

A wound-induced small polypeptide gene family is upregulated in soybean nodules

LI XiaoLin, XU Ji, YU GuanQiao & LUO Li*

State Key Laboratory of Plant Molecular Genetics, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

Received April 27, 2012; accepted August 28, 2012; published online October 29, 2012

Small peptides function as key signals in processes, such as plant cell differentiation, organ development and defenses to biotic stresses. A large number of small peptide precursor genes have been predicted from the analysis of the soybean (*Glycine max*) whole genome DNA sequence. However, most of these genes have unknown characteristics and functions. In this report, we systematically searched for the gene families of small peptide precursors that are up-regulated in soybean nitrogen-fixing root nodules. We found 212 genes (encoding peptides shorter than 150 amino acids) that were up-regulated, and among them, 79 genes belong to 38 multiple-gene families, but the other 133 genes are unique. Twenty-eight of 38 families are conserved in *Arabidopsis*, but the other 10 only exist in legumes. We also identified 16 out of the 38 members of the wound-induced polypeptide (WIP) gene family to be upregulated in nitrogen-fixing nodules. We further analyzed homologs of WIP genes in *Medicago*, *Lotus*, *Arabidopsis* and *Oryza* species and found that a few homologous genes from *Medicago truncatula* and *Lotus japonicus* were also upregulated in their nodules and some WIP genes were induced by specific fungal pathogens on soybean and rice. Structure prediction indicated that all WIP prepropeptides contain a conserved DUF3774 domain (including two hydrophobic regions) and most of them have an N-terminal signal sequence. Fluorescence microscopy analysis of two WIP prepropeptides fused to GFP revealed that these proteins are located on the plasma membrane of tobacco leaf cells. Interestingly, 34 soybean WIP genes are clustered onto three soybean chromosomes, different from known peptide gene families (such as *CLE*). Among them, 11 highly identical genes are aligned on the 6th chromosome, 12 on the 12th, and 11 on the 13th chromosomes. Most of WIP genes from the 12th chromosome share the highest identities with their homologs on the 13th chromosome, suggesting that ancestral WIP genes could have originated from the 13th chromosome, then spread onto the 12th chromosome by chromosome homologous recombination; the new WIP genes could have existed in multiple copies by gene duplication which then spread onto the 6th chromosome. In *Arabidopsis* and *Oryza* species, half of the WIP genes are also aligned on one chromosome and showed higher identity with those from the soybean 12th and 13th chromosomes, suggesting that WIP genes originated from one common ancestor.

soybean, small polypeptide, root nodule, WIP, gene expression, subcellular location, molecular evolution

Citation: Li X L, Xu J, Yu G Q, et al. A wound-induced small polypeptide gene family is upregulated in soybean nodules. *Chin Sci Bull*, 2013, 58: 1003–1009, doi: 10.1007/s11434-012-5495-2

Many small peptides have been identified recently in different plants, such as defensin [1], systemin [2], PSK (phytosulfokine) [3], ENOD40 [4], CLV3 [5], POLARIS [6], IDA [7], DVL1 [8], PSY1 [9], TDIF [10], RGF [11], CLE2 [5], CEP1 [12], RALF [13], Stomagen [14], EPF1 [15], FPF1 [16], SCR [17], LURE [18], TPD1 [19] and NCR [20]. Some of these peptides including PSK, CLV3, RGF, EPF1,

FPF1, POLARIS, ENOD40 and RALF are involved in cell differentiation and organ development, and thus are designated as peptide hormones. Other peptides, such as defensin, systemin and NCR, participate in the interactions between plants and microbes. It has been reported that small peptides, such as CLV3 and PSK, play a regulatory role in innate immunity of plants [21,22].

These active peptides are derived from <150 amino-acid precursors or prepropeptides, containing an N-terminal sig-

*Corresponding author (email: lluo@sibs.ac.cn)

nal sequence and at least one protease site. Some of these peptides carry post-translational modifications, such as tyrosine sulfation, proline hydroxylation and hydroxyproline arabinosylation. Many of them contain several conserved cysteine residues to form intramolecular disulfide bonds. In a model plant, *Arabidopsis thaliana*, analysis of its whole genome sequence yielded more than 900 small peptide precursor genes that contained the conserved signal sequences. A few new small peptide genes (*RGF*, *PSY* and *CEPI*) have been identified using an *in-silico* method.

