

Identification of tissue-specific gene clusters and orthologues of nodulation-related genes in *Vigna angularis*

Yang Jae Kang¹, Jayern Lee¹, Yong Hwan Kim² and Suk-Ha Lee^{1,3*}

¹Department of Plant Science and Research Institute for Agriculture and Life Sciences, Seoul National University, San 56-1, Sillim-dong, Gwanak-gu, Seoul 151-921, Republic of Korea, ²Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries, Anyang-si, Republic of Korea and ³Plant Genomics and Breeding Research Institute, Seoul National University, Seoul 151-921, Republic of Korea

Abstract

Nitrogen fixation in legumes is an important agricultural trait that results from symbiosis between the root and rhizobia. To understand the molecular basis of nodulation, recent research has been focused on the identification of nodulation-related genes by functional analysis using two major model legumes, *Medicago truncatula* and *Lotus japonicus*. Thus far, three important processes have been discovered, namely Nod factor (NF) perception, NF signalling and autoregulation of nodulation. Nevertheless, application of the results of these studies is limited for non-model legume crops because a reference genome is unavailable. However, because the cost of whole-transcriptome analysis has dropped dramatically due to the Next generation sequencer (NGS) technology, minor crops for which reference sequences are yet to be constructed can still be studied at the genome level. In this study, we sequenced the leaf and root transcriptomes of *Vigna angularis* (accession IT213134) and *de novo* assembled. Our results demonstrate the feasibility of using the transcriptome assembly to effectively identify tissue-specific peptide clusters related to tissue-specific functions and species-specific nodulation-related genes.

Keywords: *de novo* transcriptome assembly; nodulation; RNAseq; *Vigna angularis*

Introduction

Nodulation is an important trait of legumes that allows atmospheric nitrogen to be converted into plant-consumable ammonia by symbiosis with the rhizobia (Ferguson *et al.*, 2010). This process can be exploited in sustainable agriculture to replace chemical nitrogen fertilizers, the cost of which continues to rise along with the price of fossil fuels. For the breeding practices to improve its efficiency, it is essential to understand the molecular basis of nitrogen fixation.

Previous research has attempted to identify nodulation-related genes using mutants of two major model legumes, *Medicago truncatula* and *Lotus japonicus*, with regard to flavonoid compound secretion for attracting rhizobia and bacterial Nod factor (NF) perception of the root and auto-regulation of nodulation (AON). Several genes have been reported to be involved in the steps of these processes (Ferguson *et al.*, 2010).

In this study, we sequenced the leaf and root transcriptomes of *Vigna angularis* (accession IT213134) using Illumina HiSeq2000 and then assembled using the Trinity software. Cluster analysis of the coding sequences for each tissue revealed tissue-specific genes. The homologues of nodulation-related genes were assessed for

*Corresponding author. E-mail: sukhalee@snu.ac.kr

tissue specificity and conservation among the legume genomes.

Materials and methods

V. angularis (accession IT213134) was used in this study. The leaf triplet and root tissue samples were harvested at stage V4. RNA was extracted using TRIzol (Invitrogen, Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions. The mRNA was converted into the template-sequencing library suitable for subsequent cluster generation using the TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA, USA). For RNA sequencing, the Illumina HiSeq2000 platform was used to produce about 5 GB for each library, and *de novo* assembly using the Trinity software resulted in 49,509 and 62,922 complete coding sequences for leaf and root tissues (Table S1, available online) (Grabherr *et al.*, 2011).

