a n_a N_b n_b$ plants were Nod−, in contrast to the 25% Nod− for Mendelian inheritance. Considering the high identity (96%) of these two protein sequences, the difference in inheritance of the two *AhNSP2* genes is likely to be unrelated to the small differences in protein sequence. However, the tissue-specific expression patterns of $N_a$ and $N_b$ are different, since $N_a$ is highly expressed only in roots while $N_b$ is highly expressed in both roots and flowers (Fig. 5B). The different tissue expression patterns may have influenced their genetic inheritance. It is likely that the high expression of *NSP2* during the reproduction process is undesired and needs to be properly repressed. We propose that the expression of the $N_b$ allele is inhibited by a certain mechanism in the $n_a N_b$ female gamete and in a disadvantageous situation. The inhibition may be caused by an inhibitor gene inhibiting *AhNSP2-B07* but not *AhNSP2-A08* in female gametes. Alternatively, the inhibition may be influenced by parental or gametic imprinting, which can lead to non-Mendelian inheritance of parental alleles (Bai and Settles, 2015). Future studies are needed to validate these hypotheses. Therefore, the two *NSP2* genes in allotetraploid peanut are under different and complex regulation, which is novel compared with the single copy found in diploid model legumes.

Although both control peanut nodulation, the two *AhNSP2* genes on the two sub-genomes separately showed different expression in different tissues and during the nodulation process. In allopolyploids, the retention of homoeologs can be biased towards one progenitor copy, as reported in cotton (Chaudhary *et al.*, 2009; Flagel and Wendel, 2009; Rapp *et al.*, 2009) and *Tragopogon* (Buggs *et al.*, 2010), while the other one can gradually lose its function or diverged to a copy having a different function. As revealed from this study, a much higher percentage (~13%) of the US peanut mini core accessions carry the *AhNSP2-B07* deletion ($N_a N_a n_b n_b$) than the accessions carrying *AhNSP2-A08* SNP mutation (($n_a n_a N_b N_b$; ~4%). This frequency difference may not be a coincidence. Unlike *AhNSP2-A08*, *AhNSP2-B07* is also expressed at relatively high levels in flowers. The partition of gene expression may lead to gene retention during evolution (Adams, 2007). Although these two homoeologous genes play a redundant role in controlling peanut nodulation, *AhNSP2-B07* may play a less essential role than *AhNSP2-A08* during peanut nodulation due to the parental effect. This evidence suggests that natural selection shaped a dysfunctionalization or subfunctionalization of *AhNSP2-B07* to avoid redundancy and excessive expression of NSP2 protein and contributed to fitness during evolution after alloploidization (Chang *et al.*, 2010).

In addition to the accessions mentioned in this study, a few other accessions could also lead to the Nod− phenotype, including PI 259747, NC 17, NC Ac 2731 (Nigam *et al.*, 1980), Shantung KU No 203 (Branch *et al.*, 1984), PI 109839, and PI 405132 (Dutta and Reddy, 1988). Since these accessions were not included in the mini core collection, it would be interesting to know whether they have the $n_a$ or $n_b$ alleles. As a suite of pathway genes determine nodulation, it is also important to know whether this evolution specifically happened to *AhNSP2* or also to other genes involved in nodulation. In addition, the polymorphisms identified in the current study may not cover all the cases of non-nodulation in peanut, and other possible mutations leading to the Nod− phenotype and/or other deleterious *AhNSP2* alleles may exist. Sequencing the whole CDS of those nodulation-related genes in the mini core collection would provide more insight into the above questions.

The ortholog of *NIN* in peanut was not expressed at early stages of rhizobial infection, which is different from model legumes (Schauser *et al.*, 1999; Murakami *et al.*, 2006). In model legumes, *NSP2* plays a role in both rhizobial infection and nodule organogenesis, and is up-regulated in early stages upon rhizobial infection (Kaló *et al.*, 2005). In peanut, *AhNSP2-B07* could be induced as early as 0.5 DAI, in accordance with the early regulation in model legumes. However, a recent transcriptome study found that the two *AhNSP2* genes were only induced at late stages (21 DAI) in mature nodules. This discrepancy could be caused by the low expression of *AhNSP2* genes at early stages, involving a relatively subtle change compared with other highly expressed genes, which could be neglected. Besides, in the model legume *L. japonicus*, the expression of *NSP2* in young root regions could be down-regulated in response to rhizobial infection (Murakami *et al.*, 2013), which further contributed to the complexity of comparing gene expression levels if the whole root samples with different infection and root development stages are collected. The mutation of *NSP2* may have influenced the early induction of *NIN* in peanut. The later induction of *NIN* than *NSP2* also may demonstrate that it is downstream of *NSP2* in peanut nodulation, in agreement with model legumes (Oldroyd, 2013). The different expression patterns of *RR4* and *HAP2.1* in peanut from those in model legumes indicated a different regulation mechanism during cytokinin signaling and early transcriptional regulations of nodule organogenesis in peanut. Similar results on the differential cytokinin signaling between peanut and model legumes were reported by a recent systematic transcriptome study in peanut (Karmakar *et al.*, 2019).