Legumes including soybean, pea, and alfalfa are broadly distributed on the land, and they provide the source of protein for human and animals. Under the limitation of combined nitrogen, nodules are elicited on legumes through a symbiotic interaction with soil *Rhizobium* to fix atmospheric dinitrogen, called symbiotic nitrogen fixation. This symbiotic interaction is of great importance in agriculture and ecology. Small peptides are involved in the leguminous nodulation and symbiosis. For example, at the early stage of nodule development, the peptide of ENOD40 is required for the initiation of nodule primodium [4]. In the indeterminate nodule found in legumes including alfalfa and peas, defensin-like peptides of NCR induce the terminal differentiation of bacteroids [20]. Recently, the CLE peptides has been suggested to be recognized by LRR receptor kinase, LjHAR1/Mt SUNNY/GmNARK, involved in the autoregulation of nodulation [23,24]. Using the published genome sequences of *Glycine max* (soybean) and *Medicago truncatula*, and a large number of expressed sequence tags (EST) in public databases, small peptide genes involved in symbiosis can be identified from legumes using bioinformatical techniques [25–27]. In this report, the whole-genome transcriptomic data of soybean were analyzed, and 212 small peptide precursor genes upregulated in nitrogen-fixing nodules were identified. Among them, the structure, expression and evolution of a gene family, *WIP* (wound-induced protein) was analyzed. It is important to understand the regulation of leguminous small peptides in symbiotic nitrogen fixation.

1 Materials and methods

1.1 *In silico* screening for soybean small peptide gene families that are upregulated in nitrogen-fixing nodules

The transcriptomic data of *G. max* Williams 82 was downloaded [25,26] and searched for genes that are expressed at more than 2 folds in nitrogen-fixing nodules than in any other organs. Those genes encoding small peptide precursors with less than 150 amino acids were selected as candidate genes. Their signal peptide sequences and hydrophobic regions were predicted using online programs such as SignalP 3.1 and Jpred. The homologs and orthologs were searched using BLAST (NCBI). Motif scanning was performed using the MEME software to find the conserved

region of the small polypeptides. The phylogenetic tree was reconstructed using MAGA 3.0. Those genes clustered onto one clade of the phylogenetic tree and containing the same motifs were classified as belonging to one gene family. As a positive control, this method was tested on known *PSK* and *RGF* gene families.

1.2 Identifying members of the *WIP* gene family

We searched for homologs or orthologs of *WIP* genes by analyzing sequences using BLAST or by comparing with the genomic sequences of *G. max* and *M. truncatula* [25,27]. The multiple sequence alignment was performed using ClustalW program, and the domain of DUF3774 was scanned for known domains using Pfam and SMART programs. Those gene products containing a DUF3773 domain belong to the *WIP* gene family.

1.3 Tissue-specific expression of soybean *WIP* genes

Roots, stems, leaves, flowers, pods and nodules from soybean (Williams 82) grown in the greenhouse under the optimal conditions were collected. Their total RNA was extracted using the TransZol Plant kit (TRANS, Beijing). The purity and concentration of total RNA was determined using an ultraviolet spectrometer. The total RNA was reverse transcribed to cDNA using the PrimeScript RT reagent Kit with a DNA eraser (TaKaRa, Dalian). The primers used for the reverse-transcriptional PCR were listed in Table 1, and PCR was conducted with *TUBULIN* as an inner standard. The PCR was performed with this cycle: 94°C, 3 min (pre-denature); 94°C, 30 s (denature), 60°C, 20 s (anneal), 72°C, 30 s (extension); 72°C, 7 min.

1.4 Subcellular location of soybean *WIP* genes

The DNA fragments of *Gly13g32070* and *Gly13g32690* open reading frames were amplified by PCR using oligo primers (Table 1) and the soybean nodule cDNA as template. The PCR products were cloned into pCMBIA2300s-GFP to construct the plasmids 35S-Gly13g32070:GFP, 35S-GFP:Gly13g32070, 35S-Gly13g32690:GFP and 35S-GFP:Gly13g32690. Each plasmid was introduced into *Agrobacterium tumefaciens* EH105 using a freeze-thaw method. These strains were streaked onto YEB agar plates with appropriate antibiotics, and grown at 28°C for 2 d. The single colony was subcultured into 5 mL LB broth, and subjected to shaking at 28°C at the speed of 200 r/min for 2 d. The culture at $A_{600}=0.4-0.8$ was centrifuged at the speed of 4000 r/min for 10 min to collect bacterial cells. The cells were resuspended in 10 mmol L⁻¹ MgCl₂ with 100 mmol L⁻¹ acetonesyringone, and maintained at room temperature for 1–2 h. The injector containing 1 mL of the bacterial suspension was used to transform 2-week old tobacco leaves by high pressure. The injected sections of tobacco leaves were

