Plant Biotechnology Journal

Plant Biotechnology Journal (2012) **10**, pp. 995–1010 doi: 10.1111/j.1467-7652.2012.00729.x

Transient Nod factor-dependent gene expression in the nodulation-competent zone of soybean (*Glycine max* [L.] Merr.) roots

Satomi Hayashi[1], Dugald E. Reid[1], Michał T. Lorenc[2], Jiri Stiller[2], David Edwards[2], Peter M. Gresshoff[1,*] and Brett J. Ferguson[1]

[1]*Australian Research Council Centre of Excellence for Integrative Legume Research, The University of Queensland, St. Lucia, Brisbane, Qld, Australia*

[2]*Australian Centre for Plant Functional Genomics, School of Agriculture and Food Sciences, The University of Queensland, Brisbane, Qld, Australia*

Received 24 April 2012;
revised 26 June 2012;
accepted 28 June 2012.
**Correspondence* (Tel +617 33653550; fax +617 33653556; email p.gresshoff@uq.edu.au)

Keywords: plant development, legume, nodule, signalling, symbiosis, transcriptome.

Summary

All lateral organ development in plants, such as nodulation in legumes, requires the temporal and spatial regulation of genes and gene networks. A total mRNA profiling approach using RNA-seq to target the specific soybean (*Glycine max*) root tissues responding to compatible rhizobia [i.e. the Zone Of Nodulation (ZON)] revealed a large number of novel, often transient, mRNA changes occurring during the early stages of nodulation. Focusing on the ZON enabled us to discard the majority of root tissues and their developmentally diverse gene transcripts, thereby highlighting the lowly and transiently expressed nodulation-specific genes. It also enabled us to concentrate on a precise moment in early nodule development at each sampling time. We focused on discovering genes regulated specifically by the *Bradyrhizobium*-produced Nod factor signal, by inoculating roots with either a competent wild-type or incompetent mutant (*nodC*$^{-}$) strain of *Bradyrhizobium japonicum*. Collectively, 2915 genes were identified as being differentially expressed, including many known soybean nodulation genes. A number of unknown nodulation gene candidates and soybean orthologues of nodulation genes previously reported in other legume species were also identified. The differential expression of several candidates was confirmed and further characterized via inoculation time-course studies and qRT-PCR. The expression of many genes, including an *endo-1,4-β-glucanase*, a *cytochrome P450* and a *TIR-LRR-NBS receptor kinase*, was transient, peaking quickly during the initiation of nodule ontogeny. Additional genes were found to be down-regulated. Significantly, a set of differentially regulated genes acting in the gibberellic acid (GA) biosynthesis pathway was discovered, suggesting a novel role of GAs in nodulation.

Introduction

The genetic regulation of lateral organ development in plants is well illustrated by the legume–*Rhizobium* symbiosis (cf., Caetano-Anollés and Gresshoff, 1991; Ferguson, 2012; Ferguson *et al.*, 2010). The processes of determination, initiation, growth and development, as well as feedback regulation, commonly involve small signal molecules, namely hormones or hormone-like factors (Biswas *et al.*, 2009; Ferguson and Mathesius, 2003), short- and long-distance transport, receptors and transcriptional factors that facilitate the regulation of cell division and differentiation (cf., Beveridge *et al.*, 2007).

During nodule ontogeny, only a small portion of the legume root responds to compatible rhizobia (Bhuvaneswari *et al.*, 1980; Calvert *et al.*, 1984); this region is known as the 'Zone Of Nodulation (ZON) ' and is located between the zone of elongation and the zone where new root hairs emerge (Zone of Emerging Root Hairs [ZERH]; Bhuvaneswari *et al.*, 1980; Gresshoff and Delves, 1986). Mature root regions with elongated root hairs typically fail to respond to rhizobia (Bhuvaneswari *et al.*, 1980; Calvert *et al.*, 1984).

Nodulation is initiated by the legume exuding flavonoid molecules (Kosslak *et al.*, 1987; Peters *et al.*, 1986; Redmond *et al.*, 1986), which induce the synthesis of rhizobia-produced lipochito-oligosaccharides, called Nod factors (NF; reviewed by Caetano-Anollés and Gresshoff, 1991; Dénarié *et al.*, 1996; Spaink, 2000). The perception of NF triggers a root signalling cascade essential for rhizobia infection (Dénarié *et al.*, 1996; Desbrosses and Stougaard, 2011; Indrasumunar *et al.*, 2010, 2011; Oldroyd and Downie, 2008). Significantly, the plant also responds to the presence of the bacteria, including their exopolysaccharides, lipopolysaccharides and metabolites such as purines and quorum-sensing molecules (Giraud *et al.*, 2007; Jones *et al.*, 2008; Leigh *et al.*, 1985; Mathesius *et al.*, 2003).

The number of developing nodules is regulated by the legume through systemic 'Autoregulation Of Nodulation (AON)' (Delves *et al.*, 1986; Ferguson *et al.*, 2010; Gresshoff and Delves, 1986; Han *et al.*, 2010; Reid *et al.*, 2011a). AON is initiated during nodule primordium formation (Li *et al.*, 2009) and involves the production of CLAVATA3/ESR-related (CLE) peptides (Lim *et al.*, 2011; Mortier *et al.*, 2010; Okamoto *et al.*, 2009; Reid *et al.*, 2011b). Some CLE peptides are predicted, although not proven, to act as the ligand for a CLAVATA1-like LRR receptor kinase (Krusell *et al.*, 2002; Nishimura *et al.*, 2002; Schnabel *et al.*, 2005; Searle *et al.*, 2003). Activation of this receptor is proposed to initiate the production of a low molecu-

lar weight shoot-derived inhibitor (SDI) that is transported to the root where it inhibits further nodule initiation (Lin *et al.*, 2010, 2011b; Mathews *et al.*, 1989).

Over the last decades, forward mutagenesis approaches using legumes identified a number of genetic components required for both nodulation and AON (reviewed in Crespi and Frugier, 2008; Ferguson *et al.*, 2010; Oldroyd and Downie, 2008). In parallel, differential gene expression screens identified nodulation-enhanced genes called 'nodulins' (reviewed in Lievens *et al.*, 2001). More recently, genome-wide transcriptome analyses using microarrays were performed to investigate the gene expression with a greater emphasis on variations occurring over time or in different tissues (Colebatch *et al.*, 2004; El Yahyaoui *et al.*, 2004; Kinkema and Gresshoff, 2008; Lohar *et al.*, 2006; Maguire *et al.*, 2002; and others). It is worthwhile to note that the data may not only reflect gene expression but also integrate RNA turnover, meaning that these approaches in fact monitor total mRNA levels.

To date, limitations with both the genetic and molecular approaches used have resulted in many nodulation-related genes being overlooked. These limitations include genetic duplication and redundancy preventing mutant isolation, the lack of a proper inoculation control (usually mock inoculation with water or buffer) and the harvesting of entire root systems that harbour continuous, and often AON-regulated, developmental stages. Gene chip technology was also limited as it was hybridization based and thus failed to effectively distinguish members of multigene families. Furthermore, for soybean, the presence of a fragmented and duplicated amphidiploid genome containing diversified gene homeologues (Schmutz *et al.*, 2010) considerably complicated transcript analyses.

High-throughput RNA sequencing (RNA-seq) overcomes many of the above-mentioned shortfalls. RNA-seq uses next-generation sequencing technology to identify presumably all mRNA transcripts. Each individual transcript is then mapped to predicted genome sequences to identify how various genes are expressed.

Soybean is one of the most important crop species, providing human food, animal feed, nutritional by-products (i.e. lecithin, isoflavones) and even vegetable oil for human consumption and biofuel production. It is also important scientifically as a 'model legume species' (Ferguson and Gresshoff, 2009; Gresshoff, 2012) providing a rich database, particularly in biochemical and agronomic areas. Forward and reverse genetic methods are available (Carroll *et al.*, 1985; Hayashi *et al.*, 2008; Lin *et al.*, 2011a). Its genome has been sequenced (Lam *et al.*, 2010; Schmutz *et al.*, 2010), and initial transcriptome atlases have been published (Libault *et al.*, 2010b; Severin *et al.*, 2010). Recently, RNA-seq (Illumina, GAIIx) has been used to investigate nodulation-induced transcriptional changes in the leaf (Reid *et al.*, 2012). In addition, Libault *et al.* (2010a) used microarray chips and RNA-seq (Illumina, GAII) to identify genes regulated in root hairs harvested from entire root systems of nodulating soybean. This study provided insight into differentially expressed nodulation-related genes. However, lowly and transiently expressed transcripts would have been diluted by those in the majority of the non-nodulating part of the root. Moreover, nodulation events from across the entire root system would have been at different developmental stages (Calvert *et al.*, 1984), and genes expressed in root tissues other than the epidermis would not have been captured.

Here, we used soybean and RNA-seq approach based on the Illumina GAIIx platform to focus on the precise portion of the root that is known to respond to rhizobia inoculation, that is, the ZON (Figure 1). This enabled us to capture lowly and transiently expressed genes, in addition to genes having only subtle transcriptional differences. It also enabled us to identify transcriptional changes at a precise moment in nodulation, as opposed to the array of stages found throughout an entire root system. Furthermore, as a control, we used a *nodC*$^{-}$ mutant of *Bradyrhizobium japonicum* that is unable to synthesize NF. This avoided expression responses that were unrelated to nodulation, such as secondary infection effects and responses associated with the presence of the bacteria *per se*.

Results and discussion

Transcriptome sequencing (RNA-seq) output statistics

To generate insight into the transcriptional changes associated with soybean genes during early nodule formation, we performed RNA-seq using the Illumina GAIIx. Root samples from the ZON of soybean plants, grown in 'extended' plastic growth pouches (Figure 1), were analysed 48 h after being inoculated with either wild-type (WT) or near-isogenic *nodC*$^{-}$ mutant strain of *B. japonicum*. Using the incompatible *nodC*$^{-}$ mutant (which cannot synthesize NF owing to the absence of a chitin synthase gene) as a control, genes regulated specifically by NF were identified, while those regulated by other bacterial factors, such as exopolysaccharides or purines, were not captured.

Using 75-bp single-end reads, a total of 29 261 832 (WT-inoculated) and 27 283 006 (*nodC*$^{-}$-inoculated) reads were generated (Table 1). Approximately 50% of the total reads (13 420 549 and 14 464 439, respectively) were selected for further analysis based on a Phred quality score ≥15 (Table 1). Using this stringent filtering method, approximately 8.94 million reads from WT-inoculated roots and 7.86 million reads from *nodC*$^{-}$-inoculated roots aligned to annotated soybean genes (available at Phytozome; http://www.phytozome.net/soybean.php). Of the remaining analysed yet unmapped reads, approximately 5% aligned to the soybean genome. Approximately 3.77 million reads from the WT-inoculated samples and 6.07 million reads from the *nodC*$^{-}$-inoculated samples failed to map entirely.

Gene regulation during nodule initiation

A total of 46 360 genes were expressed (i.e. they had at least one read match) in the WT-inoculated samples compared with 46 165 genes in the *nodC*$^{-}$-inoculated samples (normalized reads ≥5e^{-07} accounting for a minimum of 5-4 reads; Table S1). This accounts for approximately 61% of all annotated soybean genes (including transcript variants) being expressed in both samples.

Using a twofold difference in expression as a minimum cut-off, 2915 differentially expressed genes were identified as being statistically significant according to the Fisher's exact test, $P < 0.05$. Of these genes, 1677 were up-regulated in response to nodule initiation (Table S2), whereas 1238 were down-regulated (Table S3). Of these genes, approximately one-quarter of all up-regulated genes (407 genes) and one-eighth of all down-regulated genes (150 genes) exhibited a greater than fivefold-change in expression. This includes 130 genes that were expressed exclusively in roots inoculated with WT *B. japonicum* and 53 genes expressed solely in roots inoculated with *nodC*$^{-}$ *B. japonicum*. This indicated that NF perception both activated and suppressed gene expression in the ZON.