Results and discussion

Transcriptome assembly and clustering

The peptide sequences derived from each tissue were clustered using the OrthoMCL software to identify tissue-specific peptides (Fig. 1(a)) (Li *et al.*, 2003). The longest peptide from each cluster was chosen as a representative gene for gene ontology (GO) enrichment analysis (Fig. 1(a)). Notably, GO enrichment of root-specific clusters revealed genes associated with 'ion binding', 'protein binding' and 'calmodulin binding' within the 'Binding' category, which is consistent with published reports that demonstrated a role for MtDMI3/PsSYM9, a calcium- and calmodulin-dependent protein kinase, in the NF signalling cascade (Levy *et al.*, 2004; Mitra *et al.*, 2004). The 'Protein binding' category includes receptor-like kinases (RLKs) involved in NF binding. The RLK-like genes in *V. angularis* were surveyed in detail following RLK classification by Shiu and Bleecker (2001) using the MapMan software (Fig. 1(b)) (Thimm *et al.*, 2004). This analysis revealed that *LRR I, III, VII, X, and XI, Crinkly4-like, DUF26, LRK10-like* and *RKF3-like* are mainly found in root-specific clusters.

Identification of nodulation-related genes

The 38 nodulation-related genes were obtained from the NCBI (National Center for Biotechnology Information) according to the literature (Ferguson *et al.*, 2010). Of these, 37 sufficiently matched to representative genes from 24 clusters (Table 1). In addition, we aligned

these nodulation-related genes against peptides derived from the entire *V. angularis* transcriptome, including those predicted from partial coding sequences, and retrieved best hits to determine the orthologues of *V. angularis* genes derived using the BLAST score (Table S2, available online). All genes related to NF perception matched to cluster C4537. The orthologues of these genes were all root-derived peptides.

Genes associated with NF signalling were matched to 16 clusters. Root-derived orthologues were identified for ten genes, namely *MtDMI2, MtDMI3/PsSYM9, MtNSP1, MtERN1, MtIPD3/LjCYCLOPS, LjCERBERUS, LjERF1* and *MtEFD* (Table S2, available online). On the other hand, leaf-derived peptides were identified as the orthologues of five genes, *MtNIN, MtENOD11, LjSIP1, MtRPG* and *MtHMGR1*. However, the expression of *MtENOD11* has been demonstrated during nodule development in *M. truncatula* (Journet *et al.*, 2001), while *MtNIN* is reportedly expressed in the leaf, root and nodule, where its expression is highest (Schäuser *et al.*, 1999). The expression of *LjSIP1* has also been shown in the leaf and root tissues of *L. japonicus* (Zhu *et al.*, 2008), while *MtRPG* has been detected at low levels in the leaf and root of *M. truncatula* (Arrighi *et al.*, 2008). Nevertheless, *MtHMGR1* is expressed in the root and nodule of *M. truncatula* (Kevei *et al.*, 2007). The differences might result from the sampling time of mRNA extraction. *MtDMI1/LjCASTOR/LjPOLLUX, LjNup85, LjNup133, MtNSP2* and *LjERF1* had orthologues that were co-expressed in both leaf and root tissues (Table S2, available online), consistent with previous research on these model legumes. *MtDMI1/LjCASTOR/LjPOLLUX* encodes potassium ion-channel proteins. Although *MtDMI1* is reportedly root specific in *M. truncatula*, *LjCASTOR* and *LjPOLLUX* are expressed in the leaf and root (Imaizumi-Anraku *et al.*, 2005). *LjNup133* and *LjNup85* encode nucleoporins, which are expressed in both the root and leaf of *L. japonicus* (Kanamori *et al.*, 2006; Saito *et al.*, 2007). Similarly, *MtNSP2* is expressed in the leaf and root of *M. truncatula* (Kalo *et al.*, 2005). The expression of *LjERF1* has been assessed in nodule-containing root tissues and found to be present at the initial stage of nodulation (Asamizu *et al.*, 2008).