Our study identified and validated the functions of a pair of homoeologous GRAS transcription factors controlling nodulation in allotetraploid peanut. We provided an example of the symbiosis gene in two sub-genomes of this allopolyploid species exhibiting Mendelian and non-Mendelian inheritance. It appears that *AhNSP2-B07*, which is less competitive in nodulation capacity than *AhNSP2-A08*, could be selected against in natural populations. The nodulation efficiency and yield of different genotypes, including $N_a N_a n_b n_b$, $n_a n_a N_b N_b$, and $N_a N_a N_b N_b$, can be evaluated in the future to select the genotype with highest performance. Our findings facilitate identifying

and exploring homoeologous genes associated with symbiosis in allopolyploid legume species. How to manipulate these homoeologous symbiosis genes to achieve a gene set leading to optimal yield is also an interesting and important topic to be investigated further.

## Supplementary data

The following supplementary data are available at *JXB* online.

Fig. S1. Morphological and yield-related traits of PI 262090, UF 487A, E6, and E7.

Fig. S2. A quantitative trait locus controlling nodulation mapped on chromosome B07 using genotypes and phenotypes of 90 $F_2$ individuals.

Fig. S3. Genetic inheritance model in the $F_2$ population derived from the cross between E6 and E7.

Fig. S4. PCR verification of transgenic positive hairy roots.

Fig. S5. Schematic showing the homologous regions between chromosome A08 and B07.

Fig. S6. Interactions between AhNSP2 and AhNSP1 and between two AhNSP2s through bimolecular fluorescence complementation assay.

Fig. S7. Expression of *NSP2* orthologs in different tissues in *Lotus japonicus*, *Medicago truncatula*, and *Arabidopsis thaliana*.

Fig. S8. Expression of AhNSP2 genes under treatment with gibberellin ($GA_3$), cytokinin (6-BA), or auxin (IAA).

Fig. S9. Expression patterns of the two *AhNSP2s*, orthologs of *NIN*, *CRE1*, *RR4*, and *HAP2.1* in E6 and E7 inoculated with water or Lb8 at eight different time points.

Fig. S10. World map showing countries of origin for US peanut mini core accessions and their genetic backgrounds.

Table S1. Primers used for mapping, Sanger sequencing, cloning, gene expression analysis, yeast two-hybrid assay and bimolecular fluorescence complementation assay.

Table S2. Segregation of nodulation in $F_1$ and $F_2$ populations from different crosses.

Table S3. Candidate genes located within identified QTL regions.

Table S4. Summary of non-nodulating and nodulating $F_2$ plants genotyped at the two *AhNSP2* loci.

Table S5. Segregation of nodulation in $F_3$ families derived from the cross between E6 and E7.

Table S6. Non-synonymous substitutions ($K_a$) and synonymous substitutions ($K_s$) of *AhNSP2*, *AhNSP1*, and *AhNSP2-like* genes in comparison with 3458 genome-wide gene pairs.

Table S7. Genotypes of the two *AhNSP2* mutations in the US peanut mini core collection.

## Acknowledgements

This study was funded by the Major Program of Guangdong Basic and Applied Research (2019B030302006), Florida Peanut Producers Association, the University of Florida IFAS Early Career Scientists Award, and Guangdong Natural Science Funds for Distinguished Young Scholars (Grant No. 2017A030306001). We thank Dr Zhonglin Mou for providing modified pCAMBIA1300 vector. The authors declare no conflict of interest.

## Author Contributions

JW conceived the research; JW and HZ designed the experiments; ZP, HZ, LT, HC, HS, AC, RKV, ZKZ, ZFZ, ZL, LW, and JM performed the experiments and analysed the data; MG and YL helped in preparing plant materials; ZP and HZ prepared the manuscript draft; JW and HZ critically revised the manuscript; all authors approved the final version of the manuscript.

## Data availability

All data supporting the findings of the study are available in the paper and within its supplementary data published online.

## References

**Adams KL.** 2007. Evolution of duplicate gene expression in polyploid and hybrid plants. The Journal of Heredity **98**, 136–141.