Table 1 PCR primers

| | Gene | Forward (5'→3') | Reverse (5'→3') |
|-------------------------|--------------------------|---------------------------------|---------------------------------|
| RT-PCR | <i>Glyma13g37780.1</i> | TGGTTCTTCTCACAAAGGTGGCTC | ACCTGCCACTACACAGTTCTTG |
| | <i>Glyma13g37750.1</i> | TGCAGGTGGAACAGTGTGATGAGA | GATTGGTGCACCTTCTGGCTCTTT |
| | <i>Glyma06g45390.1</i> | AGCACAGAAAGCTTGGATTGTGGC | TTAAGGACTCCTCTGACTGCTTCTGC |
| | <i>Glyma06g45350.1</i> | ACCACTCCGAAAACAACAAAAC | TGGGAAATCGAAGTCCAATAAG |
| | <i>Glyma06g45380.1</i> | ACCACCTTCCTCTCACTTCCTCAT | GTGTTGTGGAGCTCATAGCAGAAG |
| | <i>Glyma12g34660.1</i> | CAGCCTCAACCCAAACCAACCAA | ACCGTCTCATCGTAGCAACACCTA |
| | <i>Glyma12g32720.1</i> | TATCTGCAGGTGGAACAGCGTGAT | AGACACTTCTGTGATTTCGTGGGTG |
| | <i>Glyma12g11660.1</i> | TGGCGGTTGATGGAAGTTATTTCT | ATTGGTCGTATGGCATCGAGTTGG |
| | <i>Glyma03g27140.1</i> | ACCAAAGCTTGGATAGTGG | GGAAATCGTAGCAGAGGAC |
| | <i>Glyma12g11740.1</i> | TAGTGTGGAGTAGTGGAGGCCCTTG | CTTAAGAACTCCTCTGATTGCTTTCC |
| | <i>Glyma06g45260.1</i> | ATGAGTTCAGCACAGAGAGC | AGAAGAAAAGGAAAAGAAGGTT |
| | <i>GmTUBULIN</i> | TGCCCTGAATGTGGATGTGACTGA | AGCAGCGTTTACATCTTTGGGCAC |
| | Subcellular localization | <i>GFPGlyma13g32070</i> | ACGCGTCGACATGAGCTTGAGGTACATGAG |
| <i>Glyma13g32070GFP</i> | | GCTCTAGAAATGAGCTTGAGGTACATGAGC | CGGGATCCTCAGTTGGGACCCAGCAAC |
| <i>GFPGlyma12g32690</i> | | ACGCGTCGACATGTGTCGTGGAACCTTTGTG | GACTAGTAAAGAATTAGGACCCCATGTGCTC |
| <i>Glyma12g32690GFP</i> | | GCTCTAGAAATGTGTCGTGGAACCTTTTG | CGGGATCCCTAAGAATTAGGACCCCATG |

marked and checked under the fluorescent microscope after 3 d.

1.4 Phylogenetic analysis of *WIP* genes

The phylogenetic tree of *WIP* genes was reconstructed using MEGA 3.0. The genes with the closest relationship were marked between two chromosomes, and then lined with each other to form a synteny map.

2 Results and discussion

2.1 Identification of gene families of small polypeptides that are upregulated in soybean nodules

We found a total of 1465 genes that are dominantly expressed in soybean nodules by analyzing the whole-genome transcriptomic data of soybean. Among them, 212 gene products were less than 150 amino acids, and 31 genes were only expressed in nodules, but 18 genes (8%) existed in multiple copies. For example, *ENOD24* has 10 copies. The 51 gene products (24%) contain N-terminal signal sequences, 24 (11%) have signal anchors, and 57 (27%) carry transmembrane regions as predicted by SignalP3.0 and Jpred (Figure 1). In addition, 79 genes belong to 38 multiple-gene families, but the other 133 genes are unique by homologous sequence analysis. Twenty-eight of 38 families are conserved in *A. thaliana*, but 5 only exist in soybean, and the other five also exist in *M. truncatula* and *Lotus japonicas* (Table 2). The other six conserved gene families, such as PSK [3], RGF (root meristem growth factor) [11], EPF1 (epithelial patterning factor 1) [15], FPF1 (flowering promoting factor 1) [16], AGP (arabinoxylated glycopeptide) [28] and DVL (DEVIL 1) [8], have been studied previously.

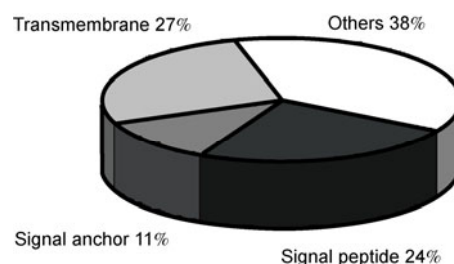


Figure 1 Classification of small polypeptide genes that are upregulated in soybean nodules. The percentage of small polypeptides containing an N-terminal signal peptide, a signal anchor and transmembrane sequences.