AON-related genes were matched to seven clusters, of which three consisted of only root-derived peptides. This analysis demonstrated that the orthologues of *GmKAPP1, GmKAPP2, PsNOD3* and *LjKLAVIER* were root-derived peptides (Table S2, available online). Grafting experiments have linked *PsNOD3* to a root-specific function in *Pisum sativum* (Li *et al.*, 2009). *LjKLAVIER* is expressed in the leaf and root of *L. japonicus* (Miyazawa *et al.*, 2010), while *GmKAPP1* and *GmKAPP2* have been found to be expressed in the soybean leaf tissue

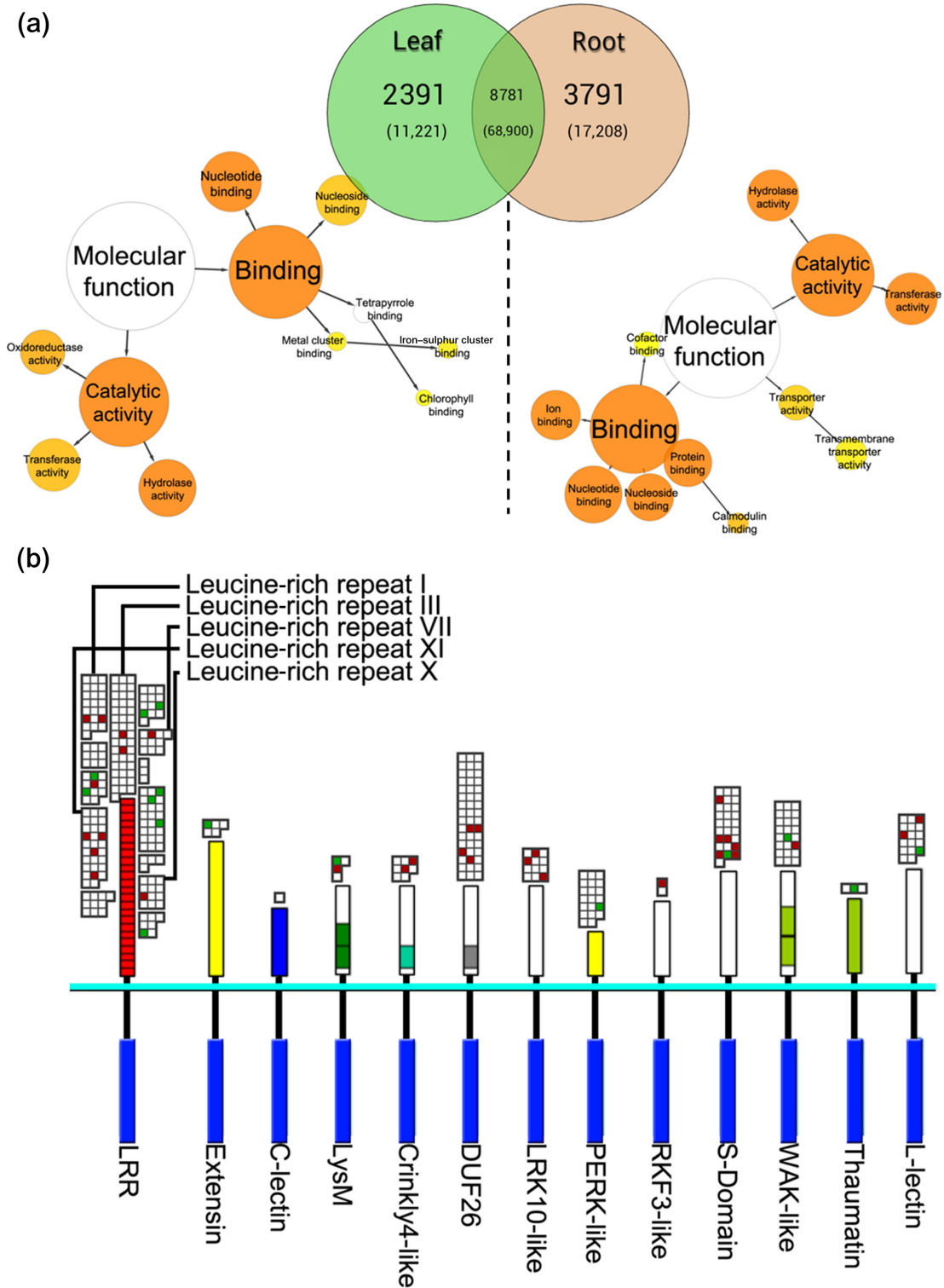


Fig. 1. (a) Venn diagram depicting the tissue-specific peptide clusters and common peptide clusters. BinGO ontology diagrams show the gene ontology enrichment pattern of leaf- and root-specific clusters. The numbers within parentheses indicate the number of peptides within the clusters. (b) Receptor-like kinases (RLKs) in leaf- and root-specific clusters. Several types of RLKs are root specific.

