THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Computational Modelling of Plant Signalling Control: a Case Study Based on Legume Autoregulation of Nodulation

Liqi Han

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in August 2010

School of Information Technology and Electrical Engineering

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the General Award Rules of The University of Queensland, immediately made available for research and study in accordance with the *Copyright Act 1968*.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material.

Statement of Contributions to Jointly Authored Works Contained in the Thesis

1. **Han L, Gresshoff PM, Hanan J** (2010a) A functional-structural modelling approach to autoregulation of nodulation. Annals of Botany **DOI:** 10.1093/aob/mcq182

    - *Han was responsible for 70% of conception and design, 100% of experimentation, 100% of data processing and analysis, and 90% of drafting and writing; Gresshoff was responsible for 15% of conception and design, and 5% of drafting and writing; Hanan was responsible for 15% of conception and design, and 5% of drafting and writing.*

2. **Han L, Hanan J, Gresshoff PM** (2010b) Computational complementation: a modelling approach to study signalling mechanisms during legume autoregulation of nodulation. PLoS Computational Biology **6:** e1000685

    - *Han was responsible for 60% of conception and design, 100% of experimentation, 100% of data processing and analysis, and 80% of drafting and writing; Hanan was responsible for 20% of conception and design, and 10% of drafting and writing; Gresshoff was responsible for 20% of conception and design, and 10% of drafting and writing.*

3. **Han L, Gresshoff PM, Hanan J** (2009) Modelling root development with signalling control: a case study based on legume autoregulation of nodulation. *In* B Li, M Jaeger, Y Guo, eds, Plant Growth Modeling and Applications, Proceedings of PMA09. IEEE Computer Society, Los Alamitos, pp 134-141

    - *Han was responsible for 70% of conception and design, 100% of experimentation, 100% of data processing and analysis, and 90% of drafting and writing; Gresshoff was responsible for 15% of conception and design, and 5% of drafting and writing; Hanan was responsible for 15% of conception and design, and 5% of drafting and writing.*

4. **Han L, Gresshoff PM, Hanan J** (2007a) Virtual soybean—a computational model for studying autoregulation of nodulation. *In* P Prusinkiewicz, J Hanan, B Lane, eds, The 5th International Workshop on Functional Structural Plant Models, Napier, New Zealand

    - *Han was responsible for 60% of conception and design, 100% of experimentation, 100% of data processing and analysis, and 90% of drafting and writing; Gresshoff was responsible for 20% of conception and design, and 5% of drafting and writing; Hanan was responsible for 20% of conception and design, and 5% of drafting and writing.*

5. **Han L, Gresshoff PM, Hanan J** (2007b) Computational approaches for studying signalling mechanisms behind legume autoregulation of nodulation. *In* Bioinformatics Australia 2007 Conference, Brisbane, Australia

    - *Han was responsible for 60% of conception and design, 100% of programming, and 90% of drafting and writing; Gresshoff was responsible for 20% of conception and design, and 5% of drafting and writing; Hanan was responsible for 20% of conception and design, and 5% of drafting and writing.*

6. **Han L, Gresshoff PM, Hanan J** (2007c) Computational models for studying signalling control mechanisms behind legume autoregulation of nodulation. *In* The 8th Asia-Pacific Complex Systems Conference, Gold Coast, Australia

    - *Han was responsible for 60% of conception and design, 100% of programming, and 90% of drafting and writing; Gresshoff was responsible for 20% of conception and design, and 5% of drafting and writing; Hanan was responsible for 20% of conception and design, and 5% of*

drafting and writing.

7. **Han L, Gresshoff PM, Hanan J** (2007d) Computational approaches for studying legume autoregulation of nodulation. *In* Annual Symposium 2007, ARC Centre of Excellence for Integrative Legume Research, Kingscliffe, Australia

   - *Han was responsible for 60% of conception and design, 100% of programming, and 90% of drafting and writing; Gresshoff was responsible for 20% of conception and design, and 5% of drafting and writing; Hanan was responsible for 20% of conception and design, and 5% of drafting and writing.*

Statement of Contributions by Others to the Thesis as a Whole

No contributions by others.

Statement of Parts of the Thesis Submitted to Qualify for the Award of Another Degree

None.

Published Works by the Author Incorporated into the Thesis

1. **Han L, Gresshoff PM, Hanan J** (2010a) A functional-structural modelling approach to autoregulation of nodulation. Annals of Botany **DOI:** 10.1093/aob/mcq182

   - *Incorporated as Chapter 4.*

2. **Han L, Hanan J, Gresshoff PM** (2010b) Computational complementation: a modelling approach to study signalling mechanisms during legume autoregulation of nodulation. PLoS Computational Biology **6:** e1000685

   - *Incorporated as Chapter 5.*

Additional Published Works by the Author Relevant to the Thesis but not Forming Part of it

1. **Han L, Hanan J, Dreccer MF** (2009) Spectral reflectance of wheat recombinant inbred lines: a computational modelling study. *In* The 3rd International Conference on Integrated Approaches to Improve Crop Production under Drought Prone Environments, Shanghai, China

Acknowledgements

I would like to thank the following people:

My advisors – Dr Jim Hanan and Prof Peter Gresshoff – for teaching me and guiding me patiently, and organising resources for me to conduct the research;

Dr Jim Hanan, Prof Peter Gresshoff, Prof Peter Lindsay, and Prof Janet Wiles for organising funds and providing a multi-disciplinary environment for me to do this project;

Prof Janet Wiles, Dr Marcus Gallagher, Prof Peter Lindsay, Dr Christine Beveridge and Dr Peter Robinson for their academic advice through the milestones;

Dr S-Zahra Hosseini Cici, Mr Yu-Hsiang Lin, Dr Qunyi Jiang, Ms Dongxue Li, Dr Sureeporn Nontachaiyapoom, Dr Pick Kuen Chan, Dr Bandana Biswas, and Dr Akira Miyahara for helping me study biology;

Mr Mikolaj Cieslak for exchanging research ideas and modelling techniques;

Ms Dongxue Li, Dr Ning Chen and Dr S-Zahra Hosseini Cici for their advice on the experiments on growing, observing and measuring soybean plants;

Mr Kuang-Hsu Wu for his participation in collection of soybean growth data;

Ms Dongxue Li for implementing the real-plant graft experiment as well as collecting and processing data;

My parents and my wife for their love, care, expectations, encouragements, and financial support that contribute greatly to my PhD study.

This thesis has been funded with scholarships from School of Information Technology and Electrical Engineering, ARC Centre of Excellence for Integrative Legume Research, ARC Centre for Complex Systems and the Graduate School based at The University of Queensland.