2.2 Identification of *WIP* gene family

The largest gene family found in our analysis of small polypeptides consisted of wound-induced proteins. This gene family, named *WIP*, was composed of 38 members in the soybean genome, and 7 homologs in the genomes of *A. thaliana* and *Oryza japonica*. The *WIP* small polypeptide is composed of about 90 amino acids, containing a conserved domain, DUF3774 (Figures 2 and S1). Most of the *WIP* members harbor an N-terminal signal sequence (Figures 2 and S1). Motif prediction showed that DUF3774 contains two conserved motifs: CRWN and WGNP (Figures 2 and S1). Secondary structure prediction showed that one hydrophobic α -helix is located onto the inside of CRWN and outside of WGNP. These analyses suggest that the *WIP* gene may encode a transmembrane small polypeptide.

To further analyze *WIP*, the two soybean *WIP* gene (*Glyma12g32690.1* and *Glyma13g32070.1*) open reading frames were fused with GFP, and transformed into young tobacco leaf cells. Fluorescent microscopy showed that GFP-*Glyma12g32690.1* fusion was located onto the plasma membrane, but *Glyma12g32690*-GFP was observed in the

Table 2 Small polypeptide gene families from soybean nodules^{a)}

| Gene family | <i>Gm</i> | <i>Mt</i> | <i>Lj</i> | <i>At</i> |
|---------------|-----------|-----------|-----------|-----------|
| <i>PSK</i> | 11 | 2 | 5 | 6 |
| <i>RGF</i> | 13 | 4 | 1 | 9 |
| <i>EPF1</i> | 33 | 12 | 3 | 9 |
| <i>FPF1</i> | 10 | 1 | 2 | 3 |
| <i>AGP</i> | 8 | 5 | 1 | 11 |
| <i>DVL</i> | 27 | 12 | 0 | 24 |
| <i>WIP</i> | 39 | 10/5 | 9 | 7 |
| <i>JPR</i> | 12 | 2 | 5 | 4 |
| <i>CRP</i> | 14 | 4 | 3 | 6 |
| <i>CKL1</i> | 5 | 3 | 1 | 1 |
| <i>CKL2</i> | 6 | 3 | 1 | 6 |
| <i>CKL3</i> | 5 | 2 | 1 | 3 |
| <i>CKL4</i> | 7 | 3 | 3 | 2 |
| <i>CKL5</i> | 4 | 1 | 0 | 2 |
| <i>CKL6</i> | 5 | 1 | 1 | 3 |
| <i>CKL7</i> | 7 | 1 | 3 | 1 |
| <i>CKL8</i> | 5 | 0 | 0 | 2 |
| <i>CKL9</i> | 6 | 2 | 2 | 1 |
| <i>CKL10</i> | 8 | 5 | 3 | 5 |
| <i>CKL11</i> | 5 | 2 | 0 | 6 |
| <i>CKL12</i> | 18 | 4 | 7 | 12 |
| <i>CKL13</i> | 3 | 0 | 2 | 1 |
| <i>CKL14</i> | 3 | 0 | 1 | 1 |
| <i>4F5</i> | 3 | 1 | 2 | 2 |
| <i>YOS</i> | 7 | 2 | 2 | 2 |
| <i>LEA</i> | 9 | 4 | 4 | 4 |
| <i>HIG</i> | 6 | 2 | 2 | 2 |
| <i>ENOD93</i> | 6 | 2 | 4 | 1 |
| <i>NRP1</i> | 3 | 2 | 2 | 0 |
| <i>NRP2</i> | 4 | 1 | 1 | 0 |
| <i>NRP3</i> | 1 | 1 | 1 | 0 |
| <i>DRP</i> | 4 | 0 | 0 | 0 |
| <i>NCR</i> | 5 | 0 | 0 | 0 |
| <i>WRP</i> | 2 | 0 | 0 | 0 |
| <i>SUN1</i> | 2 | 1 | 0 | 0 |
| <i>SUN2</i> | 2 | 1 | 1 | 0 |
| <i>SUN3</i> | 4 | 0 | 1 | 0 |
| <i>SUN4</i> | 2 | 0 | 0 | 0 |

a) *Gm*, Glycine max; *Mt*, *Medicago truncatula*; *Lj*, *Lotus japonicas*; *At*, *Arabidopsis thaliana*.

