Discovery and expression analysis of nodulation genes in soybean

Peter M. Gresshoff
ARC Centre of Excellence for Integrative Legume Research
and School of Life Sciences, The University of Queensland Qld 4072
St. Lucia, Brisbane, QLD4072, Australia; p.gresshoff@botany.uq.edu.au

Abstract:

Soybean develop new root organs that become nitrogen-fixing nodules in response to signals from rhizobacteria and then use complex systemic signalling to control cell proliferation. Recent application of functional genomics resulted in the characterization of genes encoding Leucine rich repeat (LRR) receptor kinases controlling these early nodulation steps. With short and long distance signal exchange involved, these discoveries of key genetic components signal a new era of mechanistic insights into soybean nodulation and nitrogen fixation symbiosis and plant developmental genetics in general.

Introduction:

Because of their special ability to interact with rhizobia, legumes form root organs in which they 'fix' literally inexhaustible nitrogen gas to ammonia, the substrate for comprehensive protein biosynthesis (Gresshoff, 1993). Soybean-derived protein comprises a major food and feed source for the world; accordingly scientific interest in the process by which the nodule structure is initiated and formed, then developed into a functional nitrogen fixing organ is of global significance and research interest.

The nitrogen-fixing symbiosis is the result of a complex plant-bacterium interaction, involving inter-kingdom communication, many plant processes found in other tissues as well as plant and bacterial cell differentiation (Gresshoff, 2003). It is likely that existing soybean genes were 'pirated' from related processes. Recent discoveries of the nodulation process and the isolation of essential genes in that ontogeny (Fig. 1) demonstrate that the nodule symbiosis comprises pre-existing components for lateral root initiation, pollen tube growth, and disease resistance (Szczyglowski and Amyot, 2003).

Within this specialized nodule structure, rhizobia and plant cooperate to split atmospheric N_2, and then assimilate the resultant ammonia to create a symbiosis in which the plant obtains essential nitrogen while the soil bacterium receives carbon. The symbiosis is largely controlled by the plant as evidenced by a large number of mutations isolated in the plant (see Caetano-Anollés and Gresshoff, 1991). Soil nitrate also inhibits nodulation and nitrogen fixation in an yet ill-understood mechanism. However, the discovery of the supernodulation gene in soybean (Searle et al, 2003), which also has the ability to nodulate in the presence of nitrate, is a positive step towards understanding mechanisms.

Signalling is needed for nodulation

The initial signals come from the host plant that exudes a "cocktail" of sugars, carboxylic acids and flavonic substances (Oldroyd, 2001). These feed the complex 'rhizosphere', the population of microbes living in a complex consortium within a few millimeters of the root epidermis. Different rhizobia respond to different flavonic signals (including the phytoestrogen genistein) to synthesise lipo-chito-oligosaccharides ('Nod factors'; Dénarié et al, 1996) similar to fungal and insect cell wall components and haparin sulphate ligands commonly found in mammalian cells. Nod factors decorated with a set of lipid, acetyl, sulphate, and/or fucose moieties create specificity of plant response. Nod-factor, at nanomolar level, signals the induction of root hair deformation and cell division in a compatible symbiotic partner leading to mitosis as well as to infection of cortical and pericycle cells (Fig. 1).

Radial information also governs the location of nodule initiation opposite the xylem pole with ethylene most likely is one of the key signals in that process (Lohar et al, 2003). Evidence for the control of radial positional clues comes from the analysis of transgenic *Lotus japonicus* plants that express a dominant transgene from *Arabidopsis thaliana* (*AtETR1-1*). This gene codes for the ethylene non-binding ethylene receptor, conferring ethylene insensitivity. *AtETR1-1* Lotus plants possess strong seedling ethylene insensitivity, altered floral senescence and pod shattering and nodulate more abundantly than the wild type, but appear to be still subject to autoregulation and nitrate inhibition. Increased nodulation was correlated with increased cell divisions in the region opposite the phloem poles. Thus at least two pathways regulate nodule numbers in legumes (c.f., Penmetsa et al, 2003).

Legume genetics, genomics and model legumes

Plant genetics of the symbiosis has long lagged behind bacterial analysis because of genomic complexity, long generation times, tetraploidy, and lack of high throughput technologies. For example, *Glycine max* (soybean) suffers from a duplicated tetraploid genome (DNA content about 1,100 Mb; chromosomes 2n=4x=40), so it took 17 years between the isolation of supernodulation mutants (Carroll et al, 1985) and the cloning of the autoregulation of nodulation (AON) gene (Searle et al, 2003).