Abstract

Signalling mechanisms play a vital role in plant development and function, controlling processes such as germination, branching, flowering and nodulation. However, these dynamic processes are so complex that details of their operation are still largely unknown. Endogenous signals in particular, such as those based on plant hormones and peptides, are difficult to observe and remain a critical challenge for botanic research. As an addition to conventional experimental approaches, computational modelling has emerged as a powerful tool for understanding the complexity of signalling occurring at and between different levels of organisation in plant systems. This thesis develops new methods and strategies for using computational modelling to study a typical internal shoot-root signalling system – the autoregulation of nodulation in legumes.

Nodulation is a developmental process resulting from the symbiosis of legume plants with a group of bacteria known as rhizobia. The rhizobia colonise legume roots to house themselves and provide fixed nitrogen for the host plants. Since excessive nodulation can cause overconsumption of resources and disturbs plant growth, the legumes have developed a regulatory system – autoregulation of nodulation – to maintain the balance of nodule formation. The general framework of this signalling system has been established based on experimental findings. It has been hypothesised that the nodule formation process in the roots induces a signal moving to the leaves, which triggers a shoot-derived inhibitor moving back to the root to inhibit further nodulation. However, due to the intricacy of internal signalling and absence of flux and biochemical data, detailed mechanisms during autoregulation of nodulation remain largely unclear. The shoot-root regulatory signals also remain unidentified. To address this, this thesis focuses on the inter-organ signalling of autoregulation of nodulation and uses functional-structural plant modelling to investigate its mechanisms.

At the technical level, there were two major challenges for using functional-structural modelling to study autoregulation of nodulation: one is reconstruction of the 3D architecture of legume roots; the other is coordination of the signalling and development processes. Using soybean as the target legume plant and the L-system-based software L-studio as the modelling and simulation platform, a series of methods and techniques have been developed in this thesis to collect root development data, reconstruct root architecture, and synchronise the multi-rate signalling and developmental processes.

At the strategic level, a new modelling approach called “Computational Complementation” has been developed in this research. The key idea is to use functional-structural modelling to complement, with hypothetical signalling mechanisms, the deficiency of an empirical model of a mutant plant

where the function of autoregulation of nodulation is totally lost. If the complementation leads to a regulation result the same as or similar to the wild-type phenotype, this supports the validity of the hypothesised mechanisms. The initial application of computational complementation was to investigate whether or not wild-type soybean cotyledons provide the shoot-derived inhibitor to regulate nodule progression. Two opposing hypotheses were tested with virtual experiments: (a) cotyledons function as part of the root, incapable of producing the shoot-derived inhibitor; or (b) cotyledons function as part of the shoot, involved in regulating root nodules. The virtual-experiment results suggested that hypothesis (b) was more likely to be correct, which was confirmed by a real-plant grafting experiment. This demonstrates the feasibility of computational complementation and shows its usefulness for future applications.

Suggested future research includes exploration of better techniques for model construction, application of computational complementation to help in identifying the unknown shoot-root regulatory signals and integration of lower-scale signalling models. The modelling and simulation methods as well as the computational complementation strategy developed in this thesis can be applied beyond the study of autoregulation of nodulation. They also have the potential to be used in wider studies on plant signalling, such as those on branching regulation, flowering control and lateral initiation.

Keywords

complex systems, L-systems, functional-structural plant modelling, computational biology, synchronisation

Australian and New Zealand Standard Research Classifications (ANZSRC)

080110 Simulation and Modelling 70%, 060705 Plant Physiology 30%

Contents

List of Figures

List of Tables

List of Abbreviations

1 Introduction

Plants, making up 99.9% of the Earth's biomass, play a vital role in maintenance of the ecosystem and sustaining other forms of life. Like a natural factory, they capture photons, air, water and minerals, and then synthesise all the necessary products we need to survive. For us humans, the food we eat, the oxygen we breathe, and the medicine we use all rely on plants. Although they are so important, many of their internal mechanisms are still largely unknown. Modern biology has seen great discoveries moving us closer and closer to answering the question: how do plants function?

Despite our accumulation of knowledge, the more we know, the more we realise how complex the plants are. Aside from their essential biochemical nature (as carbon-based life), a plant is also a complex system with various components interacting dynamically. To investigate the complexity occurring at different levels (from protein synthesis to phenotypic expression), interdisciplinary efforts involving computer science and systems theory have been incorporated in biological studies, forming new research fields known as computational biology and systems biology. As a part of these emerging studies, the research described in this thesis uses legumes as a basic material, and computer modelling as a basic tool, to develop computational approaches and techniques to obtain a better understanding of the complex signalling mechanisms behind legume autoregulation of nodulation.

1.1 Why plant signalling?

Plant growth is far more than increment in size and mass. Individual plants develop differently, forming various specialised structures and behaviours, due to their capability of sensing and responding to regulatory signals ranging from environmental factors to endogenous substances (Mulligan et al., 1997). These signalling mechanisms, occurring at multiple levels, compose complex networks and dynamic processes to control plant development and function, such as flowering, fruit ripening, germination, photosynthesis, branching, nodulation and so forth. However, they are "so intricate, so pervasive and so subtle" that details about them are far from being completely known (Raven et al., 1999). Endogenous signals (plant hormones, microRNA and peptides in particular), which act frequently and widely inside the plant system and are difficult to

observe and manipulate, remain a critical challenge for botanical research. Undoubtedly, molecular studies have significantly improved our understanding of signalling control by gene sequences (De Moraes et al., 2004). But knowledge about the complex intermediate-level mechanisms is far from clear and there is still a long way to go to "jump the gap between cell, tissue and whole organism" (Trewavas, 2003). Additional approaches and technologies are in demand to deal with these complexities. One natural place to look is computational modelling.