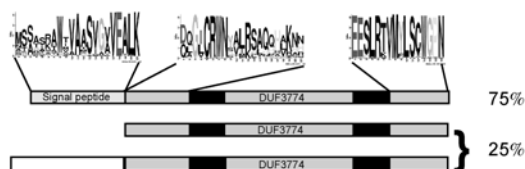


Figure 2 Structure of WIP small polypeptide. The first panel represents three conserved motifs; black boxes in the 2nd, 3rd and 4th panels represent hydrophobic regions of DUF3774 domain; 75% pertains to the ratio of WIP small polypeptides containing N-terminal signal sequence.

nucleus and on the plasma membrane (Figure 3). Both fusions of Glyma13g32070 and GFP were located on the plasma membrane. These data suggest that WIP prepropeptides are transmembrane proteins. *Glyma12g32690* encodes

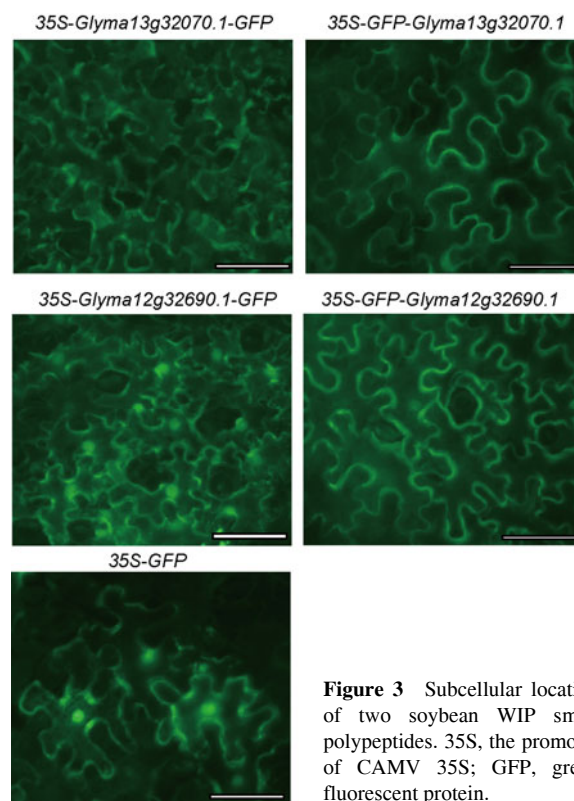


Figure 3 Subcellular location of two soybean WIP small polypeptides. 35S, the promoter of CAMV 35S; GFP, green fluorescent protein.

a WIP protein containing an N-terminal signal sequence. When the GFP was fused with the N-terminus of this WIP precursor, the signal sequence did not work, but the hydrophobic region could bring this protein onto the plasma membrane. When the GFP was linked to its C-terminus, the signal and hydrophobic sequences could function and the protease sites were recognized. Therefore, the GFP could be cleaved off the WIP protein. Glyma13g32070 encodes a WIP protein without an N-terminal sequence and conserved protease processing sites. Thus, both fusions appeared on the plasma membrane possibly through their transmembrane regions.

Additionally, the analysis of the gene structure showed that 31 soybean WIP genes consist of one exon together with a 5'-UTR and 3'-UTR; 4 genes are composed of two exons and one intron; 3 genes with three exons and two introns (Table S1). These data reveal that most of the WIP genes are composed of one exon, similar to the known small peptide gene, *CLE*. Signal peptide prediction analysis revealed that 28 gene products contain an N-terminal signal sequence (Table S1), suggesting that most of WIP peptides are secreted.

2.3 The expression of the WIP gene family

From the analysis of open transcriptomic data, 16 members of 38 soybean genes were found to be upregulated in root nodules (the expression level was more than 2 folds in nodules than in any other organs.), while 7 genes were not

transcribed in any organs [26]. To confirm this result, the total RNA of soybean roots, stems, leaves, pods and nodules was extracted for RT-PCR analysis. The genes *Glyma13g37780.1*, *Glyma13g37750.1*, *Glyma12g34660.1*, *Glyma12g32720.1* and *Glyma12g11740.1* are dominantly expressed in root nodules (Figure 4). Interestingly, *Glyma12g34660.1* and *Glyma12g32720.1* are specifically expressed in the root nodule, while *Glyma13g37780.1* and *Glyma13g37750.1* are also expressed in roots and stems. The transcripts of *Glyma12g11740.1*, *Glyma6g45380.1* and *Gly6g45390.1* were detected in all six organs, while no transcripts of *Glyma12g11660.1*, *Glyma6g45260.1* and *Glyma3g27140.1* was found in any organs. *Glyma6g45350.1* is expressed the highest in roots, though it was also detected in roots, stems and nodules (Figure 4). These 11 genes are not induced by *Bradyrhizobium japonicum* USDA 110 from our data. These data suggest that soybean *WIP* genes may play their roles at the late stage of symbiosis with *Rhizobium*. Our results were mainly consistent with the open transcriptomic data [26], and the difference was probably derived from methods, growth time and conditions of soybean plants.