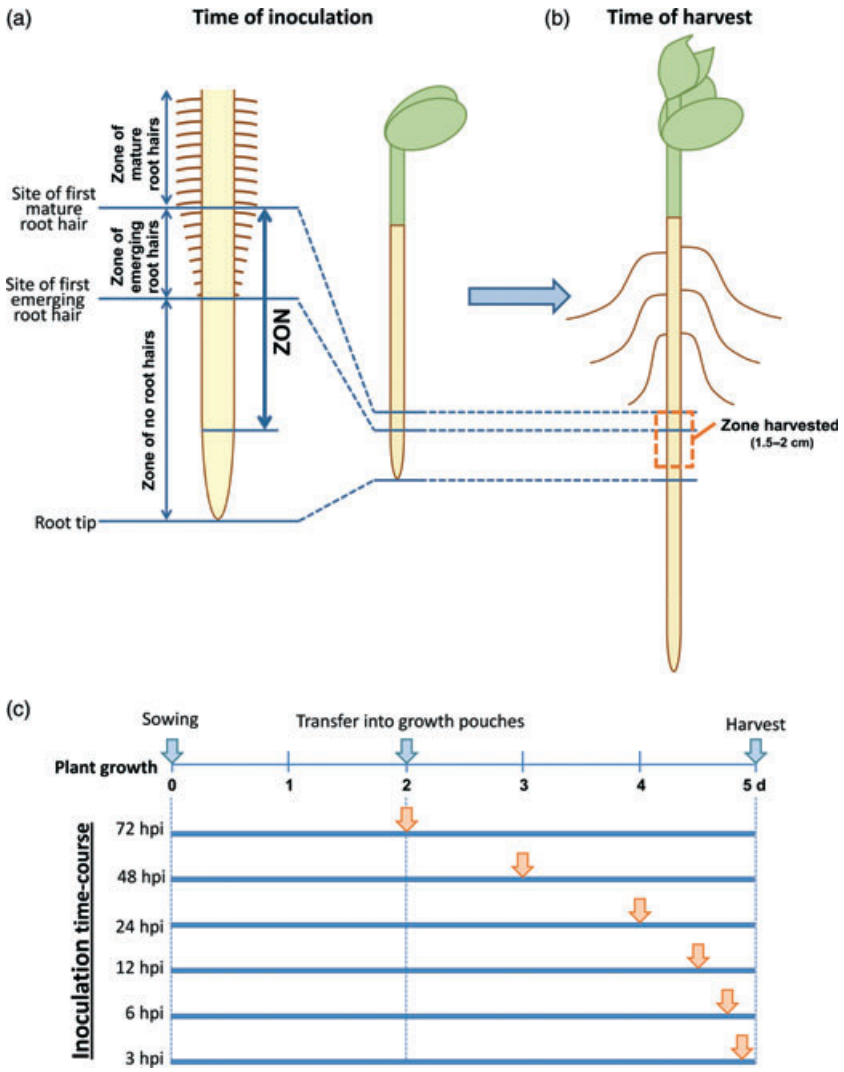

Figure 1 Location of the zone of nodulation (ZON) of soybean seedlings. (a) The ZON was marked on the growth pouch at the time of inoculation with *Bradyrhizobium japonicum* [wild-type (WT) or *nodC*⁻ mutant]. (b) The marked zone was later excised for gene expression analyses using RNA-seq or qRT-PCR. (c) For time-course experiments, the ZON was harvested from plants that were the same age, but that had been inoculated at different time-points. Illustrated is the experimental set-up using 5-day-old seedlings. Orange arrows indicate the time when plants were inoculated with rhizobia to establish an inoculation time-course.

Moreover, using edgeR program, 70% of all differentially expressed genes were identified as having a highly stringent false discovery rate (FDR) of <0.1. Hence, only 10% or less of these genes would be expected to be false positives.

To determine putative gene regulatory networks regulated during nodule initiation, each differentially expressed gene was categorized according to its putative function using Mercator software (available at http://mapman.gabipd.org/web/guest/mapman). According to annotation, each gene was assigned to one of 35 MapMan functional categories ('BINs'; Thimm *et al.*, 2004). Approximately 26% of up-regulated genes and 40% of down-regulated genes were classified as 'not assignable to BINs' (data not shown). The remaining assigned genes were widely distributed across various functional groups (Figure 2). Overall, the distribution of up- and down-regulated genes amongst the functional groups was similar. In both cases, a large proportion of differentially regulated genes were categorized into BINs designated as either 'RNA (including RNA processing, transcription and transcript regulation)', 'protein (including amino acid activation, protein synthesis, targeting, post-translational modification and degradation)' or 'signalling'. These results are consistent with other soybean transcriptome studies reported by Libault *et al.* (2010a) and Brechenmacher *et al.* (2008), and signify that NF perception leads to major metabolic restructuring at multiple levels.

Expression of known nodulation genes in the RNA-seq data set

To ensure that our methods were able to both induce and then capture expression differences caused by inoculation with our WT and *nodC*⁻ mutant *B. japonicum* strains, the expression of known nodulation-related genes was examined in the RNA-seq data set. Several nodulation genes in soybean have been conclusively identified previously, including *GmNFR1* and *GmNFR5* (Indrasumunar *et al.*, 2010, 2011), *GmRIC1*, *GmRIC2* and *GmNIC1* (Reid *et al.*, 2011b), *GmENOD2* (Franssen *et al.*, 1990)

Table 1 Illumina GAIIx RNA-seq output and mapping of 75-bp reads. Values in parentheses represent the percentage of mapped or unmapped reads of all high-quality reads used for mapping

Sample	Total number of reads	High-quality reads (Phred score ≥ 15)			Number of predicted soybean genes expressed
		Matched to annotated soybean genes	Matched to soybean genome	Unmapped to soybean genome	
Roots inoculated with wild-type *Bradyrhizobium japonicum*	29 261 832	8 937 352 (66.6)	715 097 (5.3)	3 768 100 (28.1)	46 360
Roots inoculated with *nodC*⁻ *B. japonicum*	27 283 006	7 858 305 (54.3)	537 035 (3.7)	6 069 099 (42.0)	46 165

and *GmENOD40a* and *GmENOD40b* (Kouchi and Hata, 1993). Moreover, a number of nodulation-related genes have been identified in other legume species, such as *Lotus japonicus* or *Medicago truncatula*, but have not yet been characterized in soybean (reviewed in Ferguson *et al.*, 2010). This includes genes encoding transcription factors (*NIN*, *NSP1 and NSP2*, *ERN1* and *HAP2-1*), a putative cytokinin receptor (*HK1*) and a cytokinin response regulator (*RR5*). Therefore, to increase the number of nodulation genes analysed, soybean orthologues of these genes were identified based on the sequence similarity using the known *L. japonicus* or *M. truncatula* sequences and performing a BLAST search of the soybean genome (Phytozome; http://www.phytozome.net/soybean.php). In soybean, over 75% of genes are predicted to be duplicated (i.e. have a homeologous copy) because of ancestral duplication events (Schmutz *et al.*, 2010). Thus, using the same sequence similarity approach, the homeologue of each soybean nodulation gene was also identified. The Phytozome gene ID for each soybean gene reported here, including its homeologue, is provided in Table 2.

Interestingly, four genes were identified as being similar to *LjNIN* and *MtNIN* (*GmNIN1a, GmNIN1b, GmNIN2a* and *GmNIN2b*). In comparison with the amino acid sequences of LjNIN and MtNIN, GmNIN1a and GmNIN1b share 68%–74% similarity, whereas GmNIN2a and GmNIN2b only share 34%–57% similarity.

All genes previously reported to be up-regulated in early nodulation were indeed found to be up-regulated (≥twofold) in our WT-inoculated samples compared with the *nodC*⁻-inoculated samples. Hence, these genes were induced following the onset of nodule development. This includes the expression of *GmENOD40a* and *GmENOD40b, GmENOD2, GmRIC1* and the orthologues of *L. japonicus* and/or *M. truncatula NIN*, *NSP1, NSP2, ERN1, HAP2-1* and *RR5* (Table 2). Interestingly, all *GmNIN* genes except *GmNIN1b* were up-regulated (≥twofold) in response to WT inoculation. *GmENOD40a, GmENOD40b, GmNSP2b* and *GmERN1b* and *GmRIC1a* exhibited the greatest difference in expression, with most exhibiting >10-fold up-regulation in WT-inoculated samples. In fact, *GmRIC1a* expression was only detected in WT-inoculated samples.

In contrast, small to no differences in expression (<twofold) were observed for the Nod factor receptor genes (*GmNFR1α, GmNFR1β, GmNFR5α* and *GmNFR5β*), *GmNORKa* and *GmNORKb*, *GmNIN1b*, *GmNSP1b* and the two putative cytokinin receptors, *GmHK1a* and *GmHK1b*. Intriguingly, despite their high sequence similarities (>85% at the nucleotide level), most homeologous genes exhibited different expression responses to *B. japonicum* (Table 2), possibly due to functional/cell-type-specific divergence and an early onset of neodiversification.

Interestingly, the expression of *leghaemoglobin C2* (*GmLbc*$_2$; Wiborg *et al.*, 1982), which was previously thought to act at a later stage of nodulation (Sato *et al.*, 2001), was up-regulated in WT-inoculated samples as early as 48 h after inoculation (Table 2). This suggests that the oxygen transport capability of leghaemoglobin may be used prior to the onset of nitrogen fixation. Like *GmRIC1a*, there was no *GmLbc*$_2$ expression detected in the *nodC*⁻-inoculated samples (data not shown). In contrast, transcripts of all other previously known nodulation genes and nodulation-related gene orthologues were detected at trace levels in *nodC*⁻-inoculated samples (Table 2). This may indicate that only a few of these genes are completely nodulation specific.

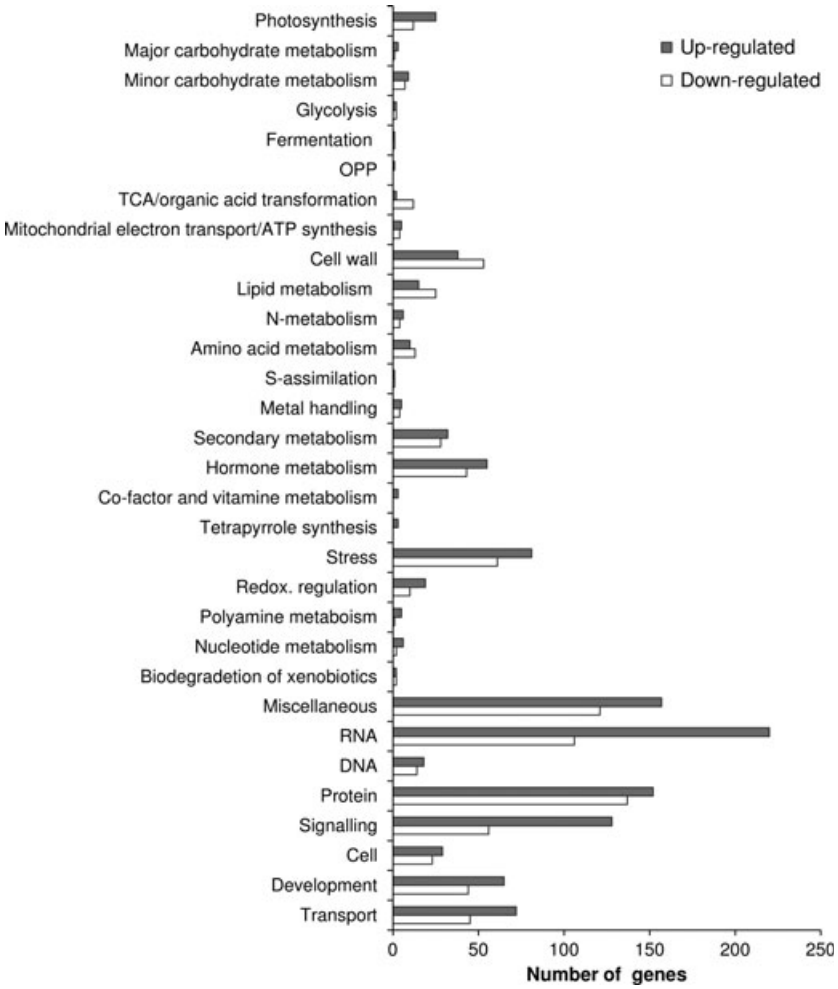

Figure 2 Functional distribution of differentially expressed soybean genes using RNA-seq. All genes regulated in transcript abundance during early nodulation (48 h p.i.) were assigned to one of 35 functional categories using Mercator software. Bars represent the number of up-regulated (shaded) and down-regulated (blank) genes responding to wild-type compared with *nodC*⁻ mutant *Bradyrhizobium japonicum* inoculation.