This situation was improved through a model plant approach, mirroring the success of *Arabidopsis thaliana* that uses the biological and genetic advantages of *Lotus japonicus* (2n=12; about 466 Mb; Kawaguchi et al, 2002) and *Medicago truncatula* (2n=16; about 470 Mb). Comparative genomics based on microsynteny and gene sequence conservation permits information transfer from the simpler model systems to crop legumes of more genetic complexity and considerably smaller databases (c.f., Krusell et al, 2002; Endré et al, 2002; Stracke et al, 2002; Nishimura et al 2002). Using technology developed originally for Arabidopsis, our laboratory was the first to use micro-array spotted Expressed Sequence Tags (ESTs) from an US gene collection to monitor the concurrent expression of 4,000 soybean genes in soybean shoots and roots (Maguire et al, 2002).

Several plant genes controlling early nodulation steps were isolated by positional cloning of chemically induced mutant loci (Gresshoff, 2003). Interestingly all turn out to be receptor-like kinases (RLKs).

RLKs are transmembrane proteins found ubiquitously across species and kingdoms being involved in many processes in plants such as pathogen interactions, hormone reception, apical meristem development and cell fate, and now nodulation. Many share extracellular leucine-rich repeat (LRR) regions that interact with external protein ligands including small peptides (Fig. 2). Significantly the newly discovered legume RLKs share substantial similarity with those found in *Arabidopsis* and other plants.

Soybean regulates organ number to attain homeostasis. This balance is distorted by nodule initiation that generates a large number of newly formed meristem clusters. Plants respond by a systemic autoregulation of nodulation (AON) feedback loop that permits the progression of first formed nodule primordia but inhibits ontogenically younger ones (Delves et al, 1986; Fig. 3). Loss-of-function mutations develop abundant supernodulation and altered lateral root morphology. In both *L. japonicus* (Krusell et al, 2002; Nishimura et al, 2002) and soybean (Searle et al, 2003; Men et al, 2003) map-based cloning succeeded in isolating the *HAR1* and *NTS1* loci, respectively, and characterizing them to code for RLKs, but unrelated to NORK/SYMRK (Endre et al, 2002). Interestingly *NARK* (nodule autoregulation receptor kinase) and *HAR1* (hypernodulation and altered root) like *AtClV1* functions by controlling the progression of nodule meristem cells but within the root. The definition of this key regulatory gene hopefully will facilitate the isolation of the root signal Q and the leaf derived inhibitor (SDI) response. Soybean also possesses a close relative to GmNARK called GmCLV1A (Searle et al, 2003). This protein may be the functional homologue to the AtCLV1 that regulates cell fate in the proliferating shoot apical meristem (SAM).

LjHAR-1, PsNARK and *GmNARK* define the same gene, sharing sequence and single intron position with the well-characterized *CLAVATA1* (*CLV1*) receptor kinase gene of *A. thaliana* (Clark et al, 1997; Fletcher, 2002). Mutations in *AtCLV1* develop enlarged meristems,

leading to a club-like (Latin: club=*clavata*) morphology. AtCLV1 (980 amino acids) interacts with a truncated version of itself, namely CLV2 (780 amino acids; lacking the kinase domain). This transmembrane complex binds the peptide CLV3 (96 amino acids) and possibly other factors to facilitate ligand-specific autophosphorylation on serine and threonine substrates. On the cytoplasmic side the kinase domain is known to interact with KAPP (kinase associated protein phosphatase) and a ROP protein (a Rho-like GTPase) that cleaves GTP and may activate a MAP kinase (Fig. 4). This kinase presumably phosphorylates and inactivates the WUSCHEL gene product (a transcription factor) that normally drives a gene needed for cell proliferation. Thus binding of CLV3 (and its undefined co-factor) results in the eventual cessation of cell proliferation and mutations in the cascade result in over-proliferation. Whether the same interacting partners exist for the NARK protein is being investigated. However, sequence conservation of the kinase domain of NARKs and CLV1-RKs suggests common serine-threonine substrates and putative interaction sites. Preliminary analyses of the kinase domain show strong sequence conservation between GmCLV1A, GmNARK and AtCLV1, suggesting that similar interactions can be expected. If so, the challenge is to define the sites of specificity that distinguish SAM versus AON circuits.