We obtained the gene-chip probes data of six soybean *WIP* genes (*Glyma06g45370.1*, *Glyma12g32690.1*, *Glyma13g37750.1*, *Glyma12g34660.1*, *Glyma13g35850.1* and *Glyma13g37700.1*) from the website Genevestigator. The expression of *Glyma06g45370.1*, *Glyma12g32690.1*, *Glyma13g37750.1* and *Glyma13g37700.1* was suppressed by *P. pachyrhizi*, while two genes of *Glyma12g34660.1* and *Glyma13g35850.1* were induced by this pathogen. The expression of *Glyma06g45370.1*, *Glyma12g32690.1*, *Glyma12g34660.1* and *Glyma13g35850.1* was significantly increased after infection by *P. sojae*. These data suggest that

some of *WIP* genes are involved in the soybean defense response to pathogens.

Some of the gene-chip probes of *Arabidopsis* and rice *WIP* genes were also obtained from Genevestigator. The expression of *AT4g10270* and *AT4g33560* was induced by *P. parasitica* and salicylic acid, while chitin suppressed their transcription (including *AT4g28240*). *AT4g28240* and *AT4g05070* were significantly suppressed by *Pseudomonas syringae* DC3000 and its elicitor, Flg22. Interestingly, the expression of two rice *WIP* genes (*Os04g0635000* and *Os08g0178500*) increased after infection by *Magnaporthe grisea*. The expression of *Os04g0635000*, *Os04g0635100*, *Os04g0635500* and *Os08g0178500* was apparently enhanced in the transgenic plants over-expressing *nahG*, but downregulated by chitin. These data suggest that some of *WIP* genes may play their roles in plant immunity.

2.4 Phylogenetic analysis of the *WIP* gene family

Among 38 soybean *WIP* genes, 34 were clustered onto three chromosomes (6, 12 and 13). There are 11 genes aligned onto the 6th chromosome, with *Glyma06g45380* and *Glyma06g45390* at the center, and the other genes distributed around them (Figure 5), suggesting that half of these genes originated from the duplication of the other-side genes. *WIP* genes (12 and 11 genes) at the 12th and 13th chromosomes showed higher identity to each other (Figure 5), suggesting that these *WIP* genes on one chromosome may be replicated from the other chromosome. Interestingly, 5, 7 and 4 *WIP* homologs on the 4th chromosome of *A. thaliana* and *O. japonica*, and the 2nd chromosome of *M. truncatula* exhibited synteny (higher identity) with *WIP* genes on the 12th and 13th chromosomes. This observation implies that *WIP* genes on both of these chromosomes could have originated earlier than those on the 6th chromosome in the soybean genome and they could be derived from one common ancestor. Additionally, 5 genes on the 4th chromosome of *A. thaliana* were not clustered, meaning that they have a long evolutionary history. Two *WIP* genes on the 13th chromosome showed lower homology with other genes, suggesting that these genes could be the oldest ones. Therefore, the *WIP* gene could have originated on the 13th chromosome, spread onto 12th chromosome by chromosome duplication, and then dispersed onto the 6th chromosome. Noticeably, the clustered characteristics of *WIP* genes are different from the known small peptide genes of CLE, which are distributed onto 18 chromosomes in the soybean genome.

3 Conclusion

In this report, 38 soybean gene families of small peptide precursors were analyzed, and some of members were upregulated in soybean nitrogen-fixing root nodules including 38 members of the gene family of *WIP*. Most of *WIP*

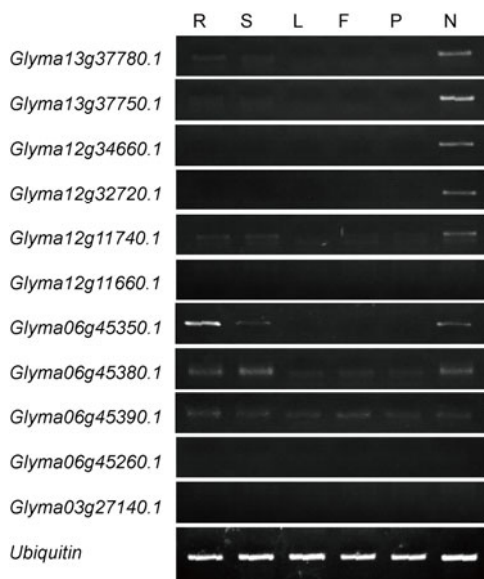


Figure 4 The tissue-specific expression of soybean *WIP* genes. R, root; S, stem; L, leaf; F, flower; P, pods; N, root nodules.