1.2 Why computational modelling?

Although complexity makes analysis of biological systems challenging (Stelling, 2007), the similarity between them and engineered system (Weng et al., 1999; van Riel, 2006; Stelling, 2007) leads to a systematic view of those biological processes (Minorsky, 2003; Hammer et al., 2004). Underlying the signalling mechanisms is a network of components connected by intricate interfaces, with activities such as "assembly, translocation, degradation, and channelling of chemical reactions" occurring simultaneously (Weng et al., 1999). These components, internal processes and their interactions – also responding to the temporally and spatially changing environment – frame dynamic and complex systems at multiple levels to orchestrate plant behaviours. Full understanding of system properties emerging from component interactions cannot be achieved by simply "drawing diagrams of their interconnections" (Kitano, 2002). Neither should the individual behaviours of constituents be over-simplified (Trewavas, 2003). The differentiation of individual components should be respected and the communication between them should be addressed dynamically over time (Stamatopoulou et al., 2007). To represent the biological systems accurately with high-quality information, powerful tools are required for managing and processing massive data sets, and for simulating complex and dynamic mechanisms (Neves and Iyengar, 2002; Minorsky, 2003). Computational modelling can well meet these requirements and is an ideal option for studying signalling networks (Neves and Iyengar, 2002; Ray et al., 2002). More than simply providing data processing and descriptive use, a key benefit of computational modelling lies in its capability of conducting virtual experiments rapidly to improve hypotheses and to predict unknown factors (Ray et al., 2002; Minorsky, 2003; van Riel, 2006). It is not only increasingly useful (Neves and Iyengar, 2002), but also recognised by some biologists as an indispensable tool on the road to full understanding of unknown signalling mechanisms such as hormonal control (Hedden and Thomas, 2006).

1.3 Why legume autoregulation of nodulation?

Legumes are one of the largest flowering plant families (Doyle and Luckow, 2003) and are second most in importance for humans (Graham and Vance, 2003). They occupy only 12% to 15% of the Earth's arable surface but provide 27% of the world's primary crop production and more than 35% of the world's processed vegetable oil (Graham and Vance, 2003). They also have great potential to be a sustainable source of biodiesel production (Graham and Vance, 2003; Scott et al., 2008). Legumes are a major natural "nitrogen-fixer" as well – providing roughly 200 million tons of nitrogen each year (Kinkema et al., 2006) equivalent to 90 billion dollars worth of fertiliser replacement value. The nitrogen is fixed through a nodule development process termed "nodulation", resulting from the symbiosis of legume roots with a type of soil bacteria broadly called rhizobia. Legume nodulation is not only important to agricultural yield (Gresshoff, 1990), but also an environmentally friendly alternative to the use of synthetic nitrogen fertilisers (Graham and Vance, 2003; Gresshoff, 2003; Ferguson et al., 2010).

But for a legume plant itself, excessive nodulation may cause over-consumption of available resources and disturb its growth (Oka-Kira and Kawaguchi, 2006). Therefore the legume plants have evolved a signalling regulatory system to maintain balance of nodule formation, which is called "autoregulation of nodulation" (AON) or "feedback inhibition of nodulation" (Carroll et al., 1985b; Delves et al., 1986; Kinkema et al., 2006; Oka-Kira and Kawaguchi, 2006). A good understanding of this regulatory system may enable us to improve nodulation and thus to amplify the practical benefits mentioned above. Moreover, as the legume growth cycle is relatively short and nodulation is an easily observed phenotype, autoregulation of nodulation (AON) can also serve as an ideal example for studying plant signalling mechanisms.

From the earlier explorations in the 1980s (Carroll et al., 1985a, b; Delves et al., 1986; Olsson et al., 1989) to the most recent discoveries in this century (Ferguson and Mathesius, 2003; Searle et al., 2003; Schnabel et al., 2005; Oka-Kira and Kawaguchi, 2006; Buzas and Gresshoff, 2007; Nontachaiyapoom et al., 2007; Hayashi et al., 2008; Kinkema and Gresshoff, 2008; Miyahara et al., 2008; Gresshoff et al., 2009; Ferguson et al., 2010; Lin et al., 2010), more than 20 years of continuous efforts have been made to study AON and have greatly enriched our knowledge of this regulatory system. The general framework of short- and long-distance control has become much clearer (Ferguson and Mathesius, 2003; Gresshoff, 2003; Oka-Kira and Kawaguchi, 2006), but the identity of the signals involved and many details about their production, transport, perception and

function still remain unknown. Current knowledge about AON provides the basis for this thesis, while a better understanding of its complexity is the goal.

1.4 Overview of this thesis

Having addressed the motives and reasons for using computational modelling to study plant signalling as well as the significance and appropriateness of choosing autoregulation of nodulation as the target signalling system, this thesis will answer the following questions in the remaining chapters:

Q1 What is known and what remains unknown about AON?
Q2 How can AON be investigated using computational modelling approaches?
Q3 What are the basic concepts for modelling a biological system?
Q4 What kind of computational model is most suitable for AON?
Q5 What are the technical challenges for modelling AON and how can these challenges be met?
Q6 What strategy can this modelling study provide to really promote the understanding of AON?
Q7 What has been discovered through the modelling techniques and strategy?
Q8 What are the advantages and disadvantages of these modelling techniques and strategy?
Q9 What are the potential future directions?

To answer the first two questions, Chapter 2 will review the biological knowledge of AON, including the processes of nodule initiation and the framework of the long-distance shoot-root signalling control, as well as the latest progress in understanding AON based on conventional biological studies. Then the question of how AON can be viewed as a dynamic network and a complex system will be addressed, clarifying the entry point for using computational modelling to study its complexity.

To answer question Q3, general concepts and classifications of biological models will be given in the first section of Chapter 3. Plant architectural modelling and functional-structural plant modelling, which are the basis for Chapter 4, will also be reviewed in Chapter 3 to give necessary background details.

Question Q4 will be answered in the first section of Chapter 4: since AON is in essence a long-distance inter-organ regulatory network, functional-structural modelling is an ideal method for this

study. To answer question Q5, Chapter 4 will point out two major challenges at the technical level for using functional-structural modelling to study AON: the reconstruction of the 3D architecture of legume roots and the coordination of signalling-development processes with various rates. The modelling and simulation techniques developed in this thesis to meet these challenges will then be presented. The content of Chapter 4 has been accepted by Annals of Botany for publication as "A Functional-structural Modelling Approach to Autoregulation of Nodulation" (Han et al., 2010a).

Chapter 5 will cope with question Q6 by presenting the key strategy developed in this thesis – computational complementation, which uses computational modelling to complement a loss-of-function mutant with hypothetical mechanisms to restore its function – and evaluating the feasibility of this strategy through its first application. Question Q7 will be answered by the result of the first application of computational complementation: wild-type soybean cotyledons are found to be involved in AON signalling. The content of Chapter 5 has been published in PLoS Computational Biology as "Computational Complementation: A Modelling Approach to Study Signalling Mechanisms during Legume Autoregulation of Nodulation" (Han et al., 2010b).

Chapter 6 is the General Discussion to address questions Q8 and Q9, where the major contributions as well as the limitations of this thesis (including its limitation in generality to other plant signalling studies) will be discussed and the future directions for modelling studies will be suggested.

2 Background of autoregulation of nodulation

This chapter firstly reviews the biological background about autoregulation of nodulation (AON), including processes of nodule initiation and nodulation control, and then discusses the characteristics of AON as an inter-organ network and as a complex system.