Table 1. Nodulation-related peptide clusters derived from root and leaf transcriptomes

Functions	Cluster ID	Nodulation-related genes	Number of complete peptides		Ratio ^a
			Root	Leaf	
Nod factor perception	C4537	<i>GmNFR5</i> , <i>LjNFR1</i> , <i>LjNFR5</i> , <i>MtLYK3</i> , <i>MtLYK4</i> , <i>MtNFP</i> and <i>PsSYM10</i>	6	1	6.00
Nod factor signalling cascade	C3624	<i>LjCASTOR</i> , <i>LjPOLLUX</i> and <i>MtDMI1</i>	6	2	3
	C1181	<i>LjCERBERUS</i>	2	40	0.05
	C14626	<i>LjCYCLOPS</i> and <i>MtIPD3</i>	2	0	Rs
	C13985	<i>LjERF1</i>	2	0	Rs
	C11450	<i>LjNup133</i>	2	1	2
	C15268	<i>LjNup85</i>	1	1	1
	C2924	<i>LjSIP1</i>	1	9	0.11
	C1666	<i>MtDMI2</i>	21	0	Rs
	C11552	<i>MtDMI3</i> and <i>PsSYM9</i>	3	0	Rs
	C15320	<i>MtEFD</i>	1	1	1
	C14666	<i>MtENOD11</i>	2	0	Rs
	C13602	<i>MtERN1</i>	2	0	Rs
	C2154	<i>MtNIN</i>	15	0	Rs
	C9428	<i>MtNSP1</i>	4	0	Rs
	C6073	<i>MtNSP2</i>	3	2	1.5
	C6915	<i>MtRPG</i>	2	2	1
	Autoregulation of nodulation	C4613	<i>GmKAPP1</i> and <i>GmKAPP2</i>	7	0
C3358		<i>GmNARK</i> , <i>LjHAR1</i> and <i>MtSUNN</i>	3	6	0.5
C10335		<i>LjASTRAY</i>	0	3	Ls
C5277		<i>LjEIN2a</i>	2	4	0.5
C2218		<i>LjETR1</i>	6	8	0.75
C14206		<i>LjKLAVIER</i>	2	0	Rs
C4531		<i>PsNOD3</i>	7	0	Rs

Rs, root specific; Ls, leaf specific.

^a Ratio refers to number of complete root peptides to number of complete leaf peptides.

(Miyahara *et al.*, 2008). Similarly, the orthologues of *GmNARK/LjHAR1/MtSUNN* were leaf-derived peptides. *GmNARK*, *LjHAR* and *MtSUNN* have been found to be expressed in the leaf and root of *Glycine max*, *L. japonicus* and *M. truncatula*, respectively (Nishimura *et al.*, 2002a; Searle *et al.*, 2003; Schnabel *et al.*, 2005). *GmNARK* controls root nodule development by recognizing the root-to-shoot Q signal and its orthologues expressed in the leaves of *V. angularis* may play a similar role. It is notable that the orthologues of *GmKAPP* were detected in the root tissue and that its corresponding cluster, C4613, consisted of only root-derived peptides (Table 1). We identified partial coding sequences in the leaf that aligned against *GmKAPP* with a score similar to that of the root orthologue (Table S2, available online). This leaf-derived partial sequence may play a role similar to that played by Kinase-associated protein phosphatases (KAPP), which interacts with Nodule autoregulation receptor kinase (NARK) in the leaf according to the model of systemic communication for nodulation in *P. sativum* (Li *et al.*, 2009). The orthologues of *LjASTRAY* were found among both the leaf- and root-derived peptides, despite a previous report showing the presence of *LjASTRAY* complementary DNA in only the root tissue

of *L. japonicus* (Nishimura *et al.*, 2002b). *LjETR1* and *LjEIN2a* are required for monitoring ethylene sensitivity. While the orthologue of *LjETR1* was co-expressed in the leaf and root, that of *LjEIN2a* was leaf specific.