Confirming the expression of known nodulation-related genes and nodulation-related gene orthologues during soybean nodule initiation

To confirm the regulation of known soybean nodulation–related genes and the soybean orthologues of *L. japonicus* and/or *M. truncatula* nodulation-related genes, qRT-PCR was performed using the same ZON samples used for RNA-seq analysis (48 h p.i.). Additional samples harvested as part of the same study at 3, 12, 24 and 72 h p.i. were also examined to determine the expression of these genes at different stages of early nodule development. The genes investigated include *GmENOD40a, GmENOD40b*, *GmRIC1a*, *GmNIN1*a, *GmNIN1b*, *GmNIN2a, GmNIN2b*, *GmNSP1a*, *GmNSP1b*, *GmNSP2a*, *GmNSP2b, GmERN1a* and *GmERN1b* (Figure 3).

With the exception of *GmNIN1b* and *GmNSP1b*, all of the nodulation-related genes examined here exhibited increased expression in WT-inoculated samples. This is consistent with the RNA-seq results. *GmENOD40a, GmENOD40b* and *GmRIC1a* expression increased dramatically over the course of a 72-h study (Figure 3), which is consistent with the timing for the onset of cell divisions in both the cortex and pericycle (Calvert *et al.*, 1984). The transcription factor, *GmNIN1a*, exhibited a similar pattern of expression, whereas the other differentially expressed transcription factors analysed (i.e. other *GmNIN*s, *GmNSP*s and the *GmERN*s) were strongly up-regulated very early (as early as 3 h p.i. for *GmERN1b*) and then began to decline (Figure 3). This decline may relate to an inhibition of nodule development because of the onset of AON (Calvert *et al.*, 1984; Delves *et al.*, 1986; Mathews *et al.*, 1989). Alternatively, it could reflect a requirement for these genes in early nodule development, followed by a decline in need, and hence expression, as nodule primordia mature.

The fact that *GmNIN1b* and *GmNSP1b* were not differentially regulated (data not shown) could indicate that changes in their expression are not required for early nodule development, or that these genes are simply not involved in nodule formation altogether. Similar expression patterns were reported for these genes in *B. japonicum* -inoculated soybean root hairs (Libault *et al.*, 2010a).

Regulation of functional categories and biological pathways in early nodule development

Several biological pathways and categories were differentially regulated during the early stages of nodulation as visualized by MapMan software (Thimm *et al.*, 2004). A selection of diagrams illustrating these pathways and categories is provided

Table 2 Expression fold-changes in soybean nodulation genes/nodulation gene orthologues and their homeologous genes (RNA-seq)

Gene name	Gene ID	Number of reads		Normalized expression fold-change	References	Note
		Wild-type inoculation	*nodC*⁻ inoculation			
GmNFR1α	Glyma02g43860	127	64	1.57	Indrasumunar *et al.* (2010)	
GmNFR1β	Glyma14g05060	8	10	0.63	Indrasumunar *et al.* (2010)	
GmNFR5α	Glyma11g06740	154	76	1.61	Indrasumunar *et al.* (2010)	
GmNFR5β	Glyma01g38560	83	45	1.46	Indrasumunar *et al.* (2010)	
GmNORKα	Glyma01g02460	538	284	1.50		
GmNORKβ	Glyma09g33510	916	426	1.71		
GmNIN-1a	Glyma04g00210	225	59	3.03		
GmNIN-1b	Glyma06g00240	5	3	1.32		
GmNIN-2a	Glyma02g48080	180	32	4.47**		
GmNIN-2b	Glyma14g00470	112	34	2.62**		
GmNSP1a	Glyma16g01020	15	6	1.98		
GmNSP1b	Glyma07g04430	44	27	1.29		
GmNSP2a	Glyma04g43090	22	7	2.50**		
GmNSP2b	Glyma06g11610	12	1	9.51**		
GmERN1a	Glyma19g29000	83	26	2.53**		
GmERN1b	Glyma16g04410	32	2	12.73**		
GmHAP2-1a	Glyma10g10240	108	13	6.59**		
GmHAP2-1b	Glyma02g35190	283	89	2.51**		
GmHK1a	Glyma08g05370	1248	738	1.34		
GmHK1b	Glyma05g34310	730	530	1.09		
GmRR5a	Glyma17g10170	75	27	2.20**		
GmRR5b	Glyma05g01730	62	27	1.82		
GmENOD40a	Glyma01g03470	1121	90	9.85**	Kouchi and Hata (1993)	
GmENOD40b	Glyma02g04180	415	35	9.38**	Kouchi and Hata (1993)	
GmENOD2	Glyma08g14020	40	5	6.36**	Franssen *et al.* (1990)	No homeologue
GmRIC1a	Glyma13g36830	23	0	$2.64e^{+09}$**	Reid *et al.* (2011b)	
GmRIC1b	Gm12:36892757..36893045		NA		Reid *et al.* (2011b)	
GmLbc$_2$a	Glyma20g33290	21	0	$2.47e^{+09}$**	Wiborg *et al.* (1982)	
GmLbc$_2$b	Glyma10g34260	0	0	NA		
GmLbc$_2$c	Glyma10g34280	0	0	NA		
GmLbc$_2$d	Glyma10g34290	0	0	NA		

Asterisks represent statitically significant difference between inoculation treatments (**$P < 0.01$).

in Figure S1. Using a Wilcoxon rank sum test (Benjamini–Hochberg corrected, provided by MapMan), 276 biological categories (BINs/sub-BINs) were identified by their content as being statistically different from all other BINs ($P < 0.05$, Table S4). These categories include polyamine biosynthesis, PR proteins (biotic stress), ATPase transport (mainly up-regulated) and the mevalonate (MVA) pathway (largely down-regulated) (Table S4, Figure S1). A large number of genes involved in signalling, such as those encoding transcription factors and receptor kinases, were also highly up-regulated in the nodulation-induced samples, including those for bHLH, MYB, WRKY, PHOR1, NAC, HSF and NIN-like transcription factors. Genes involved in hormone biosynthesis and responses [i.e. ethylene, brassinosteroids (BR), cytokinins (CK) and gibberellins (GA)] were also differentially regulated (Figure 4a,b). Several components of BR (e.g. *HYD*) and ethylene (e.g. *ACC oxidase*) biosynthesis were significantly down-regulated, whereas putative components of jasmonic acid (e.g. *lipoxygenases* and *OPDA reductase*) biosynthesis were mainly up-regulated. Although not all at statistically significant levels, the expression of a number of GA biosynthesis/catabolism genes, such as *GA 20-oxidase*, *GA 3-oxidase* and *GA 2-oxidase*, was also up-regulated in response to WT *B. japonicum* inoculation (Figure 4b).

Confirmation and further expression analysis of novel early nodulation candidate genes

Up-regulated genes

To confirm the up-regulation of several candidate genes during early nodule development and to examine their expression at different times following nodule initiation, a time-course study using qRT-PCR was performed. Samples were harvested from the ZON (detected microscopically at the time of inoculation) of WT soybean roots inoculated with either WT or *nodC*⁻ *B. japonicum* at 3, 12, 24, 48 and 96 h p.i. The strong up-regulation of known early nodulation genes (*GmENOD40b* and *GmRIC1a*) in response to WT, but not *nodC*⁻, inoculation confirmed that the samples responded appropriately to their inoculation strain (data not shown). These genes were strongly up-regulated between 12 and 48 h p.i., followed by a considerable decrease

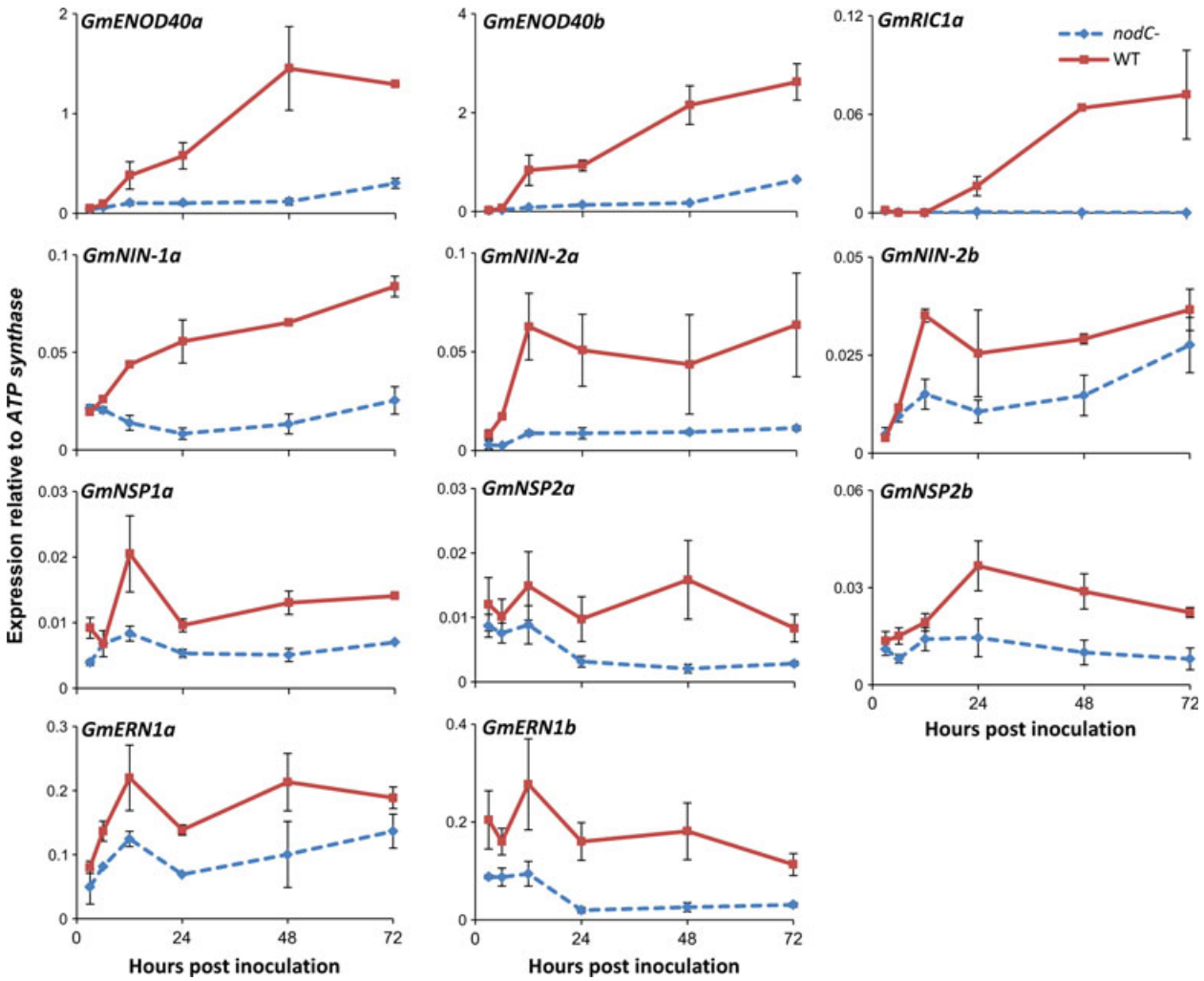

Figure 3 Expression of known soybean nodulation genes and soybean nodulation gene orthologues during early stages of nodulation. Transcript abundance was analysed by qRT-PCR using RNA isolated from *Bradyrhizobium japonicum*-inoculated soybean root tissue at 3, 6, 12, 24, 48 and 72 h p.i. The roots were inoculated with either wild-type or *nodC*⁻ mutant *B. japonicum*. Error bars indicate the standard error of the mean resulting from two independent experiments. Roots from six to eight plants were used for each experiment.

in expression to *nodC*⁻-inoculated levels by 96 h p.i. This transient increase in expression likely correlates with the stage of nodule primordium development (Lim *et al.*, 2011).