2.1 Nodulation symbiosis

Symbiosis is a close and long-term association between organisms of different species (Raven et al., 1999). If the association is beneficial to one species while harmful to the other, it is called parasitic symbiosis. On the other hand, if both species can benefit from their interactions, it is called mutualistic symbiosis. Legume nodulation is based on an established mutualistic symbiosis between rhizobia (a group of soil-living bacteria) and leguminous plants. Through the symbiosis, legumes received fixed nitrogen from rhizobia to make protein, while rhizobia get carbohydrates from plants as necessary energy to survive (Figure 2.1).

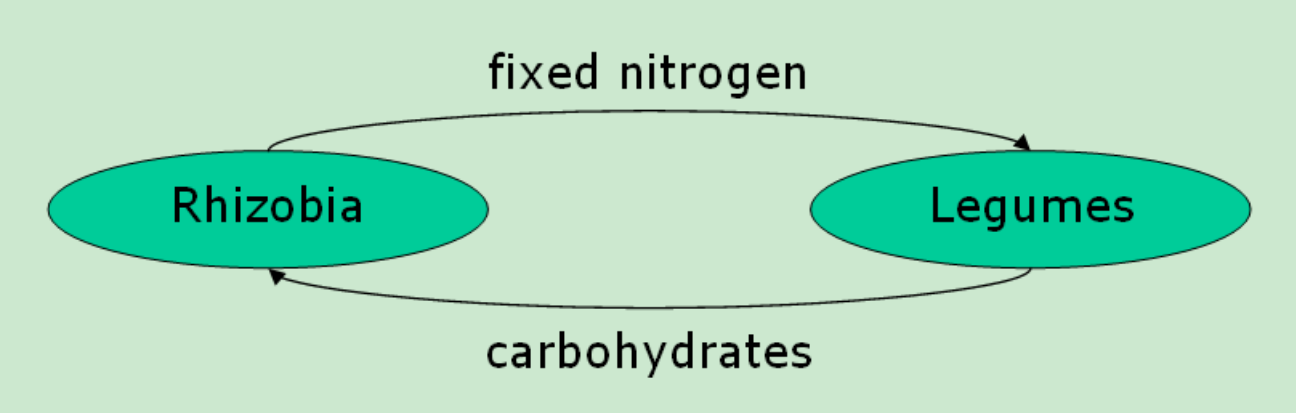


Figure 2.1. Symbiotic relationship between rhizobia and legume plants. In this symbiosis, rhizobia provide fixed nitrogen to legume plants, while legumes provide carbohydrates to rhizobia.

This interaction is initiated when rhizobia detect a type of chemical compound called a “flavonoid” released by legume roots to the soil (Redmond et al., 1986; Oldroyd and Downie, 2004; Kinkema et al., 2006). The flavonoid stimulates rhizobia to move towards host plant root hairs and to produce another signal that is widely called “Nod factor” (NF) (Spaink, 2000; Oldroyd and Downie, 2004). When the rhizobia are attached to root hairs, the NF released by them triggers a cascade of events

(Raven et al., 1999; Oldroyd and Downie, 2004; Kinkema et al., 2006). Firstly, the root hairs start swelling and deforming. Meanwhile, the NF also promotes cortical cell division inside the root, leading to formation of nodule primordium (the form of nodule in its earliest recognisable development stages). The deformed root hair then encloses the rhizobia and creates a micro-environment to facilitate their colonisation and invasion, through a tubular structure called an "infection thread", to the growing nodule primordium. As a result of the proliferation of rhizobia and root cortical cells, a tumour-like nodule is formed. The rhizobia inside the nodule then differentiate into a form known as "bacteroids". The bacteroids produce nitrogenise enzyme complex, which fixes nitrogen gas from soil air into ammonium for synthesis of amino acids.

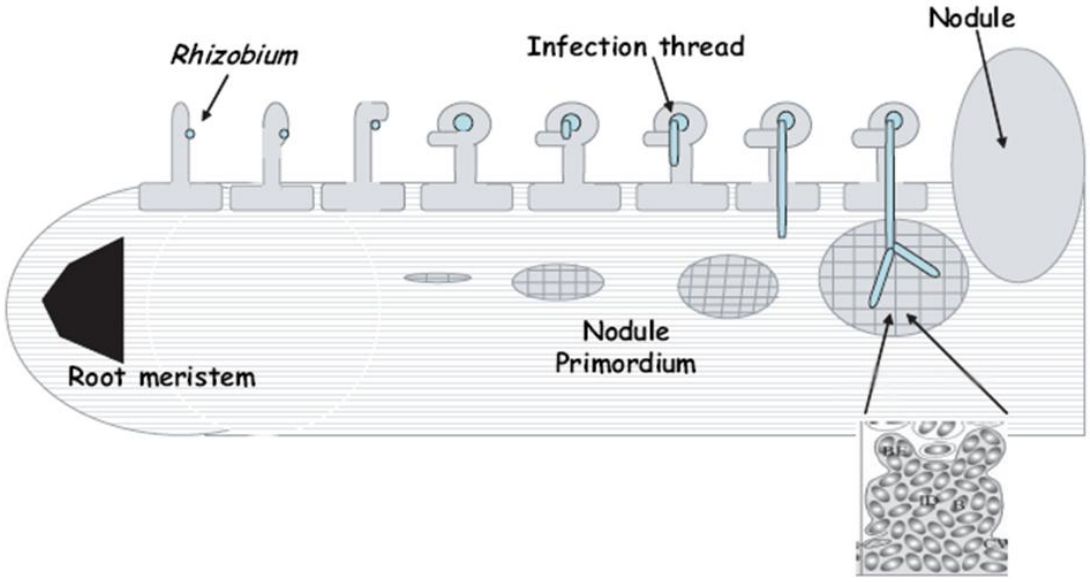


Figure 2.2. Process of rhizobia infection and nodule initiation. In response to the rhizobia attachment and perception of the signals they release, the root hairs conduct a series of behaviours, including swelling, deformation, and curling to entrap rhizobia and facilitate their colonisation through infection threads to form nodules. This figure is adapted from Kinkema et al. (2006).