**Adams KL, Cronn R, Percifield R, Wendel JF.** 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proceedings of the National Academy of Sciences, USA **100**, 4649–4654.

**Arrighi JF, Barre A, Ben Amor B, *et al.*** 2006. The *Medicago truncatula* lysin [corrected] motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. Plant Physiology **142**, 265–279.

**Bai F, Settles AM.** 2015. Imprinting in plants as a mechanism to generate seed phenotypic diversity. Frontiers in Plant Science **5**, 780.

**Bertioli DJ, Cannon SB, Froenicke L, *et al.*** 2016. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nature Genetics **48**, 438–446.

**Bertioli DJ, Jenkins J, Clevenger J, *et al.*** 2019. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. Nature Genetics **51**, 877–884.

**Bertioli DJ, Moretzsohn MC, Madsen LH, *et al.*** 2009. An analysis of synteny of *Arachis* with *Lotus* and *Medicago* sheds new light on the structure, stability and evolution of legume genomes. BMC Genomics **10**, 45.

**Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**, 2114–2120.

**Bottani S, Zabet NR, Wendel JF, Veitia RA.** 2018. Gene expression dominance in allopolyploids: hypotheses and models. Trends in Plant Science **23**, 393–402.

**Branch WD, Hammons RO, Kvien CS.** 1984. Nonnodulating allelism test in *Arachis hypogaea* (L.). Crop Science **39**, 487–488.

**Buggs RJ, Elliott NM, Zhang L, Koh J, Viccini LF, Soltis DE, Soltis PS.** 2010. Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. New Phytologist **186**, 175–183.

**Cannon SB, Sterck L, Rombauts S, *et al.*** 2006. Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. Proceedings of the National Academy of Sciences, USA **103**, 14959–14964.

**Chang PL, Dilkes BP, McMahon M, Comai L, Nuzhdin SV.** 2010. Homoeolog-specific retention and use in allotetraploid *Arabidopsis suecica* depends on parent of origin and network partners. Genome Biology **11**, R125.

**Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF.** 2009. Reciprocal silencing, transcriptional bias and functional

divergence of homeologs in polyploid cotton (*Gossypium*). Genetics **182**, 503–517.

**Chen H, Zhang Z, Ni E, *et al.*** 2020. HMS1 interacts with HMS1I to regulate very-long-chain fatty acid biosynthesis and the humidity-sensitive genic male sterility in rice (*Oryza sativa*). New Phytologist **225**, 2077–2093.

**Chen X, Li H, Pandey MK, *et al.*** 2016. Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. Proceedings of the National Academy of Sciences, USA **113**, 6785–6790.

**Combier JP, Frugier F, de Billy F, *et al.*** 2006. MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. Genes & Development **20**, 3084–3088.

**Das DR, Horváth B, Kundu A, Kaló P, DasGupta M.** 2019. Functional conservation of CYCLOPS in crack entry legume *Arachis hypogaea*. Plant Science **281**, 232–241.

**Dutta M, Reddy LJ.** 1988. Further studies on genetics of nonnodulation in peanut. Crop Science **28**, 60–62.

**Eckardt NA.** 2009. Nodulation signaling in legumes depends on an NSP1-NSP2 complex. The Plant Cell **21**, 367.

**Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, Shanmugam D, Roos DS, Stoeckert CJ Jr**. 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Current Protocols in Bioinformatics **Chapter 6**, Unit 6.12.1–Unit 6.1219.

**Flagel LE, Wendel JF.** 2009. Gene duplication and evolutionary novelty in plants. New Phytologist **183**, 557–564.

**Forrester NJ, Ashman TL.** 2018. The direct effects of plant polyploidy on the legume-rhizobia mutualism. Annals of Botany **121**, 209–220.

**Gallo-Meagher M, Dashiell KE, Gorbet DW.** 2001. Parental effects in the inheritance of nonnodulation in peanut. The Journal of Heredity **92**, 86–89.

**Geurts R, Heidstra R, Hadri AE, Downie JA, Franssen H, Van Kammen A, Bisseling T.** 1997. Sym2 of pea is involved in a nodulation factor-perception mechanism that controls the infection process in the epidermis. Plant Physiology **115**, 351–359.

**Gonzalez-Rizzo S, Crespi M, Frugier F.** 2006. The *Medicago truncatula* CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with *Sinorhizobium meliloti*. The Plant Cell **18**, 2680–2693.

**Gorbet DW, Burton JC.** 1979. A non-nodulating peanut. Agronomy Abstracts **19**, 727–728.

**Herridge DF, Peoples MB, Boddey RM.** 2008. Global inputs of biological nitrogen fixation in agricultural systems. Plant Soil **311**, 1–18.