Nine genes up-regulated by WT inoculation in the RNA-seq data set (Table 3) were confirmed to be up-regulated in this 96-h time-course study using qRT-PCR (Figure 5). Like most soybean genes, a number of these candidates have a homeologous copy/copies. Henceforth, the candidate genes are referred to as '*gene a*' and their homeologues as '*gene b*', '*gene c*', *etc*.

Two genes involved in gibberellin biosynthesis, *GA 20-oxidase a* (*GA20ox a*; Glyma04g42300) and *GA 3-oxidase 1a* (*GA3ox 1a*; Glyma15g01500), were highly up-regulated in the WT-inoculated samples. These genes exhibited a similar pattern of expression, being strongly up-regulated at 12 h p.i. and then exhibiting a strong decrease thereafter (Figure 5). Consistent with the RNA-seq data set (Table 3), the expression of *GA20ox a* was not detected in any of the *nodC*⁻-inoculated samples, and there was little to no detection of the *GA3ox 1a* (Figure 5).

The up-regulation of genes involved in GA biosynthesis is consistent with previous reports showing that GA levels were increased in developing nodules (reviewed in Ferguson and Mathesius, 2003). Additional studies have also suggested that GAs have a role in both infection thread formation and nodule primordia establishment (Ferguson *et al.*, 2005, 2011; Lievens *et al.*, 2005). Indeed, Ferguson *et al.* (2005, 2011) used pea (*Pisum sativum*) mutants altered in GA biosynthesis and response to demonstrate a direct requirement for GA in nodule development. Lievens *et al.* (2005) also showed that the expression of a *GA20ox* gene in *Sesbania rostrata* was both Nod factor-dependent and transient in an invasion-related manner, accumulating around infection pockets and young infection threads. Inhibitors of GA biosynthesis negatively influence infection thread formation in *Sesbania* root nodulation (Lievens *et al.*, 2005). Fascinatingly, translated products of the *S. rostrata*'s *GA20ox* and the up-regulated soybean *GA20ox a* found in our study share high sequence similarity (75% similarity). This indicates that these two genes may be orthologous to one another and that they likely share similar function during nodulation. Moreover, both of the up-regulated GA biosynthesis genes identified here were also up-regulated in an RNA-seq data set generated using sheared soybean root hairs inoculated with *B. japonicum* (Libault *et al.*,

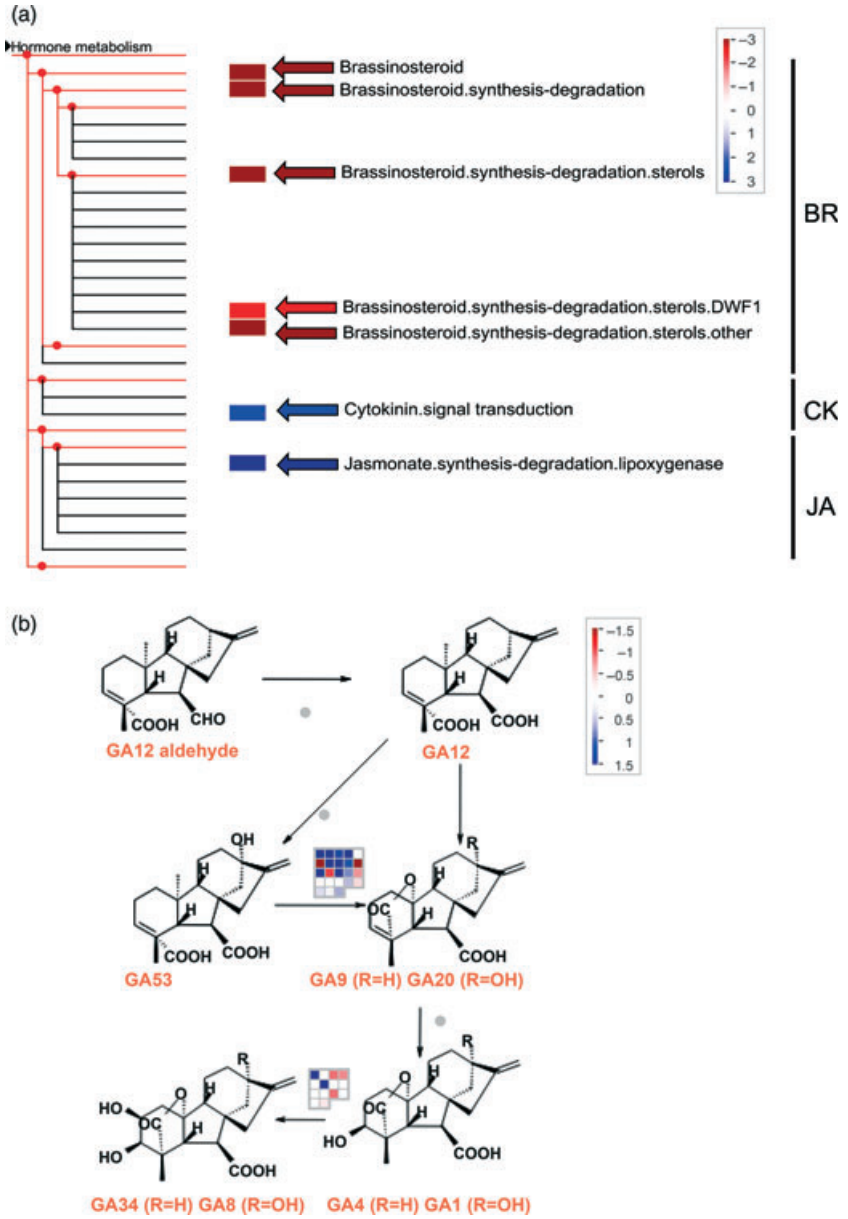

Figure 4 MapMan/PageMan diagrams representing hormonal responses during early soybean nodulation (48 h p.i.) as determined using RNA-seq. (a) An overview of the hormonal response (PageMan), where a Wilcoxon rank sum test (Benjamini–Hochberg corrected) was employed to identify BINs whose contents were differentially regulated. (b) Later stages of the gibberellin biosynthetic pathway (MapMan). Each coloured block represents an individual BIN/sub-BIN or gene in the PageMan or MapMan analysis, respectively. The colour represents the direction and strength of their regulation (Logarithmic colour scale; red, highly down-regulated and blue, highly up-regulated).

2010a). This further suggests that GAs have a significant role in rhizobia infection, such as root hair invasion and infection thread formation.

A member of the large cytochrome P450 gene family, *cytochrome P450* (Glyma11g37110), exhibited a >30-fold increase in expression 48 h p.i. in WT inoculated samples (Figure 5). This level of expression remained increased but declined to about half of that level by 96 h p.i. (Figure 5). In contrast, its expression was nearly undetectable in *nodC*$^{-}$-inoculated samples. *Cytochrome P450* genes typically encode enzymes that catalyse the oxidation of organic substances.

A gene encoding an endo-1,4-β-glucanase (*endo-1,4-β-glucanase a*; Glyma18g03470) was significantly up-regulated at just 12 h p.i. (Figure 5). By 48 h p.i., it exhibited a marked 140-fold increase in expression. Endo-1,4-β-glucanases, also known as cellulases, are typically involved in the hydrolysis of cellulose. Thus, it is tempting to speculate that this gene has roles in root hair invasion and possibly in infection thread development and progression through the root cortex.

Two of the candidates encode products that typically function in signal transduction, including a putative TIR-NBS-LRR protein (*TIR-NBS-LRR*; Glyma12g03040) and a putative histidine kinase (*histidine kinase a*; Glyma04g06190). Both were highly up-regulated in samples induced to form nodules, peaking at 24 h p.i., with a striking 194-fold increase for the *TIR-NBS-LRR* and a 13-fold increase for the *histidine kinase a* (Figure 5). RNA levels of

Table 3 Expression of up-regulated nodulation candidate genes and their homeologous gene copies at 48 h p.i. (RNA-seq)

Gene function/category	Gene ID	Number of reads		Normalized expression fold-change	Biological pathway/function	Note
		Wild-type inoculation	*nodC*⁻ inoculation			
GA 20-oxidase a	**Glyma04g42300**	**12**	**0**	**1.38e^{+09}****	Gibberellin biosynthesis	
GA 20-oxidase b	Glyma06g12510	3	0	28.36	Gibberellin biosynthesis	
GA 3-oxidase 1a	**Glyma15g01500**	**117**	**8**	**11.60****	Gibberellin biosynthesis	
GA 3-oxidase 1b	Glyma13g43850	57	25	0.85	Gibberellin biosynthesis	
Cytochrome P450	**Glyma11g37110**	**242**	**2**	**95.97****	Oxidase	No homeologue
Endo-1,4-β-glucanase a	**Glyma18g03470**	**281**	**7**	**31.84****	Cellulase	
Endo-1,4-β-glucanase b	Glyma11g34850	0	0	NA		
Histidine kinase a	**Glyma04g06190**	**18**	**3**	**4.76****	Signalling	
Histidine kinase b	Glyma06g06180	6	1	2.25	Signalling	
TIR-NBS-LRR	**Glyma12g03040**	**12**	**0**	**1.38e^{+09}****	Signalling	No homeologue
β-expansin a	**Glyma11g17160**	**44**	**6**	**5.82****	Cell wall loosening	
β-expansin b	Glyma01g16140	60	14	1.77	Cell wall loosening	
MPBQ/MSBQ methyltransferase	**Glyma02g00440**	**10**	**3**	**2.64**	Tocopherol biosynthesis	No homeologue
ABC transporter a	**Glyma02g34070**	**34**	**3**	**8.99****	Transporter	
ABC transporter b	Glyma10g11000	22	3	2.54**	Transporter	

Asterisks represent statistically significant differences between inoculation treatments (**$P < 0.01$). Data for the candidate genes are represented in boldface.