Significantly the two circuits, while closely related, differ significantly in the final target, contrasting short distance versus long distance regulation. However, in both cases cell proliferation, an essential feature of the multi-cellular status, is controlled. It is noteworthy that despite extensive mutational screens, all supernodulation mutants are altered in the *NARK* gene (20 total so far sequenced, see Fig. 2).

Expression analysis using quantitative RT-PCR demonstrated that the *GmNARK* gene is expressed independent of nodulation or mutant status in all major plant tissues; in soybean it is highly expressed in young and mature leaves, consistent with the autoregulation of nodulation model (Gresshoff, 1993; Searle et al, 2003), but also in the internodes, epicotyl and root (D. Buzas, unpubl. data). This latter finding presents a paradox; why does the root NARK transcript fail to cross-feed the NARK deficiency in the shoot of chimeric, grafted plants?? One possibility is that post-transcriptional controls provide specificity. Intracellular

trafficking may be regulated, as could be post-translational modifications. Is the pronounced branched root phenotype (Wopereis et al, 2000) of the *Lotus har1* mutants an indication of root activity?

The paradigm shift allowing model legumes to lead to gene discovery in large genome crop legumes such as soybean provides strong signals that a new era of mechanistic analysis of the nodule symbiosis is with us. Whether the new gene discoveries will lead to better legumes and sustainable agricultural practices of benefit to environmental and human health, will be the challenge of the next few years.

References:

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

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.

Clark S. E., Williams, R.W., and Meyerowitz, E.M. (1997) The CLAVATA1 gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. *Cell* **89**: 575-585.

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

Endré, G., Kereszt, A., Kevei, Z., Mihacea, S., Kalo, P. and Kiss, G.P. (2002) A receptor kinase regulating symbiotic nodule development. *Nature* **417**: 962-966.

Fletcher, J.C. (2002) Shoot and floral meristem maintenance in Arabidopsis. *Annu. Rev. Plant Biol.* **53**: 45-66.

Gresshoff, P.M. (1993) Molecular genetic analysis of nodulation genes in soybean. *Plant Breeding Reviews* **11**: 275-318.

Gresshoff, P.M. (2003) Post-genomic insights into nodulation. *Genome Biology* **4 (1)**:201

Kawaguchi, M., Imaizumi-Anraku, H., Koiwa, H., Niwa, S., Ikuta, A., Syono, K., and Akao, S. (2002) Root, root hair and symbiotic mutants of the model legume *Lotus japonicus*. *Mol. Plant. Microbe Interact*. **15**:17-26.

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 is mediated by a receptor-like kinase. *Nature* **420**:422-426.

Lohar, D., Stiller, J., Hababunga, S., Kam, J.H.W., Dunlap, J., Stacey, G. and Gresshoff, P.M. (2003) Ethylene insensitivity conferred by the *Arabidopsis etr1-1* receptor gene alters the nodulation response of transgenic *Lotus japonicus.* In review

Maguire, T.L., Grimmond, S., Forrest, A., Iturbe-Ormaetxe, I., Meksem, K., Gresshoff, P.M. (2002) Tissue-specific gene expression monitored by cDNA microarray analysis of soybean (*Glycine max*). *Journal of Plant Physiology* **159**:1361-1374.

Men, A.E., Linaya, T.S., Searle, I.R., Iturbe-Ormaetxe, Hussain, A.K.M., Gresshoff, I., Jiang, Q., Carroll, B.J., Gresshoff, P.M. (2002) Fast neutron mutagenesis produces a supernodulating mutant containing a large deletion in linkage group H of soybean (*Glycine soja* L.). *Genome Letters* **1**:147-155.

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.

Oldroyd, G.E.D. (2001) Dissecting symbiosis: developments in nod factor perception. *Annals of Botany* **87**:709-718.

Penmetsa, R.V., Frugoli, J.A., Smith, L.S., Long, S.R., and Cook, D.R. (2003) Dual genetic pathways controlling nodule number in *Medicago truncatula*. *Plant Physiology* **131**:1-11.

Searle, I.R., Men, A.M., Laniya, T.S., Buzas, D.M., Iturbe-Ormaetxe, I., Carroll, B.J., and Gresshoff, P.M. (2003) Long distance signalling for nodulation control in legumes requires a CLAVATA1-like receptor kinase. *Science* **299**:108-112 (first published on line 31. Oct. 2002. 10.1126/science.1077937).

Stracke, S., Kistner, C., Yoshida, S., Mulder, I., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J., Szczyglowski, K. and Parniske, M. (2002) A plant receptor-like kinase required for both bacterial and fungal symbiosis, *Nature* **417**:959-962.