2.2 Long-distance signalling control

For a legume plant itself, excessive nodulation could over-consume its metabolic resources and cause disproportional distribution of internal growth regulators (Oka-Kira and Kawaguchi, 2006). Therefore the legume plants have developed a signalling regulatory system known as "autoregulation of nodulation" (AON) to maintain the balance of nodule formation (Carroll et al., 1985a, b; Delves et al., 1986; Gresshoff, 2003; Oka-Kira and Kawaguchi, 2006). It has been hypothesised that a signal "Q" is induced by rhizobial invasion of nodule primordium, which then

moves through a root-shoot pathway to the leaves (Gresshoff, 2003; Hayashi et al., 2008; Ferguson et al., 2010). A leucine-rich repeat (LRR) receptor kinase in the phloem parenchyma of leaf vascular tissue (Nontachaiyapoom et al., 2007), related in structure to CLAVATA1 in Arabidopsis, detects the Q signal or an intermediate. This LRR receptor kinase is referred to as *GmNARK* in soybean (Searle et al., 2003; Miyahara et al., 2008), *HAR1* in Lotus (Krusell et al., 2002; Nishimura et al., 2002), and *SUNN* in Medicago (Schnabel et al., 2005). The perception of Q by the LRR receptor kinase triggers production of a shoot-derived inhibitor (SDI) that is transported to the root to inhibit further nodulation (Ferguson et al., 2010). Wild-type legume plants perform AON to keep their balance of nodule formation well maintained, while mutants (Carroll et al., 1985a, b; Wopereis et al., 2000) lacking this regulation demonstrate a phenotype called "supernodulation" or "hypernodulation" with many more nodules. For example, the wild-type soybean (*Glycine max* L. Merrill) genotype Bragg is controlled by AON and exhibits a normal number of nodules (Figure 5.1*A* and *C*), while its hypernodulation mutant *nts1116* (Carroll et al., 1985a, b; Hansen et al., 1989) cannot produce SDI due to the absence of *GmNARK* in its leaves and therefore allows a larger number of nodules to be formed (Figure 5.1*B* and *D*).

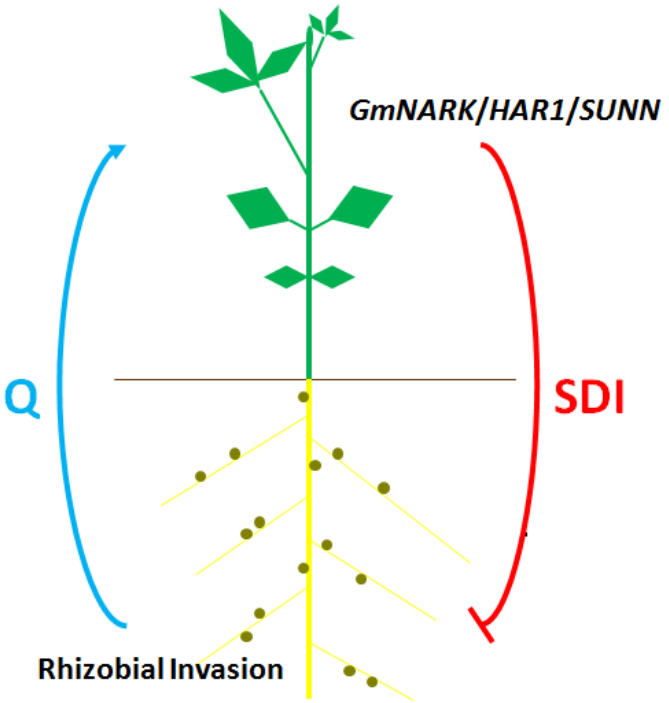


Figure 2.3. Autoregulation of nodulation. The rhizobial invasion of the legume roots triggers production of a signal (Q) that moves through a root-shoot pathway to the leaves. In the leaves, the Q signal activates a leucine-rich repeat (LRR) receptor kinase (known as *GmNARK* in soybean, *HAR1* in Lotus and *SUNN* in Medicago) and then stimulates production of a shoot-derived inhibitor (SDI) that is transported to the root to inhibit further nodulation.

According to recent studies, the Q signal is presumed to be a CLV3/ESR-related (CLE) peptide (Gresshoff et al., 2009; Okamoto et al., 2009). A root-specific gene *PsNOD3* may be involved in the production or transport of Q (Li et al., 2009; Ferguson et al., 2010). SDI has been shown to be a small, water-soluble, heat-stable and inoculation-dependant molecule and is unlikely to be an RNA or a protein (Lin et al., 2010). It has also been proposed that a plant hormone, abscisic acid, is not directly involved in AON (Biswas et al., 2009). Another plant hormone, auxin, might play the role of SDI for regulation of nodulation (Mathesius, 2008), but this hypothesis has not been confirmed. That is to say, the two major signals – Q and SDI – for AON remain unidentified and the detailed mechanisms of their production, transport, perception and function are still largely unknown.

2.3 A complex system

Aside from its subtle and intricate biochemical nature, the long-distance signalling of AON is also a network with intercommunicating components (Figure 2.4 *A*): the Q signal is produced by nodule primordia in the root, moves up to the leaves and induces SDI that is transported down to the root to inhibit further nodulation.

In early stages of this AON, when more and more new nodule primordia are initialised and developing in the root, the sources for Q production are gradually increased (Figure 2.4 *B*). Meanwhile, the leaf number and biomass also keep increasing (Figure 2.4 *C*), providing more vasculature containing the LRR receptor kinase to perceive Q and to trigger the production of SDI. The enhanced production of SDI increases the strength of its downward flows and intensifies the inhibition of nodulation. The strengthened regulation of nodulation prevents more potential nodules from formation and consequently weakens Q production (Figure 2.4 *D*). If the supply of Q is reduced, the production of SDI may also be abated (Figure 2.4 *E*). The decreased amount of SDI arriving in the root would allow more new nodules to be formed. The plant's development not only keeps changing the sources and targets for signal production, perception and function, but also keeps extending the distance of signal transport and ramifying the signal pathways (Figure 2.4 *F*). These factors and processes interact dynamically, leading to an emergent result – the pattern of nodulation, including number, size and distribution of nodules.

From this point of view, AON has all necessary characteristics to be treated as a complex system (Nicolis and Nicolis, 2007): dynamics, non-linearity, self-organisation and unpredictability. Its

complexity can neither be analysed only by intuition, nor can it be fully revealed by a few diagrams (such as Figure 2.3 and Figure 2.4). With its advantages in handling these kinds of complexities (as addressed in Section 1.2), computational modelling becomes an indispensible tool for studying AON.