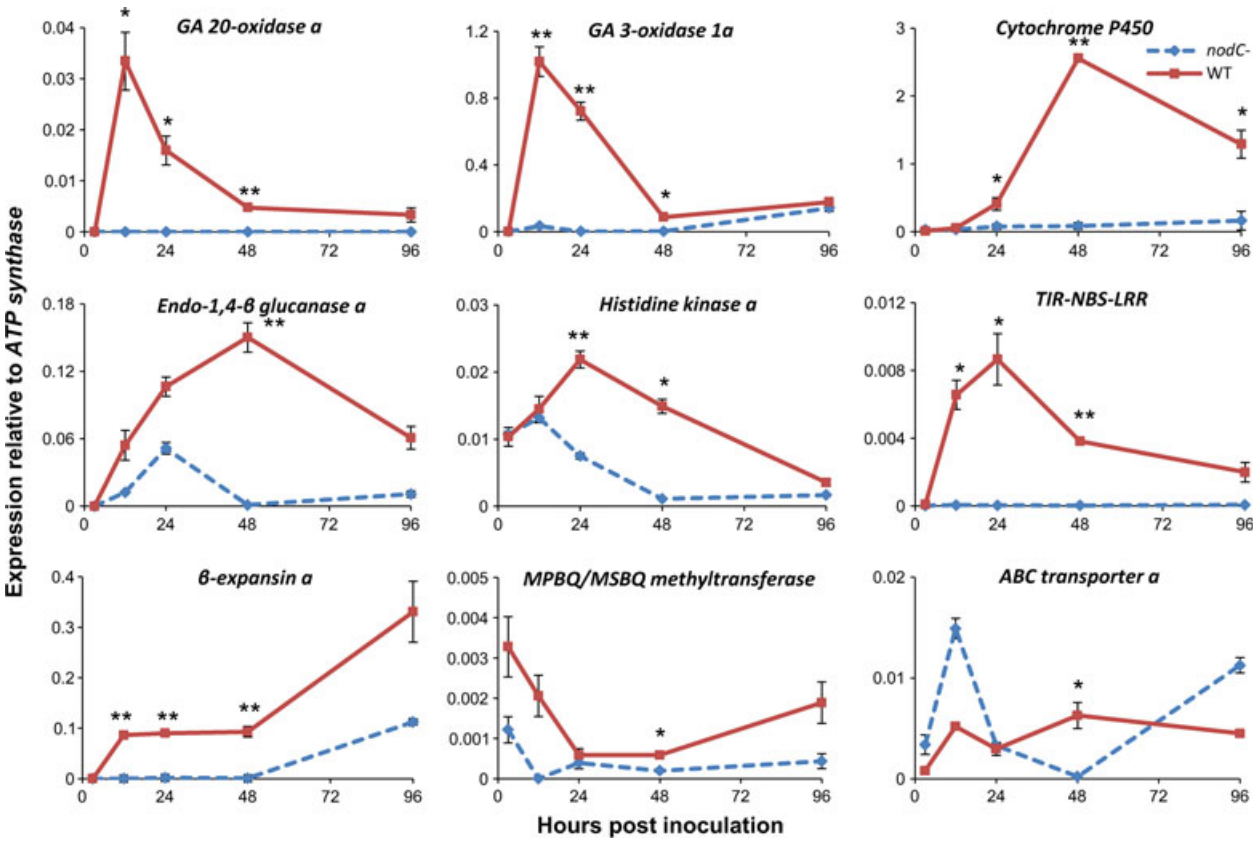

Figure 5 Expression of up-regulated nodulation gene candidates during the early stages of soybean nodulation. Transcript abundance was analysed by qRT-PCR using RNA isolated from *Bradyrhizobium japonicum*-inoculated soybean root tissues at 3, 12, 24, 48 and 96 h p.i. The roots were inoculated with either wild-type or *nodC*⁻ mutant *B. japonicum*. Error bars indicate the standard error of the mean resulting from three biological replicates, each consisting of roots (zone of nodulation) from four to six plants. Asterisks represent statistically significant differences between inoculation treatments at the same time-point (Student's *t*-test; *$P < 0.05$ and **$P < 0.01$).

each decreased considerably thereafter. *TIR-NBS-LRR* expression was nearly undetectable in *nodC*⁻-inoculated samples, whereas *histidine kinase a* was expressed in both WT and *nodC*⁻-inoculated samples. Because these genes exhibited a similar pattern of expression, it is possible that they function in the same pathway. TIR-NBS-LRR proteins commonly act in pathogen detection and plant disease resistance (Dinesh-Kumar *et al.*, 2000), so this candidate may recognize NF in a chitin-like manner, possibly to suppress pathogen-related defence mechanisms that would otherwise prevent successful *Bradyrhizobium* infection.

Unlike the abovementioned candidates, the expression of a *β-expansin* gene (*β-expansin a*; Glyma11g17160) continued to increase over the course of the 96-h study (Figure 5). Expansins are generally associated with cell wall growth, expansion and loosening (Cosgrove, 2000); thus, this gene may function in invasion, infection thread progression or even cell division. In contrast, the expression of a MPBQ/MSBQ methyltransferase gene (*MPBQ/MSBQ methyltransferase*; Glyma11g17160) was greatest at 3 h p.i. (Figure 5). Similar genes act in the biosynthesis of compounds such as tocopherol, commonly known as vitamin E, is thought to act in stress tolerance (reviewed in Munné-Bosch, 2005). Interestingly, an ABC transporter gene (*ABC transporter a*; Glyma02g34070) that was up-regulated in the 48 h p.i. RNA-seq data set was also up-regulated 48 h p.i. in this qRT-PCR time-course study, yet was not up-regulated at any other time-point examined (Figure 5), indicating tight control of its expression pattern.

Down-regulated genes

Approximately half of the differentially regulated genes identified via RNA-seq were down-regulated. Nine of these candidates (Table 4) were investigated further using qRT-PCR and the 96-h time-course study. These genes were selected either because they exhibited the strongest level of down-regulation or because of their putative function and possible role in nodulation. The candidates include ACC oxidase (*ACC oxidase a;* Glyma14g05350), pectinesterase (*pectinesterase a;* Glyma09g09050), GA 3-oxidase (*GA3ox 2a;* Glyma04g07520), protein of unknown function (*unknown;* Glyma01g17590), *MVA kinase a* (*MVA kinase a*; Glyma03g39890), AP2-EREBP-type transcription factor (*AP2-EREBP a*; Glyma05g32040), KH domain containing protein (*KH domain containing protein a*; Glyma06g09460), γ-tocopherol methyltransferase (*γ-tocopherol methyltransferase a*; Glyma12g01680) and HYD1 (*HYD1 a*; Glyma09g31870). These genes were all confirmed to be down-regulated in samples inoculated with WT *B. japonicum* compared with those inoculated with the *nodC*⁻ mutant (Figure 6). Unlike the up-regulated genes investigated here, however, it is difficult to assign biological relevance to these genes because the level to which they were down-regulated was minimal and often not significant (Figure 6).

In contrast to the abovementioned genes, biological relevance can putatively be assigned to a down-regulated *GA 3-oxidase* gene (Figure 6). As opposed to the two up-regulated GA biosynthesis genes described earlier, this *GA3ox 2a* gene was not identified as being differentially regulated in the soybean root hair RNA-seq data set. This suggested that it may act in the root itself. The expression differences observed support previous findings that GA levels need to be tightly regulated during nodule development, both transiently and spatially, as levels that are either too high or too low impair proper nodule formation (Ferguson *et al.*, 2005, 2011).

Expression of candidate gene homeologous genes

The expression of homeologous copies of the candidate genes was examined using the RNA-seq data set. Of the 18 candidates examined, only *cytochrome P450*, *TIR-NBS-LRR*, *MPBQ/MSBQ*

Table 4 Expression of down-regulated nodulation candidate genes and their homeologous gene copies at 48 h p.i. (RNA-seq)

Gene function/category	Gene ID	Number of reads		Normalized expression fold-change	Biological pathway/function	Note
		Wild-type inoculation	*nodC*⁻ inoculation			
ACC oxidase a	**Glyma14g05350**	**45**	**118**	**−3.31****	Ethylene biosynthesis	
ACC oxidase b	Glyma14g05360	104	130	−0.66	Ethylene biosynthesis	
Pectinesterase a	**Glyma09g09050**	**3**	**16**	**−6.72****	Cell wall modification	
Pectinesterase b	Glyma09g08960	0	0	0.00	Cell wall modification	
GA 3-oxidase 2a	**Glyma04g07520**	**2**	**11**	**−6.93***	Gibberellin biosynthesis	
GA 3-oxidase 2b	Glyma06g07630	0	1	−27.11	Gibberellin biosynthesis	
Unknown	**Glyma01g17590**	**3**	**17**	**−7.14****	Unknown	No homeologue
MVA kinase a	**Glyma03g39890**	**52**	**118**	**−2.86****	Mevalonate (MVA) pathway	
MVA kinase b	Glyma19g42440	36	42	−0.56	MVA pathway	
AP2-EREBP a	**Glyma05g32040**	**1**	**16**	**−20.11****	Transcription factor	
AP2-EREBP b	Glyma08g15350	4	5	−0.66	Transcription factor	
KH domain containing protein a	**Glyma06g09460**	**54**	**184**	**−4.30****	Unknown	
KH domain containing protein b	Glyma04g09300	183	184	−0.34	Unknown	
γ-Tocopherol methyltransferase a	**Glyma12g01680**	**11**	**27**	**−3.09****	Tocopherol biosynthesis	
γ-Tocopherol ***methyltransferase*** *b*	Glyma12g01690	112	71	0.32	Tocopherol biosynthesis	
HYD1 a	**Glyma09g31870**	**173**	**335**	**−2.44****	Brassinosteroid biosynthesis	
HYD1 b	Glyma09g31880	2	5	−1.66	Brassinosteroid biosynthesis	
HYD1 c	Glyma09g31890	32	34	−0.42	Brassinosteroid biosynthesis	

Asterisks represent statistically significant differences between inoculation treatments (*$P < 0.05$, **$P < 0.01$). Data for the candidate genes are represented in boldface.

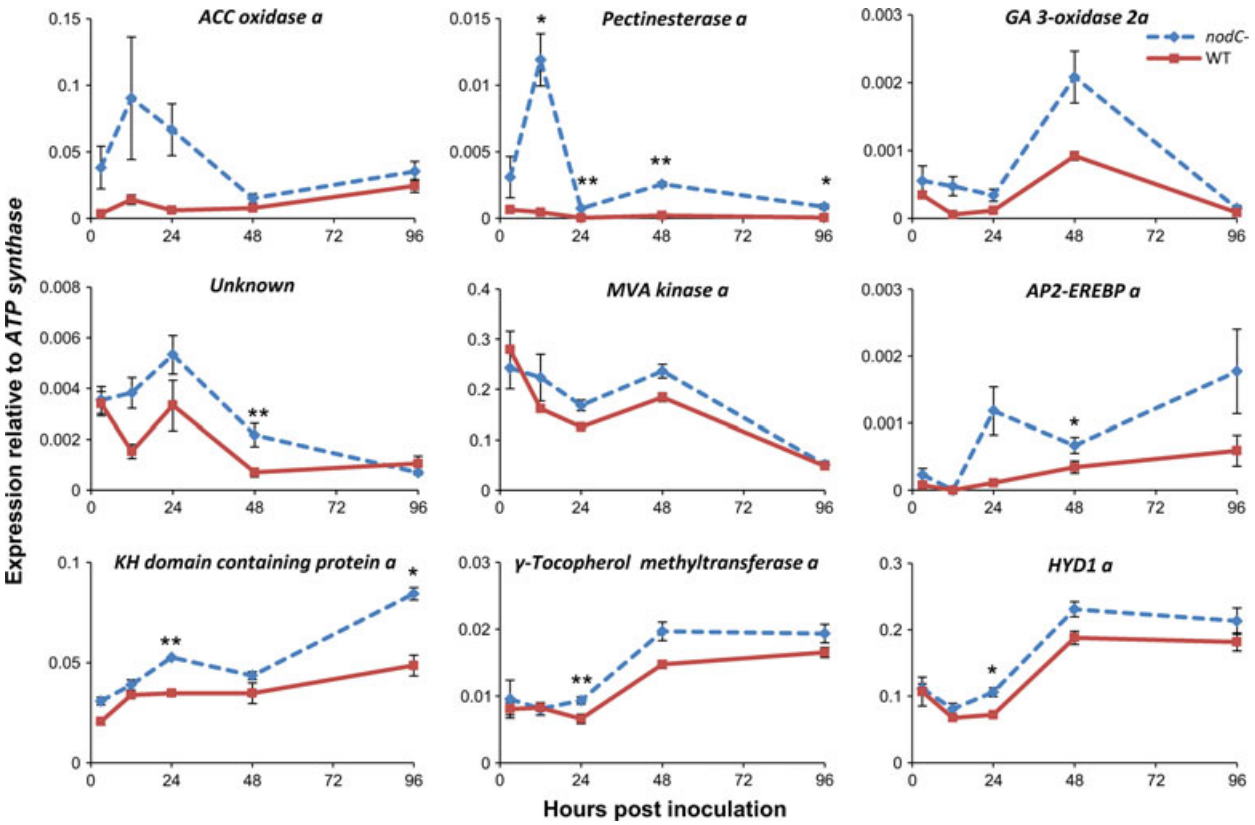

Figure 6 Expression of down-regulated nodulation gene candidates at the early stages of soybean nodulation. Transcript abundance was analysed by qRT-PCR using RNA isolated from *Bradyrhizobium japonicum*-inoculated soybean root tissues at 3, 12, 24, 48 and 96 h p.i. The roots were inoculated with either wild-type or *nodC*⁻ mutant *B. japonicum*. Error bars indicate the standard error of the mean resulting from three biological replicates, each consisting of roots (zone of nodulation) from four to six plants. Asterisks represent statistically significant differences between inoculation treatments at the same time-point (Student's t-test; $*P < 0.05$ and $**P < 0.01$).

methyltransferase and *unknown* did not have at least one homeologous copy. Despite having high sequence similarities, no significant differences in expression were observed for most homeologous genes (Tables 3 and 4). Moreover, although *endo-1,4-β-glucanase a* and *pectinesterase a* candidate genes exhibited a significant number of reads, expression of their homeologous genes was not detected in either sample. This may be due to functional divergence of the homeologous genes as part of neodiversification. Indeed, many of the candidates could have specialized functions in nodulation, or other processes, that are not shared by their homeologous copy.