**Holbrook CC, Dong WB.** 2005. Development and evaluation of a mini core collection for the US peanut germplasm collection. Crop Science **45**, 1540–1544.

**Horváth B, Domonkos Á, Kereszt A, *et al.*** 2015. Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the *Medicago truncatula dnf7* mutant. Proceedings of the National Academy of Sciences, USA **112**, 15232–15237.

**Ibáñez F, Wall L, Fabra A.** 2017. Starting points in plant-bacteria nitrogen-fixing symbioses: intercellular invasion of the roots. Journal of Experimental Botany **68**, 1905–1918.

**Kaló P, Gleason C, Edwards A, *et al.*** 2005. Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. Science **308**, 1786–1789.

**Kang YJ, Kim SK, Kim MY, *et al.*** 2014. Genome sequence of mungbean and insights into evolution within *Vigna* species. Nature Communications **5**, 5443.

**Karmakar K, Kundu A, Rizvi AZ, Dubois E, Severac D, Czernic P, Cartieaux F, DasGupta M.** 2019. Transcriptomic analysis with the progress of symbiosis in 'Crack-Entry' legume *Arachis hypogaea* highlights its contrast with 'Infection Thread' adapted legumes. Molecular Plant-Microbe Interactions **32**, 271–285.

**Katoh K, Misawa K, Kuma K, Miyata T.** 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research **30**, 3059–3066.

**Kereszt A, Li D, Indrasumunar A, Nguyen CD, Nontachaiyapoom S, Kinkema M, Gresshoff PM.** 2007. *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. Nature Protocols **2**, 948–952.

**Kim M, Chen Y, Xi J, Waters C, Chen R, Wang D.** 2015. An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. Proceedings of the National Academy of Sciences, USA **112**, 15238–15243.

**Kundu A, DasGupta M.** 2018. Silencing of putative cytokinin receptor histidine kinase1 inhibits both inception and differentiation of root nodules in *Arachis hypogaea*. Molecular Plant-Microbe Interactions **31**, 187–199.

**Levy J, Bres C, Geurts R, *et al.*** 2004. A putative $Ca^{2+}$ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. Science **303**, 1361–1364.

**Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**, 1754–1760.

**Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R**; 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics **25**, 2078–2079.

**Limpens E, Franken C, Smit P, Willemse J, Bisseling T, Geurts R.** 2003. LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. Science **302**, 630–633.

**Lin L, Zhang C, Chen Y, Wang Y, Wang D, Liu X, Wang M, Mao J, Zhang J, Xing W.** 2019. PAWH1 and PAWH2 are plant-specific components of an Arabidopsis endoplasmic reticulum-associated degradation complex. Nature Communications **10**, 3492.

**Madsen EB, Madsen LH, Radutoiu S, *et al.*** 2003. A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. Nature **425**, 637–640.

**Maku J, Wang L, Liu F, Liu L, Kelley K, Peng Z, Wang J.** 2018. Involvement of root hair during rhizobial invasion in cultivated peanut (*Arachis hypogaea* L.). American Journal of Plant Sciences **9**, 1646.

**Mansfeld BN, Grumet R.** 2018. QTLseqr: an R package for bulk segregant analysis with next-generation sequencing. Plant Genome **11**, 5.

**Marsh JF, Rakocevic A, Mitra RM, Brocard L, Sun J, Eschstruth A, Long SR, Schultze M, Ratet P, Oldroyd GE.** 2007. *Medicago truncatula* NIN is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. Plant Physiology **144**, 324–335.

**Meng L, Li HH, Zhang LY, Wang JK.** 2015. QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. Crop Journal **3**, 269–283.

**Messinese E, Mun JH, Yeun LH, Jayaraman D, Rougé P, Barre A, Lougnon G, Schornack S, Bono JJ, Cook DR.** 2007. A novel nuclear protein interacts with the symbiotic DMI3 calcium-and calmodulin-dependent protein kinase of *Medicago truncatula*. Molecular Plant-Microbe Interactions **20**, 912–921.

**Middleton PH, Jakab J, Penmetsa RV, *et al.*** 2007. An ERF transcription factor in *Medicago truncatula* that is essential for Nod factor signal transduction. The Plant Cell **19**, 1221–1234.

**Mitra RM, Gleason CA, Edwards A, Hadfield J, Downie JA, Oldroyd GE, Long SR.** 2004. A $Ca^{2+}$/calmodulin-dependent protein kinase required for symbiotic nodule development: gene identification by transcript-based cloning. Proceedings of the National Academy of Sciences, USA **101**, 4701–4705.