Conclusion

We used clustering and GO enrichment analysis to identify candidate RLK genes that may participate in NF signalling and the orthologues of nodulation-related genes. Transcriptome-derived peptides of *V. angularis* possessed the orthologues of nodulation-related genes with a high similarity. Tissue specificity was slightly different between the model legumes; however, possible paralogues that clustered with the orthologues exhibited root- or leaf-specific expression, suggesting that the evolution of *V. angularis* subfunctionalization differs from that of other model legumes. These results demonstrate the feasibility of applying the NGS technology to effectively discover agriculturally important genes in minor crops that do not have a complete reference genome sequence.

Supplementary material

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/S1479262114000185>

Acknowledgements

This study was supported by a grant from the Next-Generation BioGreen 21 Program (no. PJ008060) of the Rural Development Administration, Republic of Korea.

References

- Arrighi JF, Godfroy O, de Billy F, Saurat O, Jauneau A and Gough C (2008) The RPG gene of *Medicago truncatula* controls Rhizobium-directed polar growth during infection. *Proceedings of the National Academy of Sciences of the United States of America* 105: 9817–9822.
- Asamizu E, Shimoda Y, Kouchi H, Tabata S and Sato S (2008) A positive regulatory role for LjERF1 in the nodulation process is revealed by systematic analysis of nodule-associated transcription factors of *Lotus japonicus*. *Plant Physiology* 147: 2030–2040.
- Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE and Gresshoff PM (2010) Molecular analysis of legume nodule development and autoregulation. *Journal of Integrative Plant Biology* 52: 61–76.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N and Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.
- Imaizumi-Anraku H, Takeda N, Charpentier M, Perry J, Miwa H, Umehara Y, Kouchi H, Murakami Y, Mulder L, Vickers K, Pike J, Downie JA, Wang T, Sato S, Asamizu E, Tabata S, Yoshikawa M, Murooka Y, Wu GJ, Kawaguchi M, Kawasaki S, Parniske M and Hayashi M (2005) Plastid proteins crucial for symbiotic fungal and bacterial entry into plant roots. *Nature* 433: 527–531.
- Journet EP, El-Gachtouli N, Vernoud V, de Billy F, Pichon M, Dedieu A, Arnould C, Morandi D, Barker DG and Gianinazzi-Pearson V (2001) *Medicago truncatula* ENOD11: a novel RPRP-encoding early nodulin gene expressed during mycorrhization in arbuscule-containing cells. *Molecular Plant–Microbe Interactions* 14: 737–748.
- Kalo P, Gleason C, Edwards A, Marsh J, Mitra RM, Hirsch S, Jakab J, Sims S, Long SR, Rogers J, Kiss GB, Downie JA and Oldroyd GED (2005) Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science* 308: 1786–1789.
- Kanamori N, Madsen LH, Radutoiu S, Frantescu M, Quistgaard EMH, Miwa H, Downie JA, James EK, Felle HH, Haaning LL, Jensen TH, Sato S, Nakamura Y, Tabata S, Sandal N and Stougaard J (2006) A nucleoporin is required for induction of Ca²⁺ spiking in legume nodule development and essential for rhizobial and fungal symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* 103: 359–364.
- Kevei Z, Loughon G, Mergaert P, Horvath GV, Kereszt A, Jayaraman D, Zaman N, Marcel F, Regulski K, Kiss GB, Kondorosi A, Endre G, Kondorosi E and Ane JM (2007) 3-Hydroxy-3-methylglutaryl coenzyme A reductase 1 interacts with NORK and is crucial for nodulation in *Medicago truncatula*. *Plant Cell* 19: 3974–3989.