Expression of candidate genes in a range of soybean tissues

To better characterize the expression of the eight up-regulated candidate genes across the whole plant, their transcripts were quantified in different soybean tissues using qRT-PCR. The location of expression varied significantly amongst the genes (Figure 7). Some were expressed strongly in a particular tissue, whereas others were expressed at similar levels across several tissues.

The *GA20ox a*, *cytochrome P450*, *endo-1,4,-β glucanase a*, *TIR-NBS-LRR* and *β-expansin a* genes exhibited significantly higher expression in nodules compared to other parts of the plant examined here (Figure 7). This further indicates that they are predominantly (if not entirely in some cases) nodulation specific. Of these genes, *GA20ox a* showed the highest level of expression in young developing nodules (1 week old) and diminished in more mature nodules (3 weeks old). In contrast, the expression of *cytochrome P450*, *TIR-NBS-LRR* and *β-expansin a* was highest in 3-week-old nodules, while *endo-1,4,-β glucanase a* expression remained similarly increased in 1- and 3-week-old nodules. The timing of expression could be indicative of the gene's role in nodulation, with some being required for nodule establishment and development and others being required for nodule maintenance or plant–microbe perception and/or communication.

GA3ox 1a, *histidine kinase a* and *MPBQ/MSBQ methyltransferase* were not found to be nodulation specific in their expression (Figure 7). Thus, despite being up-regulated in the ZON following nodule initiation, these genes may have more general roles in the plant, such as in cell division and elongation or bacteria perception and response.

Of the two GA biosynthesis genes, *GA3ox 1a* was expressed in all tissues tested, whereas *GA20ox a* was almost exclusively detected in young, developing nodules (Figure 7). *GA 20-oxidase* acts downstream of *GA 3-oxidase* in the GA biosynthetic pathway (Ferguson *et al.*, 2011), indicating that *GA20ox a* may play a more critical and highly specified role in regulating the production of the hormone during nodule development.

Localization of candidate gene expression in roots

By comparing the genes identified in our RNA-seq data set with those identified in the data set generated using soybean root hairs (Libault *et al.*, 2010a), we can begin to establish which

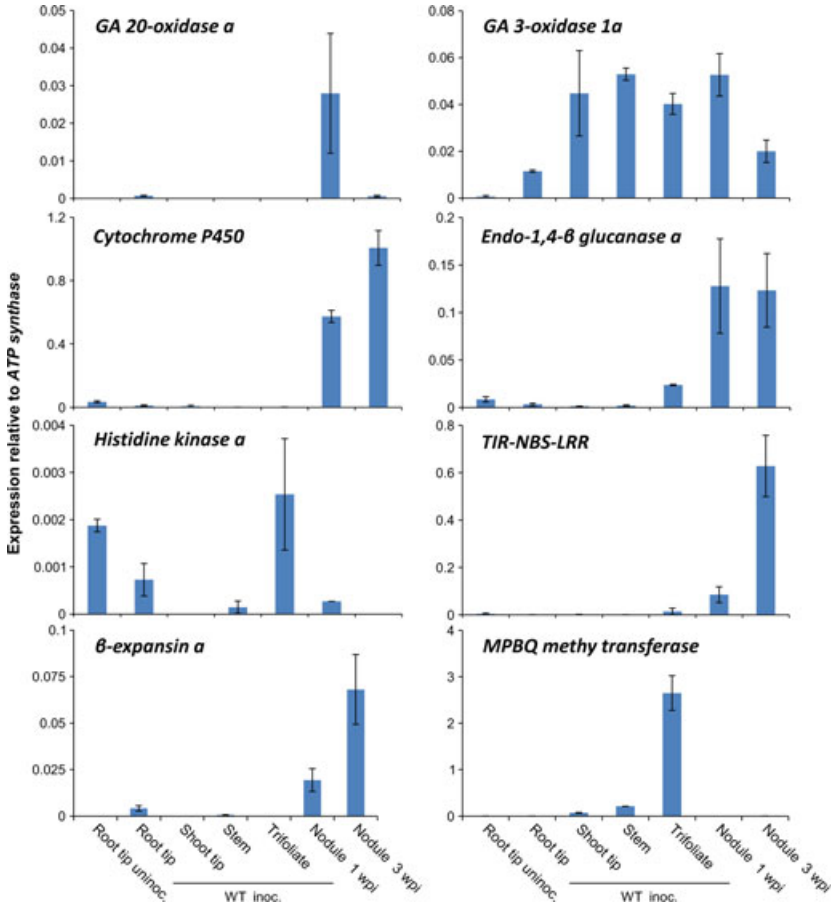

Figure 7 Expression of up-regulated nodulation gene candidates during the early stages of soybean nodulation in various soybean tissues. Transcript abundance was determined by qRT-PCR using RNA isolated from several tissues of soybean plants that were either inoculated with compatible wild-type *Bradyrhizobium japonicum* or were uninoculated. Error bars indicate the standard error of the mean resulting from three biological replicates, each consiting of tissues from two to four plants.

genes may function in the epidermal layer versus those that act in the inner root tissues. This comparison revealed 162 genes that were commonly regulated in the two data sets (112 up-regulated and 50 down-regulated), including the two GA biosynthesis genes, Glyma04g42300 and Glyma15g01500.

The remaining 2793 genes identified in our data set may be regulated in other parts of the root (i.e. cortex, pericycle, vasculature) or in root hairs located specifically in the ZON. The latter could particularly be true for lowly expressed genes whose subtle transcriptional differences are captured in the ZON but are diluted by tissue collected across the entire root system. Indeed, nodule development is presumed to be at different stages throughout the root system (cf., Calvert *et al.*, 1984; Mathews *et al.*, 1989) and mature root hairs, which make up the majority, do not respond to *B. japonicum* inoculation.

In conclusion, using RNA-seq, we generated a transcriptomic data set of genes differentially and transiently expressed in the ZON of soybean roots during nodule initiation. This data set represents a resource for the molecular signalling responses occurring during nodule initiation and early nodule formation. Further analyses are now required to confirm which of the approximately 3000 genes identified are required for early nodule development and how they function in the nodulation pathway.

Experimental procedures

Plant growth conditions

In all experiments conducted, WT soybean (*Glycine max* [L.] Merr. cv. Bragg) was used. For transcriptome sequencing and qRT-PCR inoculation time-course experiments, sterilized seeds were germinated in beakers containing paper towel that had been rolled, moistened and sterilized. The beakers were then placed in the dark at 28 °C for 2 day to encourage germination. The resulting emerging seedlings were then transferred into extended plastic growth pouches (Mega International, West St. Paul, MN) with each pouch [double standard length; 35 cm (L) × 16 cm (W)] containing three seedlings. The pouches were surrounded with cardboard to keep the root system in dark at all times. Both germination and seedling growth were carried out in a plant growth chamber (light/-dark = 16/8 h; T = 28/25 °C and 80% humidity). For qRT-PCR experiments using various tissue types, plants were grown in pots (175 mm diameter) containing sterile grade 2 vermiculite in a temperature-controlled glasshouse (light/-dark = 16 h/8 h, T = 28/23 °C). These plants were watered twice per week with low-nitrogen Herridge solution (Herridge, 1977) containing 0.5 mM KNO_3.

Bradyrhizobium inoculation and plant tissue harvest

To specifically target ZON for harvesting, 2–6-day-old seedlings were observed under a stereoscopic microscope (Nikon, SMZ800, Tokyo, Japan). Three sites along the tap root were marked on the growth pouch: the root tip, the first emerging root hairs and the first mature root hairs. The region of the root system below the first mature root hairs was then inoculated with 500 μL of either *B. japonicum* strain CB1809 (WT) or AN122 (*nodC*⁻ mutant; Nieuwkoop *et al.*, 1987; Sanjuan *et al.*, 1992) grown in Yeast Manitol Broth at 28 °C for 3 day and diluted to OD_{600} = ~0.1.

To establish an inoculation time-course experiment, the plants were inoculated at different times prior to harvest (Figure 1c). The root section (1.5–2 cm) containing the ZERH (between the first emerging root hair and the first mature root hair) at the time of inoculation, which was previously identified under stereomicroscope, was then harvested from 5-day-old (RNA-seq experiment and qRT-PCR analysis of soybean nodulation genes) or 6-day-old plants (qRT-PCR analysis of candidate genes). This ensured that the root samples were taken from plants of the same developmental stage, and thus, only the stage of nodulation varied between treatments. Inoculation of 6-day-old seedlings was conducted similar to that for 5-day-old seedlings (Figure 1c); however, the time-point of inoculation was shifted so that the samples representing the desired inoculation time-course (up to 96 h p.i.) were achieved.

Tissues for qRT-PCR expression analysis were harvested from 14-day-old plants inoculated with *B. japonicum* strain CB1809 at the time of sowing (root tip, shoot tip, stem and trifoliate-leaf samples) or left uninoculated (root tip). Additional nodule tissue samples were harvested from roots inoculated with *B. japonicum* strain CB1809 14 day after sowing. All samples were snap-frozen in liquid nitrogen at the time of harvest and stored at −80 °C.

Illumina sample preparation and sequencing run

Total RNA was extracted (method described in the section below) from roots of six to eight plants inoculated with either WT or *nodC*⁻ mutant *B. japonicum* at 48 h p.i. These samples were subsequently used for library generation by the Australian Genome Research Facility (AGRF, Brisbane, Qld, Australia). The mRNA-seq libraries were constructed following Illumina's standard protocol. Briefly, mRNA was isolated from total RNA using Sera-Mag Magnetic Oligo(dT) Beads. Isolated mRNA was fragmented using divalent cations under increased temperatures followed by ethanol precipitation. The fragmented mRNA was reverse-transcribed into cDNA using Superscript III and random primers. The cDNA was end-repaired to create blunt-ended fragments, and an 'A' base was ligated to the 3′ ends of the fragments to create an 'A' base overhang. Adapters with a 'T' base overhang were ligated to each end of the cDNA fragments. The ligated fragments were run on an agarose gel, and a thin slice of DNA fragments corresponding to approximately 200 bp was excised and purified from the gel. The purified size-selected fragments were enriched by PCR to generate the final library fragments.

Owing to the evolutionary background of their genome, most soybean genes exist in duplicate copies, which shares high sequence homology. To differentiate the reads that arise from one gene to those from the duplicated gene, sequencing was run with a relatively long read length of 75 bp. Furthermore, each library was run on two lanes for greater sensitivity. Sequencing reactions were carried out on the Illumina GAIIx platform.

Read alignment and bioinformatics analysis of Illumina sequencing output

For each sample, approximately 30 million 75-bp reads were returned from the Illumina transcriptome sequencing run. Illumina sequence data were processed using a custom script trimConverter.py to produce reads with a minimum length of 50 bp and Phred quality score 15 or greater at each nucleotide position. The resulting data set was mapped to predicted soybean cDNAs (Phytozome release 4.01; Schmutz *et al.*, 2010) using SOAP 2.20 (Li *et al.*, 2008) with default parameters. The resulting mapping file was processed using another custom script diff_display_predgene.pl to produce a table of gene names, genomic location and number of mapped reads for each experiment. Differentially expressed genes were identified by comparing the number of reads matching each gene between experiments. Reads that did not map to the soybean cDNAs were mapped against the complete soybean genome (Phytozome release 4.0) using SOAP 2.20 and with default parameters.