**Murakami Y, Miwa H, Imaizumi-Anraku H, Kouchi H, Downie JA, Kawaguchi M, Kawasaki S.** 2006. Positional cloning identifies *Lotus japonicus* NSP2, a putative transcription factor of the GRAS family, required for NIN and ENOD40 gene expression in nodule initiation. DNA Research **13**, 255–265.

**Murakami Y, Yokoyama H, Fukui R, Kawaguchi M.** 2013. Down-regulation of NSP2 expression in developmentally young regions of *Lotus japonicus* roots in response to rhizobial inoculation. Plant and Cell Physiology **54**, 518–527.

**Nigam S, Arunachalam V, Gibbons R, Bandyopadhyay A, Nambiar P.** 1980. Genetics of non-nodulation in groundnut (*Arachis hypogaea* L.). Oleagineux **35**, 453–455.

**Oldroyd GE.** 2013. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. Nature Reviews. Microbiology **11**, 252–263.

**Oldroyd GED, Long SR.** 2003. Identification and characterization of nodulation-signaling pathway 2, a gene of *Medicago truncatula* involved in Nod factor signaling. Plant Physiology **131**, 1027–1032.

**Oldroyd GE, Murray JD, Poole PS, Downie JA.** 2011. The rules of engagement in the legume-rhizobial symbiosis. Annual Review of Genetics **45**, 119–144.

**Peng Z, Liu FX, Wang LP, Zhou H, Paudel D, Tan LB, Maku J, Gallo M, Wang JP.** 2017. Transcriptome profiles reveal gene regulation of peanut (*Arachis hypogaea* L.) nodulation. Scientific Reports **7**, 12.

**Peng Z, Tan LB, Lopez Y, *et al.*** 2018. Morphological and genetic characterization of non-nodulating peanut recombinant inbred lines. Crop Science **58**, 540–550.

**Peoples MB, Brockwell J, Herridge DF, *et al.*** 2009. The contributions of nitrogen-fixing crop legumes to the productivity of agricultural systems. Symbiosis **48**, 1–17.

**Powell AF, Doyle JJ.** 2016. Enhanced rhizobial symbiotic capacity in an allopolyploid species of *Glycine* (Leguminosae). American Journal of Botany **103**, 1771–1782.

**Radutoiu S, Madsen LH, Madsen EB, *et al.*** 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. Nature **425**, 585–592.

**Rapp RA, Udall JA, Wendel JF.** 2009. Genomic expression dominance in allopolyploids. BMC Biology **7**, 18.

**Schauser L, Roussis A, Stiller J, Stougaard J.** 1999. A plant regulator controlling development of symbiotic root nodules. Nature **402**, 191–195.

**Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC.** 2004. Mining EST databases to resolve evolutionary events in major crop species. Genome **47**, 868–876.

**Schmutz J, Cannon SB, Schlueter J, *et al.*** 2010. Genome sequence of the palaeopolyploid soybean. Nature **463**, 178–183.

**Sinharoy S, DasGupta M.** 2009. RNA interference highlights the role of CCaMK in dissemination of endosymbionts in the Aeschynomeneae legume *Arachis*. Molecular Plant-Microbe Interactions **22**, 1466–1475.

**Smit P, Raedts J, Portyanko V, Debellé F, Gough C, Bisseling T, Geurts R.** 2005. NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. Science **308**, 1789–1791.

**Sprent JI, James EK.** 2007. Legume evolution: where do nodules and mycorrhizas fit in? Plant Physiology **144**, 575–581.

**Stamatakis A.** 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**, 1312–1313.

**Tseng YC, Tillman BL, Peng Z, Wang JP.** 2016. Identification of major QTLs underlying tomato spotted wilt virus resistance in peanut cultivar Florida-EP (TM) '113'. BMC Genetics **17**, 14.

**Wang YS, Pi LY, Chen X, *et al.*** 2006. Rice XA21 binding protein 3 is a ubiquitin ligase required for full Xa21-mediated disease resistance. The Plant Cell **18**, 3635–3646.

**Yang Z.** 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution **24**, 1586–1591.

**Yano K, Yoshida S, Müller J, *et al.*** 2008. CYCLOPS, a mediator of symbiotic intracellular accommodation. Proceedings of the National Academy of Sciences, USA **105**, 20540–20545.

**Young ND, Debellé F, Oldroyd GE, *et al.*** 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature **480**, 520–524.

**Zhou H, Zhou M, Yang Y, *et al.*** 2014. RNase Z(S1) processes UbL40 mRNAs and controls thermosensitive genic male sterility in rice. Nature Communications **5**, 4884.

**Zhuang W, Chen H, Yang M, *et al.*** 2019. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nature Genetics **51**, 865–876.