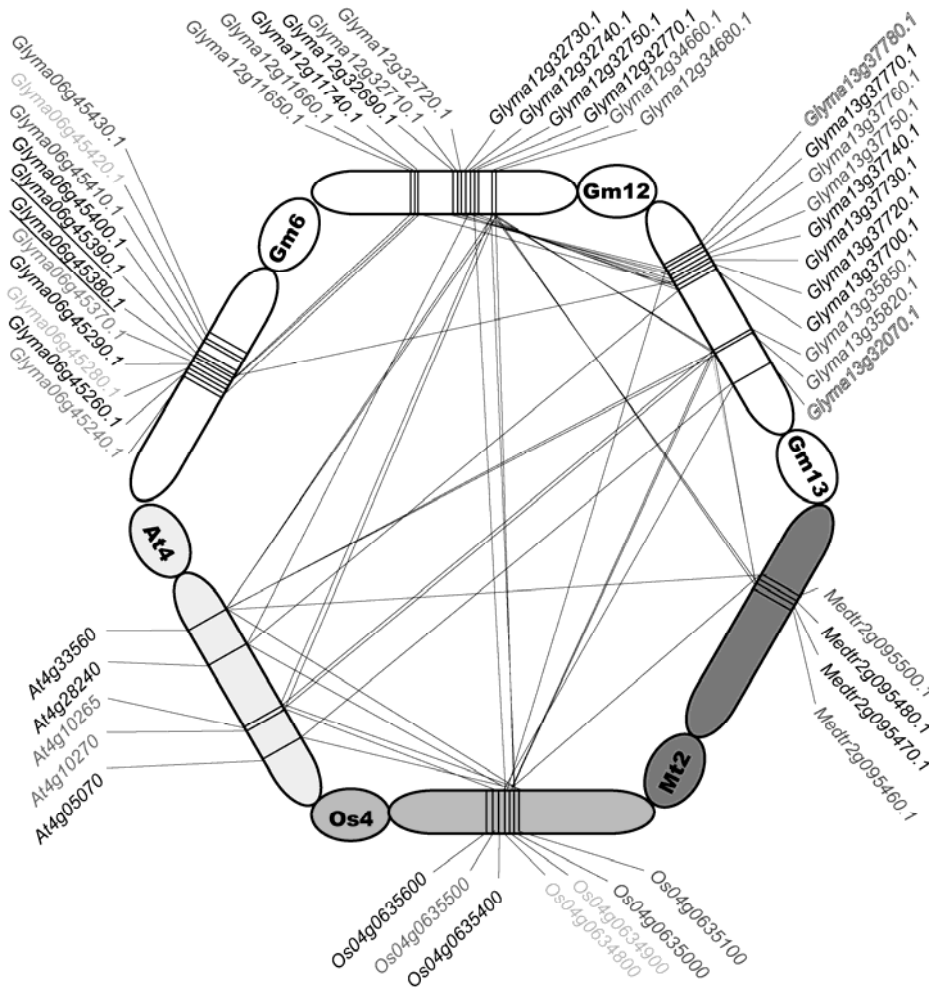


Figure 5 Synteny of *WIP* genes on different chromosomes.

prepeptides consist of one N-terminal signal sequence and one DUF3774 domain, located onto the plasma membrane. 34 soybean *WIP* genes were clustered onto three chromosomes, suggesting that they have a specific evolutionary pathway. About half of the *WIP* genes were dominantly expressed in root nodules, implying that they are involved in symbiotic interactions between soybean and *Rhizobium*. The active form of *WIP* will be identified and its functions in symbiosis or immunity will be studied in the immediate future.

This work was supported by National Key Basic Research Program of China (2011CB100702 and 2010CB126501), National Natural Science Foundation of China (31070218), Shanghai Municipal Natural Science Foundation (09ZR1436500) and Knowledge Innovation Program of Chinese Academy of Sciences (2009KIP206).

- 1 Stotz H U, Thomson J G, Wang Y. Plant defensins: Defense, development and application. *Plant Signal Behav*, 2009, 4: 1010–1012
- 2 Pearce G. Systemin, hydroxyproline-rich systemin and the induction of protease inhibitors. *Curr Protein Pept Sci*, 2011, 12: 399–408
- 3 Yang H, Matsubayashi Y, Hanai H, et al. Phytosulfokine- α , a

peptide growth factor found in higher plants: Its structure, functions, precursor and receptors. *Plant Cell Physiol*, 2000, 41: 825–830