RNA-seq data that mapped to the annotated genes were normalized according to the method described in the study by Robinson and Oshlack (2010) using the edgeR package of Bioconductor v 2.5 (available at http://www.bioconductor.org/). To screen for genes that were differentially expressed between the two samples, statistical analysis provided by the software was applied. Differentially expressed genes were selected against cut-off values of greater than twofold with Fisher exact *P*-value < 0.05 or FDR <0.05.

Resulting lists of candidate genes were annotated and functionally categorized using Mercator software and allowed for visualization using MapMan software v.3.5.0 (both available at http://mapman.gabipd.org/web/guest/mapman; Thimm *et al.*, 2004). Wilcoxon rank sum test with Benjamini–Hochberg correction provided in MapMan software was applied to isolate BINs that display significant differences in BIN contents compared to all other BINs. BINs with $P < 0.05$ were regarded significant. Over-representation analysis was performed via PageMan tool (Usadel *et al.*, 2006) integrated in MapMan software. Annotations of genes were manually confirmed by performing protein Basic Local Alignment Search Tool (BLASTp) against GenBank database (National Centre for Biotechnology Information) on the predicted amino acid sequences.

RNA extraction and cDNA synthesis

RNA was extracted using TRIzol® Reagent (Invitrogen, Carlsbad, CA) following manufacturer's instruction. To remove contaminating genomic DNA from the sample, approximately 1 μg of RNA was treated with 1 U of DNaseI (Fermentas, Burlington, VT, Canada) at 37 °C for 40–60 min. These reactions were terminated by the addition of 1 μL of 25 mM EDTA (Invitrogen) and incubation at 65 °C for 10 min. cDNA synthesis was carried out in 20-μL reaction mixture containing approximately 0.5 μg of DNase-treated RNA, 1 μL of 50 μM oligo(dT) primers, 0.5 mM deoxynucleoside triphosphates (dNTPs), 5× first-strand buffer (Invitrogen), 5 mM dithiothreitol (DTT), 40 U of RNaseOUT™ (Invitrogen) and 100 U of SuperScript III® reverse transcriptase (Invitrogen) incubated at 50 °C for an hour. The reaction was verified by PCR using *GmATP synthase* (Glyma20g25920) primers.

Quantitative real-time PCR

Quantitative real-time PCR (qRT-PCR) primers were designed using either Primer Express® v. 2.0 (Applied Biosystems, Foster City, CA) or Primer3 (available at http://biotools.umassmed.edu/bioapps/primer3_www.cgi) software. A list of primers used in this study is provided in Table S5. Primers for soybean genes that have been described in the literature were designed based on the DNA sequence available on the GenBank database. In the case of genes reported in other species, but not yet identified in soybean, putative orthologues were first determined by utilizing BLAST against the soybean genome (Phytozome v. 4.0; The United States Department of Energy Joint Genome Institute and Centre for Integrative Genomics; available at http://www.phytozome.net). Primers were then designed for the predicted gene sequences exhibiting the highest sequence similarity to the gene of interest.

Gene expression analysis was carried out similar to that described in the study by Hayashi *et al.* (2008). PCR was carried out using SYBR® Green PCR Master Mix (Applied Biosystems), and the 384-well plates for qRT-PCR analysis were set up using an Eppendorf® epMotion™ 5075 Robotics System. The reactions were run on an ABI Prism® 7900 Sequence Detection System (Applied Biosystems). Each plate contained no template (water) controls. Genomic DNA contamination of the cDNA samples was verified by including the reverse transcription negative (RT–) controls in the qRT-PCR. All qRT-PCR were carried out in duplicate (technical replicates) and run for 40 cycles using an annealing temperature of 60 °C. The qRT-PCR was run using the following cycle conditions: 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min. A dissociation stage of 95 °C for 2 min was added at the end of the cycle in order to assess the specificity of the PCR. PCR efficiency for each sample was calculated using LinRegPCR 7.5 software (Ramakers *et al.*, 2003), and the relative expression for gene of interest was measured against that of *GmATP synthase* (Glyma20g25920).

Inoculation time-course expression analysis of nodulation genes was conducted on two biological replicates, each consisting of the ZON of six to eight plants, while three biological replicates, each consisting of the ZON from four to six plants, were used to investigate the candidate gene expression. Expression of candidate genes in several tissues was studied using three biological replicates, each consisting of tissue from two to four plants.

Acknowledgements

This work was supported by the Australian Research Council Centre of Excellence program and Australian Research Council linkage projects (LP0882095 and LP0989200). Additional support was provided to S. Hayashi by The University of Queensland Research Scholarship. We would like to thank A. McGrath, D. Miller, T. Bruxner and K. Hanson (AGRF, Brisbane) for consultation regarding RNA-seq, the initial bioinformatics work and for GAIIx sequencing, and also B. Simpson (School of Molecular and Microbial Science, The University of Queensland) for help with qRT-PCR. We also thank J. Dalton-Morgan and D. Li for their technical support.

References

Beveridge, C.A., Mathesius, U., Rose, R.J. and Gresshoff, P.M. (2007) Common regulatory themes in meristem development and whole plant homeostasis. *Curr. Opin. Plant Biol.* **10**, 44–51.

Bhuvaneswari, T.V., Turgeon, B.G. and Bauer, W.D. (1980) Early events in the infection of soybean (*Glycine max* L. Merr.) by *Rhizobium japonicum*: 1. Localization of infectible root cells. *Plant Physiol.* **66**, 1027–1031.

Biswas, B., Chan, P.K. and Gresshoff, P.M. (2009) A novel ABA insensitive mutant of *Lotus japonicus* with a wilty phenotype but unaltered nodulation regulation. *Mol. Plant*, **2**, 487–499.

Brechenmacher, L., Kim, M.Y., Benitez, M., Li, M., Joshi, T., Calla, B., Lee, M.P., Libault, M., Vodkin, L.O., Xu, D., Lee, S.H., Clough, S.J. and Stacey, G. (2008) Transcription profiling of soybean nodulation by *Bradyrhizobium japonicum*. *Mol. Plant–Microbe Interact.* **21**, 631–645.

Caetano-Anollés, G. and Gresshoff, P.M. (1991) Plant genetic-control of nodulation. *Annu. Rev. Microbiol.* **45**, 345–382.

Calvert, H.E., Pence, M.K., Pierce, M., Malik, N.S.A. and Bauer, W.D. (1984) Anatomical analysis of the development and distribution of *Rhizobium* infections in soybean roots. *Can. J. Bot.* **62**, 2375–2384.

Carroll, B.J., McNeil, D.L. and Gresshoff, P.M. (1985) Isolation and properties of soybean mutants which nodulate in the presence of high nitrate concentrations. *Proc. Natl Acad. Sci. USA*, **82**, 4162–4166.

Colebatch, G., Desbrosses, G., Ott, T., Krusell, L., Montanari, O., Kloska, S., Kopka, J. and Udvardi, M.K. (2004) Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. *Plant J.* **39**, 487–512.

Cosgrove, D.J. (2000) Loosening of plant cell walls by expansins. *Nature*, **407**, 321–326.

Crespi, M. and Frugier, F. (2008) De novo organ formation from differentiated cells: root nodule organogenesis. *Sci. Signal.* **1**, 1–8.

Delves, A.C., Mathews, A., Day, D.A., Carter, A.S., Carroll, B.J. and Gresshoff, P.M. (1986) Regulation of the soybean-*Rhizobium* nodule symbiosis by shoot and root factors. *Plant Physiol.* **82**, 588–590.

Dénarié, J., Debelle, F. and Prome, J.C. (1996) R*hizobium* lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annu. Rev. Biochem.* **65**, 503–535.

Desbrosses, G.J. and Stougaard, J. (2011) Root nodulation: a paradigm for how plant-microbe symbiosis influences host developmental pathways. *Cell Host Microbe*, **10**, 348–358.

Dinesh-Kumar, S.P., Tham, W.H. and Baker, B.J. (2000) Structure-function analysis of the tobacco mosaic virus resistance gene N. *Proc. Natl Acad. Sci. USA*, **97**, 14789–14794.

El Yahyaoui, F., Küster, H., Ben Amor, B., Hohnjec, N., Pühler, A., Becker, A., Gouzy, J., Vernié, T., Gough, C., Niebel, A., Godiard, L. and Gamas, P. (2004) Expression profiling in *Medicago truncatula* identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program. *Plant Physiol.* **136**, 3159–3176.

Ferguson, B.J. (2012) Rhizobia and legume nodulation genes. In *Brenner's Online Encyclopedia of Genetics*, 2nd edn (Maloy, S. and Hughes, K., eds). Oxford, UK: Elsevier, In press.

Ferguson, B.J. and Gresshoff, P.M. (2009) Soybean as a model legume. *Grain Legumes*, **53**, 7.

Ferguson, B.J. and Mathesius, U. (2003) Signaling interactions during nodule development. *J. Plant Growth Regul.* **22**, 47–72.

Ferguson, B.J., Ross, J.J. and Reid, J.B. (2005) Nodulation phenotypes of gibberellin and brassinosteroid mutants of pea. *Plant Physiol.* **138**, 2396–2405.

Ferguson, B.J., Indrasumunar, A., Hayashi, S., Lin, M.-H., Lin, Y.-H., Reid, D.E. and Gresshoff, P.M. (2010) Molecular analysis of legume nodule development and autoregulation. *J. Integr. Plant Biol.* **52**, 61–76.

Ferguson, B.J., Foo, E., Ross, J.J. and Reid, J.B. (2011) Relationship between gibberellin, ethylene and nodulation in *Pisum sativum*. *New Phytol.* **189**, 829–842.

Franssen, H.J., Thompson, D.V., Idler, K., Kormelink, R., Vankammen, A. and Bisseling, T. (1990) Nucleotide-sequence of two soybean *ENOD2* early nodulin genes encoding Ngm-75. *Plant Mol. Biol.* **14**, 103–106.

Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, J.C., Jaubert, M., Simon, D., Cartieaux, F., Prin, Y., Bena, G., Hannibal, L., Fardoux, J., Kojadinovic, M., Vuillet, L., Lajus, A., Cruveiller, S., Rouy, Z., Mangenot, S., Segurens, B., Dossat, C., Franck, W.L., Chang, W.S., Saunders, E., Bruce, D., Richardson, P., Normand, P., Dreyfus, B., Pignol, D., Stacey, G., Emerich, D., Verméglio, A., Médigue, C. and Sadowsky, M.

(2007) Legumes symbioses: absence of nod genes in photosynthetic bradyrhizobia. *Science*, **316**, 1307–1312.

Gresshoff, P.M. (2012) *Glycine max* (Soybean). In *Brenner's Online Encyclopedia of Genetics*, 2nd edn (Maloy, S. and Hughes, K., eds). Oxford, UK: Elsevier, In press.

Gresshoff, P.M. and Delves, A.C. (1986) Plant genetic approaches to symbiotic nodulation and nitrogen fixation in legumes. In *Plant Gene Research. III. A Genetical Approach to Plant Biochemistry* (Blonstein, A.D. and King, P.J., eds), pp. 156–206. Vienna, Austria: Springer-Verlag.

Han, L., Hanan, J. and Gresshoff, P.M. (2010) Computational Complementation: a modelling approach to study signalling mechanisms during legume autoregulation of nodulation. *PLoS Comput. Biol.* **6**, e1000685.

Hayashi, S., Gresshoff, P.M. and Kinkema, M. (2008) Molecular analysis of lipoxygenases associated with nodule development in soybean. *Mol. Plant–Microbe Interact.* **21**, 843–853.