Szczyglowski, K. and Amyot, L. (2003) Symbiosis; inventiveness by recruitment? Plant Physiol. **131**, xxx-yyy.

Wopereis, J., Pajuelo, E., Dazzo, F.B., Jiang, Q., Gresshoff, P.M., de Bruijn, F.J., Stougaard, J. and Szczyglowski, K. (2000) Short root mutant of *Lotus japonicus* with a dramatically altered symbiotic phenotype. *Plant J.* **23**:97-114.

Figures:

Figure 1:

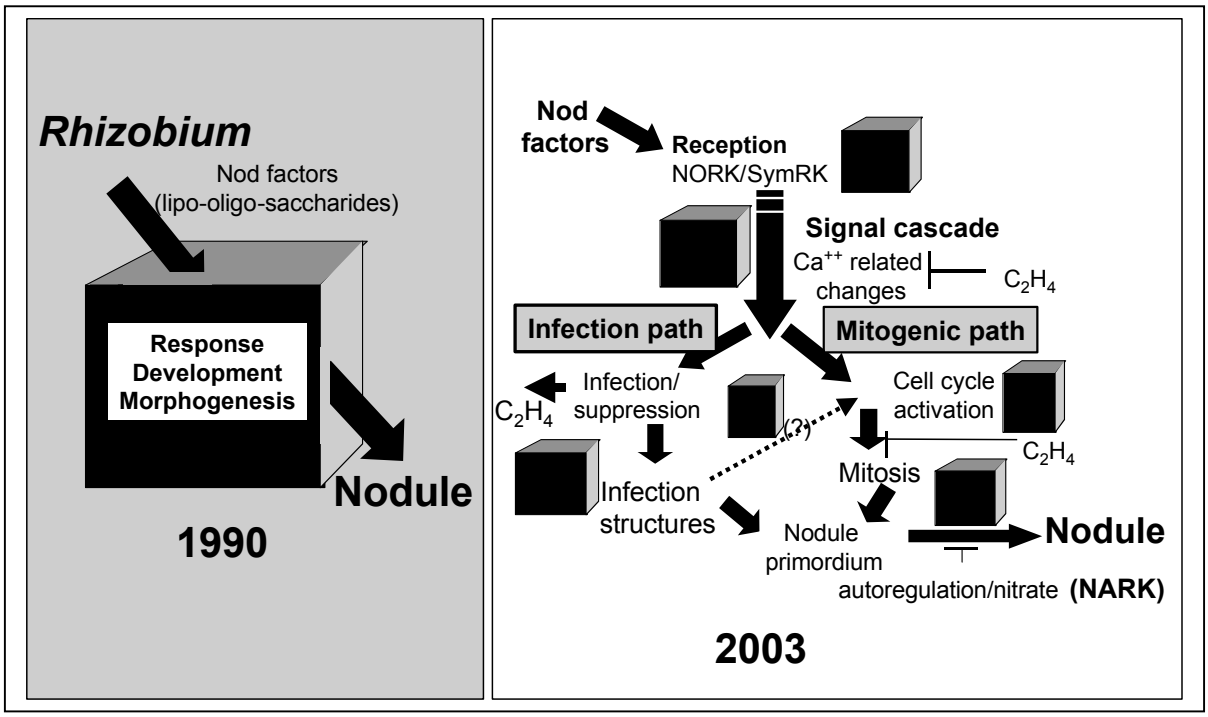


Figure 2:

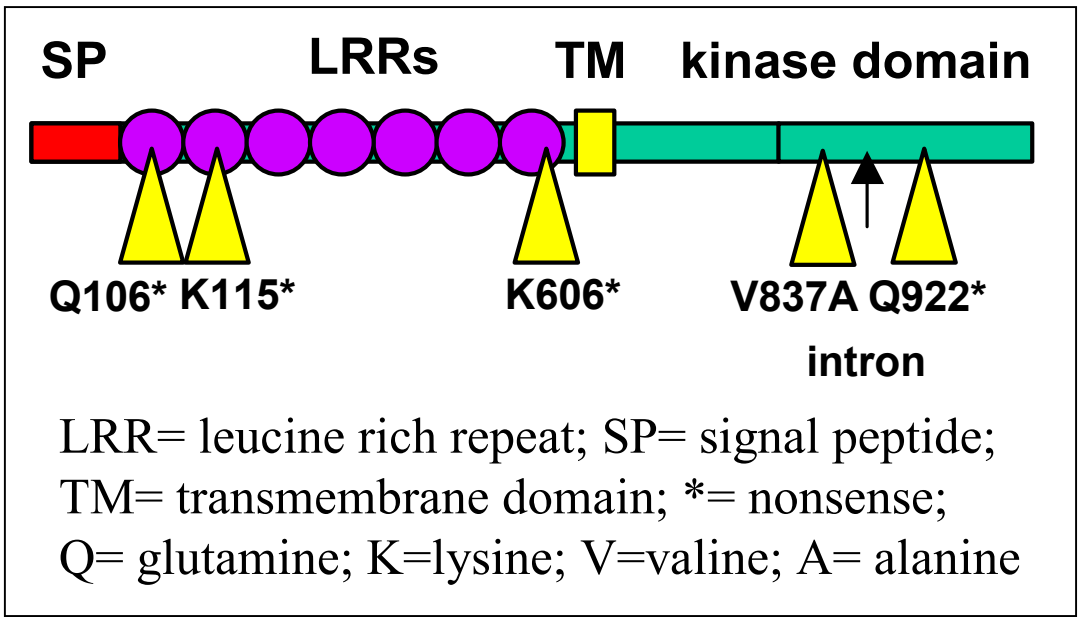

SP
LRRs
TM
kinase domain
Q106* K115*
K606*
V837A Q922*
intron
LRR= leucine rich repeat; SP= signal peptide;
TM= transmembrane domain; *= nonsense;
Q= glutamine; K=lysine; V=valine; A= alanine

Figure 3:

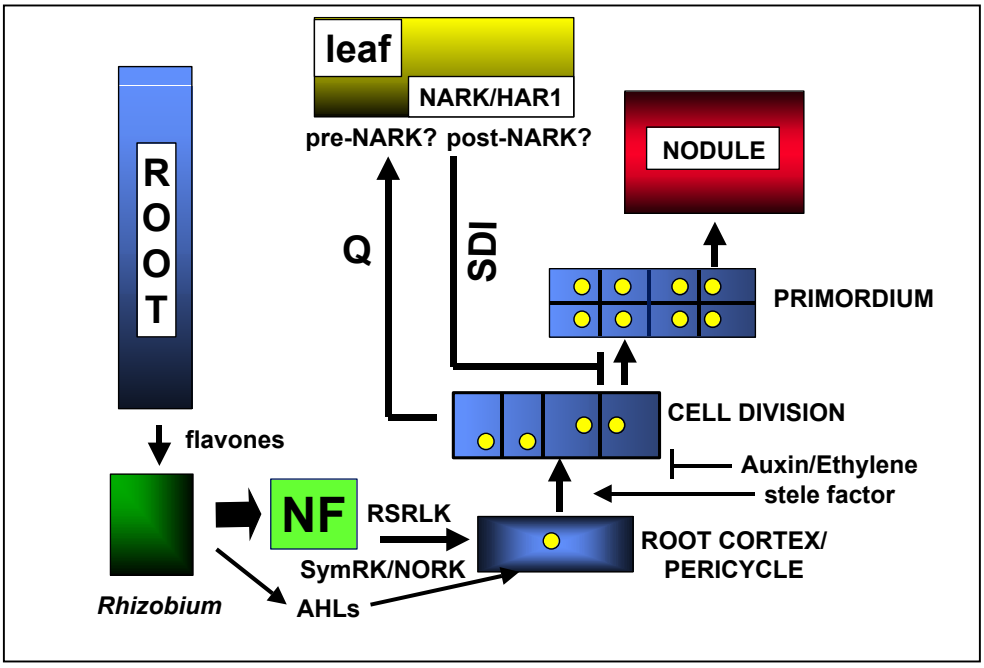
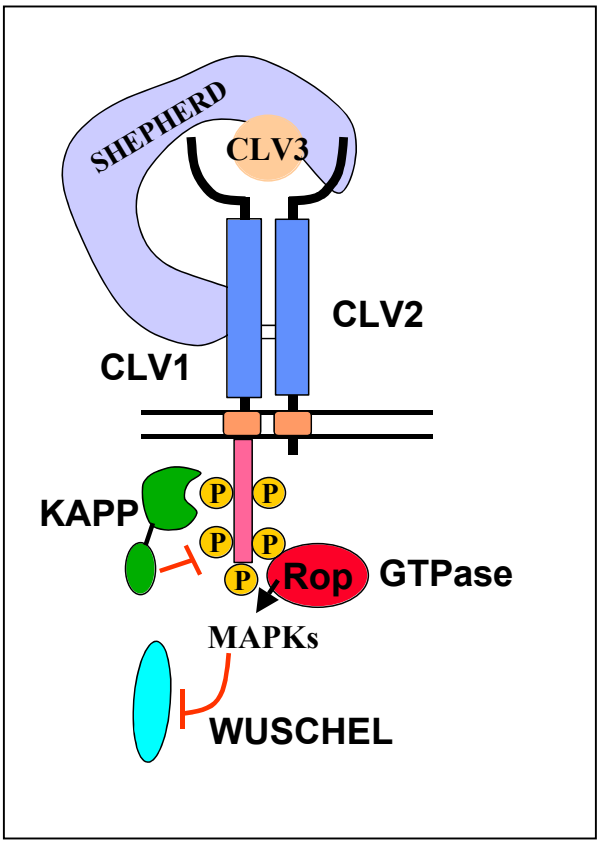