- 4 Batut J, Mergaert P, Masson-Boivin C. Peptide signalling in the rhizobium-legume symbiosis. *Curr Opin Microbiol*, 2011, 14: 181–187
- 5 Wang G, Fiers M. CLE peptide signaling during plant development. *Protoplasma*, 2010, 240: 33–43
- 6 Casson S A, Chilley P M, Topping J F, et al. The POLARIS gene of *Arabidopsis* encodes a predicted peptide required for correct root growth and leaf vascular patterning. *Plant Cell*, 2002, 14: 1705–1721
- 7 Stenvik G E, Butenko M A, Urbanowicz B R, et al. Overexpression of *INFLORESCENCE DEFICIENT IN ABSCISSION* activates cell separation in vestigial abscission zones in *Arabidopsis*. *Plant Cell*, 2006, 18: 1467–1476
- 8 Wen J, Lease K A, Walker J C. DVL, a novel class of small polypeptides: Overexpression alters *Arabidopsis* development. *Plant J*, 2004, 37: 668–677
- 9 Amano Y, Tsubouchi H, Shinohara H, et al. Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in *Arabidopsis*. *Proc Natl Acad Sci USA*, 2007, 104: 18333–18338
- 10 Hirakawa Y, Kondo Y, Fukuda H. TDIF peptide signaling regulates vascular stem cell proliferation via the WOX4 homeobox gene in *Arabidopsis*. *Plant Cell*, 2010, 22: 2618–2629
- 11 Matsuzaki Y, Ogawa-Ohnishi M, Mori A, et al. Secreted peptide signals required for maintenance of root stem cell niche in *Arabidopsis*. *Science*, 2010, 329: 1065–1067
- 12 Ohyama K, Ogawa M, Matsubayashi Y. Identification of a biologically active, small, secreted peptide in *Arabidopsis* by *in silico* gene

- screening, followed by LC-MS-based structure analysis. *Plant J*, 2008, 55: 152–160
- 13 Bedinger P A, Pearce G, Covey P A. RALFs: Peptide regulators of plant growth. *Plant Signal Behav*, 2010, 5: 1342–1346
- 14 Sugano S S, Shimada T, Imai Y, et al. Stomagen positively regulates stomatal density in *Arabidopsis*. *Nature*, 2010, 463: 241–244
- 15 Hara K, Kajita R, Torii K U, et al. The secretory peptide gene EPF1 enforces the stomatal one-cell-spacing rule. *Genes Dev*, 2007, 21: 1720–1725
- 16 Kania T, Russenberger D, Peng S, et al. FPF1 promotes flowering in *Arabidopsis*. *Plant Cell*, 1997, 9: 1327–1338
- 17 Schopfer C R, Nasrallah M E, Nasrallah J B. The male determinant of self-incompatibility in *Brassica*. *Science*, 1999, 286: 1697–1700
- 18 Okuda S, Tsutsui H, Shiina K, et al. Defensin-like polypeptide LUR-ES are pollen tube attractants secreted from synergid cells. *Nature*, 2009, 458: 357–361
- 19 Yang S L, Xie L F, Mao H Z, et al. *T APETUM DETERMINANT1* is required for cell specialization in the *A rabidopsis* anther. *Plant Cell*, 2003, 15: 2792–2804
- 20 Van de Velde W, Zehirov G, Szatmari A, et al. Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science*, 2010, 327: 1122–1126
- 21 Igarashi D, Tsuda K, Katagiri F. The peptide growth factor, phyto-sulfokine, attenuates pattern-triggered immunity. *Plant J*, 2012, 21, doi: 10.1111/j.1365-313X.2012.04950.x
- 22 Lee H, Chah O K, Sheen J. Stem-cell-triggered immunity through CLV3p-FLS2 signalling. *Nature*, 2011, 473: 376–379
- 23 Mortier V, Fenta B A, Martens C, et al. Search for nodulation-related CLE genes in the genome of *Glycine max*. *J Exp Bot*, 2011, 62: 2571–2583
- 24 Reid D E, Ferguson B J, Gresshoff P M. Inoculation- and nitrate-induced CLE peptides of soybean control NARK-dependent nodule formation. *Mol Plant Microbe Interact*, 2011, 24: 606–618
- 25 Schmutz J, Cannon S B, Schlueter J, et al. Genome sequence of the paleopolyploid soybean (*Glycine max* (L.) Merr.). *Nature*, 2010, 463: 178–183
- 26 Libault M, Farmer A, Joshi T, et al. An integrated transcriptome atlas of the crop model *Glycine max* and its use in comparative analyses in plants. *Plant J*, 2010, 63: 86–99
- 27 Young N D, Debellé F, Oldroyd G E, et al. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature*, 2011, 480: 520–524
- 28 Ellis M, Egelund J, Schultz C J, et al. Arabinogalactan-proteins: Key regulators at the cell surface? *Plant Physiol*, 2010, 153: 403–419

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Supporting Information

Table S1 Constitution of soybean *WIP* genes

Figure S1 Multiple-sequence alignment of *WIP* small polypeptides.

Figure S2 Phylogenetic tree of *WIP* genes from *Arabidopsis thaliana* and *Oryza sativa*.

The supporting information is available online at csb.scichina.com and www.springerlink.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.