Herridge, D.F. (1977) Carbon and nitrogen nutrition of two annual legumes. PhD dissertation. Perth, WA, Australia: University of Western Australia.

Indrasumunar, A., Kereszt, A., Searle, I., Miyagi, M., Li, D., Nguyen, C.D.T., Men, A., Carroll, B.J. and Gresshoff, P.M. (2010) Inactivation of duplicated Nod factor receptor 5 (NFR5) genes in recessive loss-of-function non-nodulation mutants of allotetraploid soybean (*Glycine max* L. Merr.). *Plant Cell Physiol.* **51**, 201–214.

Indrasumunar, A., Searle, I., Lin, M.-H., Kereszt, A., Men, A., Carroll, B.J. and Gresshoff, P.M. (2011) Nodulation factor receptor kinase 1a controls nodule organ number in soybean (*Glycine max* L. Merr.). *Plant J.* **65**, 39–50.

Jones, K.M., Sharopova, N., Lohar, D.P., Zhang, J.Q., VandenBosch, K.A. and Walker, G.C. (2008) Differential response of the plant *Medicago truncatula* to its symbiont *Sinorhizobium meliloti* or an exopolysaccharide-deficient mutant. *Proc. Natl Acad. Sci. USA*, **105**, 704–709.

Kinkema, M. and Gresshoff, P.M. (2008) Investigation of downstream signals of the soybean autoregulation of nodulation receptor kinase GmNARK. *Mol. Plant–Microbe Interact.* **21**, 1337–1348.

Kosslak, R.M., Bookland, R., Barkei, J., Paaren, H.E. and Applebaum, E.R. (1987) Induction of *Bradyrhizobium japonicum* common *nod* genes by isoflavones isolated from *Glycine max*. *Proc. Natl Acad. Sci. USA*, **84**, 7428–7432.

Kouchi, M. and Hata, S. (1993) Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. *Mol. Gen. Genet.* **238**, 106–119.

Krusell, L., Madsen, L.H., Sato, S., Aubert, G., Genua, A., Szczyglowski, K., Duc, G., Kaneko, T., Tabata, S., de Bruijn, F., Pajuelo, E., Sandal, N. and Stougaard, J. (2002) Shoot control of root development and nodulation is mediated by a receptor-like kinase. *Nature*, **420**, 422–426.

Lam, H.M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F.L., Li, M.-W., He, W., Qin, N., Wang, B., Li, J., Jian, M., Wang, J., Shao, G., Wang, J., Sun, S.S. and Zhang, G. (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* **42**, 1053–1059.

Leigh, J.A., Signer, E.R. and Walker, G.C. (1985) Exopolysaccharide-deficient mutants of *Rhizobium meliloti* that form ineffective nodules. *Proc. Natl Acad. Sci. USA*, **82**, 6231–6235.

Li, R.Q., Li, Y.R., Kristiansen, K. and Wang, J. (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics*, **24**, 713–714.

Li, D., Kinkema, M. and Gresshoff, P.M. (2009) Autoregulation of nodulation (AON) in *Pisum sativum* (pea) involves signalling events associated with both nodule primordia development and nitrogen fixation. *J. Plant Physiol.* **166**, 955–967.

Libault, M., Farmer, A., Brechenmacher, L., Drnevich, J., Langley, R.J., Bilgin, D.D., Radwan, O., Neece, D.J., Clough, S.J., May, G.D. and Stacey, G. (2010a) Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. *Plant Physiol.* **152**, 541–552.

Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R.J., Franklin, L.D., He, J., Xu, D., May, G. and Stacey, G. (2010b) An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *Plant J.* **63**, 86–99.

Lievens, S., Goormachtig, S. and Holsters, M. (2001) A critical evaluation of differential display as a tool to identify genes involved in legume nodulation: looking back and looking forward. *Nucleic Acids Res.* **29**, 3459–3468.

Lievens, S., Goormachtig, S., Den Herder, J., Capoen, W., Mathis, R., Hedden, P. and Holsters, M. (2005) Gibberellins are involved in nodulation of *Sesbania rostrata*. *Plant Physiol.* **139**, 1366–1379.

Lim, C.W., Lee, Y.W. and Hwang, C.H. (2011) Soybean nodule-enhanced CLE peptides in roots act as signals in GmNARK-mediated nodulation suppression. *Plant Cell Physiol.* **52**, 1613–1627.

Lin, Y.-H., Ferguson, B.J., Kereszt, A. and Gresshoff, P.M. (2010) Suppression of hypernodulation in soybean by a leaf-extracted, NARK- and Nod factor-dependent small molecular fraction. *New Phytol.* **185**, 1074–1086.

Lin, M.-H., Gresshoff, P.M., Indrasumunar, A. and Ferguson, B.J. (2011a) *pHairyRed*: a novel binary vector containing the *DsRed2* reporter gene for visual selection of transgenic hairy roots. *Mol. Plant*, **4**, 537–545.

Lin, Y.-H., Lin, M.-H., Gresshoff, P.M. and Ferguson, B.J. (2011b) An efficient petiole-feeding bioassay for introducing aqueous solutions into dicotyledonous plants. *Nat. Protoc.* **6**, 36–45.

Lohar, D.P., Sharopova, N., Endre, G., Penuela, S., Samac, D., Town, C., Silverstein, K.A.T. and VandenBosch, K.A. (2006) Transcript analysis of early nodulation events in *Medicago truncatula*. *Plant Physiol.* **140**, 221–234.

Maguire, T.L., Grimmond, S., Forrest, A., Iturbe-Ormaetxe, I., Meksem, K. and Gresshoff, P.M. (2002) Tissue-specific gene expression in soybean (*Glycine max*) detected by cDNA microarray analysis. *J. Plant Physiol.* **159**, 1361–1374.

Mathesius, U., Mulers, S., Gao, M., Teplitski, M., Caetano-Anollés, G., Rolfe, B.G. and Bauer, W.D. (2003) Extensive and specific responses of a eukaryote to bacterial quorum-sensing signals. *Proc. Natl Acad. Sci. USA*, **100**, 1444–1449.

Mathews, A., Carroll, B.J. and Gresshoff, P.M. (1989) Development of *Bradyrhizobium* infection in supernodulating and non-nodulating mutants of soybean (*Glycine max* [L.] Merrill). *Protoplasma*, **150**, 40–47.

Mortier, V., Den Herder, G., Whitford, R., Van de Velde, W., Rombauts, S., D'Haeseleer, K., Holsters, M. and Goormachtig, S. (2010) CLE peptides control *Medicago truncatula* nodulation locally and systemically. *Plant Physiol.* **153**, 222–237.

Munné-Bosch, S. (2005) The role of α-tocopherol in plant stress tolerance. *J. Plant Physiol.* **162**, 743–748.

Nieuwkoop, A.J., Banfalvi, Z., Deshmane, N., Gerhold, D., Schell, M.G., Sirotkin, K.M. and Stacey, G. (1987) A locus encoding host range is linked to the common nodulation gene of *Bradyrhizobium japonicum*. *J. Bacteriol.* **169**, 2631–2638.

Nishimura, R., Hayashi, M., Wu, G.J., Kouchi, H., Imaizumi-Anraku, H., Murakami, Y., Kawasaki, S., Akao, S., Ohmori, M., Nagasawa, M., Harada, K. and Kawaguchi, M. (2002) HAR1 mediates systemic regulation of symbiotic organ development. *Nature*, **420**, 426–429.

Okamoto, S., Ohnishi, E., Sato, S., Takahashi, H., Nakazono, M., Tabata, S. and Kawaguchi, M. (2009) Nod factor/nitrate-induced *CLE* genes that drive HAR1-mediated systemic regulation of nodulation. *Plant Cell Physiol.* **50**, 67–77.

Oldroyd, G.E.D. and Downie, J.M. (2008) Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu. Rev. Plant Biol.* **59**, 519–546.

Peters, N.K., Frost, J.W. and Long, S.R. (1986) A plant flavone, luteolin, induces expression of Rhizobium meliloti nodulation genes. *Science*, **233**, 977–980.

Ramakers, C., Ruijter, J.M., Deprez, R.H.L. and Moorman, A.F.M. (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* **339**, 62–66.

Redmond, J.W., Batley, M., Djordjevic, M.A., Innes, R.W., Kuempel, P.L. and Rolfe, B.G. (1986) Flavones induce expression of nodulation genes in *Rhizobium*. *Nature*, **323**, 632–635.

Reid, D.E., Ferguson, B.J., Hayashi, S., Lin, Y.-H. and Gresshoff, P.M. (2011a) Molecular mechanisms controlling legume autoregulation of nodulation. *Ann. Bot.* **108**, 789–795.

Reid, D.E., Ferguson, B.J. and Gresshoff, P.M. (2011b) NARK-dependent nodule regulation is activated in the shoot and root in response to inoculation and nitrate induced CLE peptides of soybean. *Mol. Plant–Microbe Interact.* **24**, 606–618.

Reid, D.E., Hayashi, S., Lorenc, M., Stiller, J., Edwards, D., Gresshoff, P.M. and Ferguson, B.J. (2012) Identification of systemic responses in soybean nodulation by xylem sap feeding and complete transcriptome sequencing reveal a novel component of the autoregulation pathway. *Plant Biotechnol. J.* **10**, 680–689.

Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25.

Sanjuan, J., Carlson, R.W., Spaink, H.P., Bhat, U.R., Barbour, W.M., Glushka, J. and Stacey, G. (1992) A 2-O-methylfucose moiety is present in the lipo-oligosaccharide nodulation signal of *Bradyrhizobium japonicum*. *Proc. Natl Acad. Sci. USA*, **89**, 8789–8793.

Sato, T., Onoma, N., Fujikake, H., Ohtake, N., Sueyoshi, K. and Ohyama, T. (2001) Changes in four leghemoglobin components in nodules of hypernodulating soybean (*Glycine max* L. Merr.) mutant and its parent in the early nodule developmental stage. *Plant Soil*, **237**, 129–135.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C. and Jackson, S.A. (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178–183.

Schnabel, E., Journet, E.P., 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 Mol. Biol.* **58**, 809–822.

Searle, I.R., Men, A.E., Laniya, T.S., Buzas, D.M., Iturbe-Ormaetxe, I., Carroll, B.J. and Gresshoff, P.M. (2003) Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase. *Science*, **299**, 109–112.

Severin, A.J., Woody, J.L., Bolon, Y.T., Joseph, B., Diers, B.W., Farmer, A.D., Muehlbauer, G.J., Nelson, R.T., Grant, D., Specht, J.E., Graham, M.A., Cannon, S.B., May, G.D., Vance, C.P. and Shoemaker, R.C. (2010) RNA-seq atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biol.* **10**, 160.

Spaink, H.P. (2000) Root nodulation and infection factors produced by rhizobial bacteria. *Annu. Rev. Microbiol.* **54**, 257–288.

Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L.A., Rhee, S.Y. and Stitt, M. (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **37**, 914–939.

Usadel, B., Nagel, A., Steinhauser, D., Gibon, Y., Bläsing, O.E., Redestig, H., Sreenivasulu, N., Krall, L., Hannah, M.A., Poree, F., Fernie, A.R. and Stitt, M. (2006) PageMan: an interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. *BMC Bioinformatics*, **7**, 535.

Wiborg, O., Hyldignielsen, J.J., Jensen, E.O., Paludan, K. and Marcker, K.A. (1982) The nucleotide sequence of two leghemoglobin genes from soybean. *Nucleic Acids Res.* **10**, 3487–3494.

Supporting information

Additional Supporting information may be found in the online version of this article:

Figure S1 MapMan diagram illustrating differentially regulated functional categories.

Table S1 Expression of soybean genes (RNA-seq data) in roots inoculated with WT *B. japonicum* and *nodC*⁻ *B. japonicum* at 48 h p.i.

Table S2 List of all up-regulated genes and their expression (RNA-seq data).

Table S3 List of all down-regulated genes and their expression (RNA-seq data).

Table S4 MapMan functional categories (BINs) with contents that are statistically different from other BINs.

Table S5 qRT-PCR primer sequences.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.