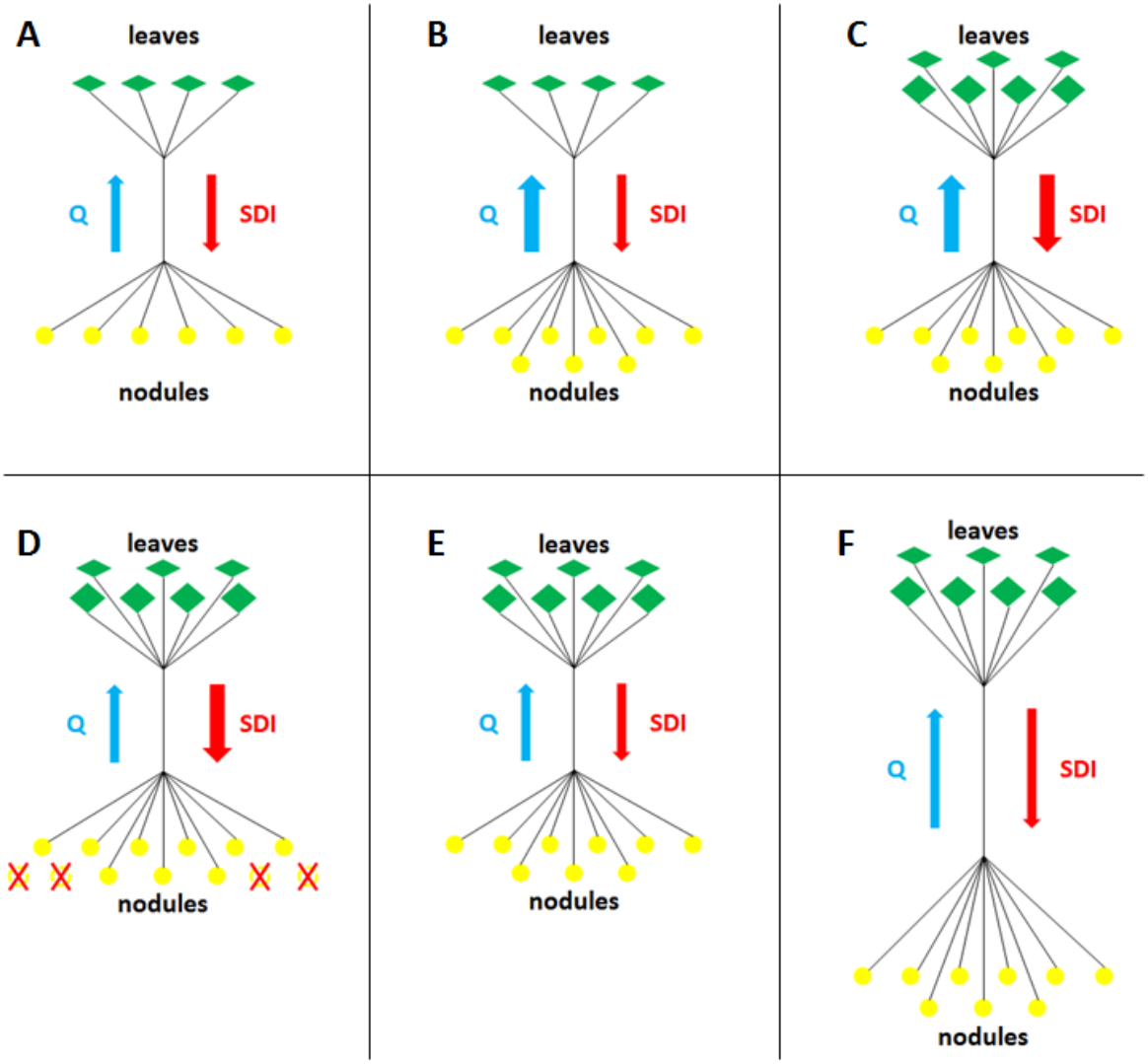


Figure 2.4. Illustration of autoregulation of nodulation as a dynamic network. AON is a network of leaves and nodules that are linked by Q and SDI signals (A). The strengths of the upward and downward signal flows (illustrated by the width of relevant arrows) as well as the distances for signal transport (illustrated by the length of arrows) keep modifying and being modified by the network. The addition of new nodules to the network provides more sources for Q production (B) and increases the amount of Q signal moving from root to shoot. The increment of leaf biomass as well as leaf number results in more places to perceive Q and synthesise SDI (C). The increased supply of SDI strengthens the inhibition of nodule formation and weakens Q production (D). Less Q arrives at the leaves, less SDI is triggered (E). Aside from changes to the network topology, the lengths of the shoot-root signal pathways also keep changing (F) as a result of plant elongation.

3 Basis of computational plant modelling

In this chapter, the basic modelling concepts related to this thesis are introduced, including the definition and classification of biological models, the philosophy of plant architectural models and the background of functional-structural plant modelling.

3.1 Basic concepts of biological models

A scientific model is broadly regarded as a representation of a system composed by interrelated objects (Haefner, 2005). Such representations could be built in different forms, focused on different hierarchical levels, supported by different strategies and used for different purposes.

Haefner (2005) classified biological models into four forms: verbal, diagrammatic, physical and formal. A verbal model is a description of a system using human languages. A diagrammatic model is a graphical representation, abstracting the interrelated objects and their relations into a diagram. A physical model is a mock-up of the real system. A formal model is a mathematical model expressed in mathematical languages, usually as a group of mathematical equations. When a mathematical model uses algorithms to organise its structure and requires computer programs to handle its implementation, it evolves to a new form – a computational model. Hill (2001) defines a computational model as "a set of computational codes, executable in some software/hardware environment, that transform a set of input data into a set of output data, with the input, output, and transformation typically having some interpretation in terms of real-world phenomena".

Biological systems are highly organised with multiple hierarchical levels progressing from atoms to the whole biosphere (Krogh, 2009) (Figure 3.1). An entity at a given level is a system composed by some of its lower-level objects, and in turn functions as an object for a higher-level system. Thus biological models, including models of plant development and function, are classifiable into these different levels (Prusinkiewicz, 1998). For example, the long-distance shoot-root framework of AON (Figure 2.3) is an inter-organ model, while the description of signal perception by LRR receptor kinase is an inter-cell model. If a model uses lower-level processes to investigate higher-level

phenomena, it is called a "bottom-up" model; if it views the system behaviour as a result of "a phenomenological relation" with external factors, it is called a "top-down" model (Haefner, 2005).