- Levy J, Bres C, Geurts R, Chalhoub B, Kulikova O, Duc G, Journet EP, Ane JM, Lauber E, Bisseling T, Denarie J, Rosenberg C and Debelle F (2004) A putative Ca²⁺ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science* 303: 1361–1364.
- Li L, Stoeckert CJ and Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* 13: 2178–2189.
- Li DX, Kinkema M and Gresshoff PM (2009) Autoregulation of nodulation (AON) in *Pisum sativum* (pea) involves signalling events associated with both nodule primordia development and nitrogen fixation. *Journal of Plant Physiology* 166: 955–967.
- Mitra RM, Gleason CA, Edwards A, Hadfield J, Downie JA, Oldroyd GED and Long SR (2004) A Ca²⁺/calmodulin-dependent protein kinase required for symbiotic nodule development: gene identification by transcript-based cloning. *Proceedings of the National Academy of Sciences of the United States of America* 101: 4701–4705.
- Miyahara A, Hirani TA, Oakes M, Kereszt A, Kobe B, Djordjevic MA and Gresshoff PM (2008) Soybean nodule autoregulation receptor kinase phosphorylates two kinase-associated protein phosphatases *in vitro*. *Journal of Biological Chemistry* 283: 25381–25391.
- Miyazawa H, Oka-Kira E, Sato N, Takahashi H, Wu GJ, Sato S, Hayashi M, Betsuyaku S, Nakazono M, Tabata S, Harada K, Sawa S, Fukuda H and Kawaguchi M (2010) The receptor-like kinase KLAVER mediates systemic regulation of nodulation and non-symbiotic shoot development in *Lotus japonicus*. *Development* 137: 4317–4325.
- Nishimura R, Hayashi M, Wu GJ, Kouchi H, Imaizumi-Anraku H, Murakami Y, Kawasaki S, Akao S, Ohmori M, Nagasawa M, Harada K and Kawaguchi M (2002a) HAR1 mediates systemic regulation of symbiotic organ development. *Nature* 420: 426–429.
- Nishimura R, Ohmori M, Fujita H and Kawaguchi M (2002b) A Lotus basic leucine zipper protein with a RING-finger motif negatively regulates the developmental program of nodulation. *Proceedings of the National Academy of Sciences of the United States of America* 99: 15206–15210.
- Saito K, Yoshikawa M, Yano K, Miwa H, Uchida H, Asamizu E, Sato S, Tabata S, Imaizumi-Anraku H, Umehara Y, Kouchi H, Murooka Y, Szczygłowski K, Downie JA, Parniske M, Hayashi M and Kawaguchi M (2007) NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in *Lotus japonicus*. *Plant Cell* 19: 610–624.
- Schauser L, Roussis A, Stiller J and Stougaard J (1999) A plant regulator controlling development of symbiotic root nodules. *Nature* 402: 191–195.
- Schnabel E, Journet EP, de Carvalho-Niebel F, Duc G and Frugoli J (2005) The *Medicago truncatula* SUNN gene encodes a CLV1-like leucine-rich repeat receptor kinase that regulates nodule number and root length. *Plant Molecular Biology* 58: 809–822.

- Searle IR, Men AE, Laniya TS, Buzas DM, Iturbe-Ormaetxe I, Carroll BJ and Gresshoff PM (2003) Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase. *Science* 299: 109–112.
- Shiu SH and Bleeker AB (2001) Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proceedings of the National Academy of Sciences of the United States of America* 98: 10763–10768.
- Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY and Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant Journal* 37: 914–939.
- Zhu H, Chen T, Zhu MS, Fang Q, Kang H, Hong ZL and Zhang ZM (2008) A novel ARID DNA-binding protein interacts with SymRK and is expressed during early nodule development in *Lotus japonicus*. *Plant Physiology* 148: 337–347.