leaf
NARK/HAR1
pre-NARK?
post-NARK?
NODULE
ROOT
Q
SDI
PRIMORDIUM
CELL DIVISION
flavones
Auxin/Ethylene
stele factor
NF
RSRLK
ROOT CORTEX/
PERICYCLE
SymRK/NORK
Rhizobium
AHLs

Figure 4:


SHEPHERD
CLV3
CLV2
CLV1
KAPP
P
P
P
P
P
Rop
GTPase
MAPKs
WUSCHEL

Legends to figures:

Figure 1: Nodulation as an ontogenic pathway involving multiple plant processes. By 1990 extensive information existed on the nature of the development, and several biochemical functions within nodules. However, a large "black box" continued to exist except for the fact that bacterial genetics of various *Rhizobium* species delineated the genes needed for the synthesis of the Nod factors (see Oldroyd, 2001 for review). By 2003, numerous "black boxes" continue to exist, but insight has been gained into some of the plant's contribution to the perception of the signals and the development of the nodule.

Figure 2: Structure and alleles of the soybean *NARK* gene: mutations lead to the loss of the activity and result in abundant nodulation. All mutations are in the Bragg background except for *En6500* (K606*) which is in Enrei (Japan). Mutation *nts1116* (V837A) is close to the active kinase site and leads to hypernodulation and less pronounced growth effects on roots.

Figure 3: Signal transmission controlling cell proliferation of nodules. Root exudes flavones that induce the synthesis of nod factor(s) (NF) in the *Rhizobium*. Concurrently, acyl homo lactone, affects the plant root. Receptor kinases including NORK/SYMRK and the gene product encoded by *Ljsym1*, lead to cell division that progress to a nodule primordium that continues to develop to a nodule. Ethylene and auxin are presumed internal negative signals controlling the initiation of cell proliferation. Ethylene insensitive transgenics and mutants are still subject to autoregulation, indicating that ethylene controls nodulation before autoregulation. Stele factor, comprised of uridine, is a positive regulator. Early cell division cluster signals the leaf by signal Q, which is perceived by the NARK/HAR1 receptor kinase in the leaf (Searle et al, 2003; Nishimura et al, 2002; Krusell et al, 2002). NARK may be associated with tissue-specific pre-NARK and post-NARK proteins, possibly resembling proteins found in the CLAVATA1 cascade (Fletcher, 2002; Fig. 4). This activates signal SDI that is translocated to the root and inhibits progression cell proliferation to nodule primordia. NARK activity in the root, coupled with root-specific pre- and post-NARK factors may regulate lateral root initiation from pericycle.

Figure 4: Possible structure of the nodulation autoregulation receptor kinase (GmNARK) and its associated complex: (bsed on the CLAVATA1 model; modified from Fletcher, 2002). GmNARK is closely related to GmCLV1A that in turn is very similar to AtCLV1 (Searle et al, 2003). Extensive research in *Arabidopsis thaliana* suggests that several peptides and proteins interact with AtCLV1 (Becraft, 2002; Fletcher, 2002). WUSCHEL is a transcription factor regulating a cell proliferation gene; SHEPHERD is a chaperonin-like protein facilitating the assembly to the CLV1-CLV2 complex. ROP is a Rho-like GTPase and KAPP is a kinase associated protein phosphatase. CLV3 is a short peptide ligand. Orthologues or paralogues of relevant *Arabidopsis* genes have now been found in soybean.