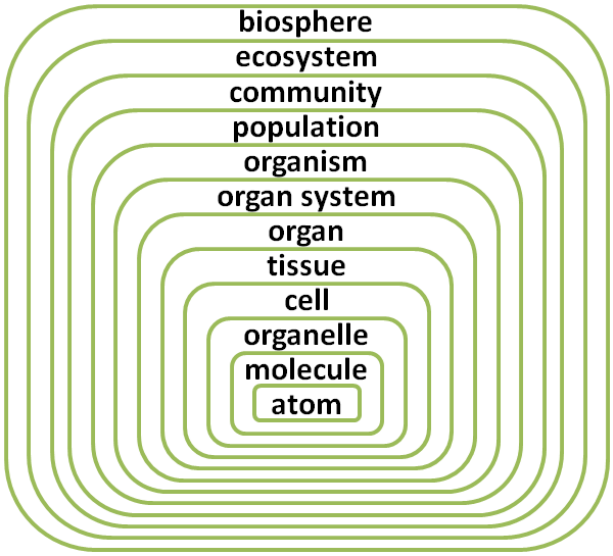


Figure 3.1 Hierarchy of organisation in biological systems. In this paradigm, the atoms are the most fundamental elements. At any higher level than the atoms, an entity can either be modelled as a system with lower-level objects or as one of the objects composing a higher-level system. This figure is redrawn from Krogh (2009).

A model can also be characterised as "empirical" or "mechanistic" depending on how it is built and thereby how it is used to study a system. If it simply describes the observed data or phenomena, it is an empirical or descriptive model; if it is process-oriented and represents the known or hypothesised mechanisms that cause the observed system behaviours, it is a mechanistic model (Haefner, 2005).

3.2 Plant architectural modelling

A plant can be viewed as a topological connection of different components or modules (Halle et al., 1978; Room et al., 1994; Hanan and Room, 1996), such as internodes, leaves and flowers, produced by apical meristems (for shoot growth) or root tips (for root growth). Each component has its own "shape, size, orientation and spatial location", referred to as geometric information (Godin et al., 1999). The organisation of plant components, represented by their topological and geometric attributes, is called plant architecture (Godin et al., 1999). To investigate the complex patterns and dynamic changes of plant architecture, computational modelling approaches have been developed since earlier works in the 1980s (Honda et al., 1981, 1982; de Reffye et al., 1988; Prusinkiewicz et al., 1988), forming a research domain known as plant architectural modelling.

Although plant architectural models have been widely built with different techniques or tools, such as L-systems (Prusinkiewicz and Lindenmayer, 1990), AMAP (Jaeger and de Reffye, 1992) and LIGNUM (Perttunen et al., 1996), they share the same underlying philosophy (Prusinkiewicz, 2004b): they all describe "a growing branching structure in terms of the activities of individual plant modules". For example, the plant shoot architecture can be modelled with different modules that are produced iteratively in hierarchical orders (Figure 3.2). Although the root architecture looks much more irregular than the shoot, it can also be decomposed into axes with different orders (e.g. primary root and lateral roots) and each axis can be decomposed into different segments (Danjon and Reubens, 2008).

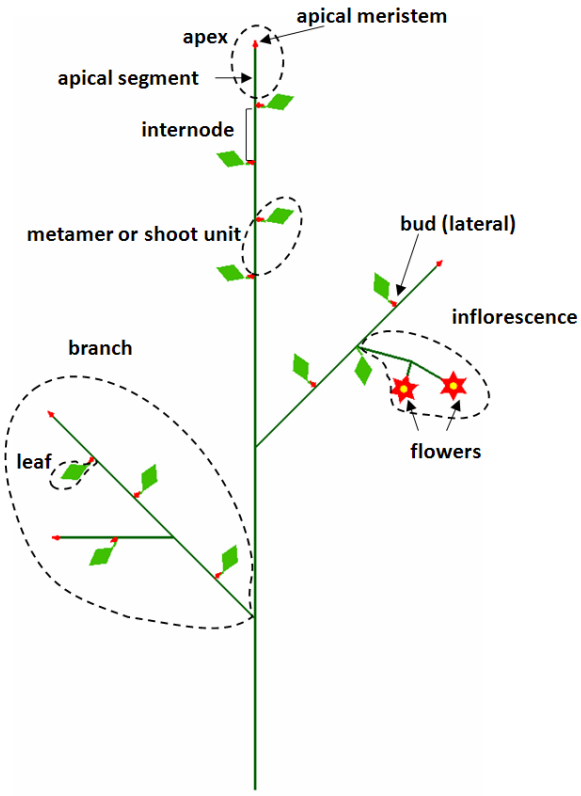


Figure 3.2. Representation of plant shoot architecture. In this architectural model, the shoot components, such as buds, leaves, internodes and flowers, are represented by relevant types of modules. The modules are produced iteratively, forming axes with similar patterns at different orders. This figure is redrawn from Prusinkiewicz (1998).

The representation of plant components can be coded into different types of data structures, such as a string of symbols (Prusinkiewicz, 1998), a list of elementary length units (Jourdan and Rey, 1997) or a multi-scale tree graph (Godin et al., 1999), to describe plant architecture (Figure 3.3). Each type of

data structure has its own advantage depending on different modelling emphases. The use of symbols (Figure 3.3 *A*) allows categorisation of architectural modules, so that all the modules under the same category can be processed in the same way. This reduces the number of growth rules and makes model specifications concise (Prusinkiewicz, 2004b). Instead of treating internodes as single modules (such as in Figure 3.3 *A*), the use of elementary length units divides an inter-branch section into lower-scale elements. This is more suitable to simulate the irregular and flexible architecture of a root. In the multi-scale tree graph paradigm, where a branching structure can be represented by different modules at multiple scales, an internode is also broken down into a number of successive components to address the diversity of directions and diameters observed in a tree axis.

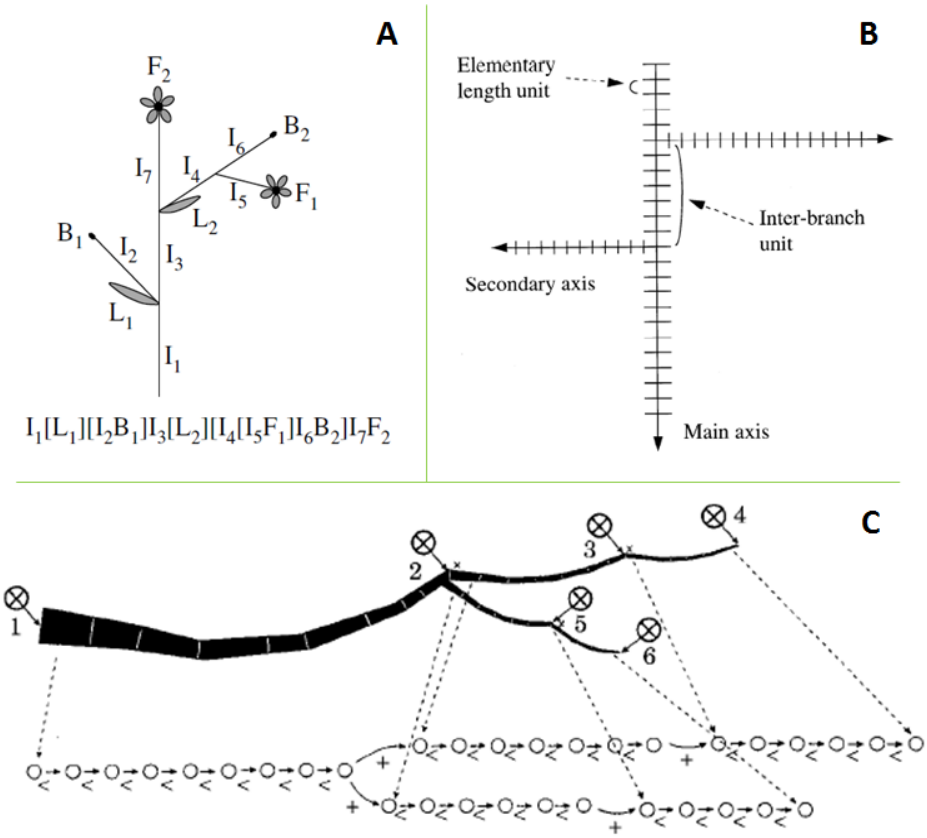


Figure 3.3. Data structures for coding plant representations. The shoot architecture in panel (A) is coded with a string of symbols, where each symbol represents a particular organ ("I" for internode, "L" for leaf, "B" for bud, and "F" for flower) and the symbols embraced by a pair of square brackets represent a branching structure. The root architecture in panel (B) is decomposed into axes and each axis is further split into a number of discrete elementary units of the same length. The multi-scale tree graph in panel (C) describes the successive and branching links between plant components, where "x < y" means component y is generated by the terminal bud of component x and "x + y" means component y branches from the axillary bud of component x. This figure is adapted from Prusinkiewicz (1998), Jourdan and Rey (1997) and Godin et al. (1999).

3.3 Functional-structural plant modelling

Far more than just topological and geometric rules underlie the complex behaviours of plant development. Plant architecture is also constrained by external factors and internal physiological processes (Godin and Sinoquet, 2005; Barthélémy and Caraglio, 2007; Vos et al., 2010), such as light environment, resource allocation and signalling regulation. Functional-structural plant modelling (Prusinkiewicz, 2004b; Godin and Sinoquet, 2005; Vos et al., 2010), also known as virtual plant modelling (Room et al., 1996; Hanan, 1997), integrates representation of plant function and structure in computational models.

In functional-structural plant models involving environmental factors, plant architecture serves as an interface between internal and external systems (Godin and Sinoquet, 2005): on one hand, the architectural components sense environmental changes that have further impacts on the endogenous physiological processes; on the other hand, the spatial distribution and dimension of these components modify the local environment. For example, a leaf not only captures radiant energy for photosynthesis and thereby provides a source of carbon, but also creates shade that affects the local light environment of its neighbouring leaves. This feedback characteristic is also addressed by the focus of functional-structural modelling on internal processes, where the interconnected architectural components provide a network for the function of physiological fluxes. Such a network includes where the fluxes start (the sources), where they pass through (the channels) and where they stop to work (the sinks). Plant development changes the networks (adding or dropping sources and sinks, extending channels, etc.), and the changed networks regulate plant growth in return. These complex interactions occurring over space and time lead to an emergent result: the appearance of a plant represented by its final architecture. Therefore, plant architecture is also a direct reporter of the underlying processes. This reporter can either be compared with data from real-plant experiments directly to test the reasonability of the hypothesised physiological mechanisms, or be used to predict the consequences of changes to environmental parameters.

There have been a variety of software tools developed for building functional-structural plant models. One of the mostly widely used platforms is L-studio (Prusinkiewicz, 2004a), which provides a plant modelling environment with two L-system-based (Lindenmayer, 1968; Prusinkiewicz and Lindenmayer, 1990) plant simulators: *cpfg* and *lpfg*. It not only allows users to define biological rules in the form of programs, but also facilitates 3D visualisation based on the step-by-step

implementation of these rules. The context-sensitive functionality of L-studio, in particular, enables the plant modules to exchange information with their neighbouring modules, thus allowing simulations of flows in the plant. The Linux-based Virtual Laboratory (VLab) has been developed by the same team as the Windows-based L-studio and supports most of its functionality (Prusinkiewicz, 2004a). Another L-system-based modelling package is GroIMP (Hemmerling et al., 2008), where XL is the programming language. Since XL is an extension of Java, which is cross-platform, the programs developed with GroIMP are not dependant on a specific operating system environment and can be executed on both Windows and Linux without changes. GREENLAB (Yan et al., 2004) is a functional-structural modelling tool that is based on the AMAP approach (Jaeger and de Reffye, 1992) rather than on L-systems. Compared with L-systems, GREENLAB also treats a plant as having modular composition and supports 3D visualisation, but integrates statistical functionality that makes it more straightforward for parameter optimisation. However, it is "simple in its physiological components" at present and needs further development to be "fully mechanistic with regards to physiological processes" (Yan et al., 2004). OpenAlea is an open-source software platform with ease of use, reusability and extendibility as well as collaborative development as its typical features (Pradal et al., 2008). Although OpenAlea "only partially addresses" the question "regarding the construction of comprehensive models that incorporates several aspects of plant functioning with intricate interactions between functions" (Pradal et al., 2008), it has the capability to integrate models built with different tools or languages. This advantage will make it easier to develop more comprehensive functional-structural models based on programs contributed by different researchers using different platforms.

Functional-structural modelling has experienced a rapid growth in the past two decades. Owing to close collaborations between plant and computer scientists, various approaches and tools focusing on different aspects of plant modelling and simulation have been developed. While much effort has been put into developing functional-structural models related to the external light environment (Chelle et al., 1998; Chelle and Andrieu, 1999; Chelle et al., 2004; Chelle et al., 2007; Cieslak et al., 2008) and internal resource allocation (Bidel et al., 2000; Drouet and Pagès, 2003; Allen et al., 2005; Drouet and Pagès, 2007; Lopez et al., 2008), the signalling function through plant architecture wait for better exploration by plant modellers (Vos et al., 2010). L-systems (Prusinkiewicz and Lindenmayer, 1990), with its mature techniques for architectural representation and context-sensitive information transfer, could serve as an ideal tool to integrate signalling and developmental processes. Further critical analysis of why new functional-structural modelling techniques and strategy need to be developed for this study is included in Section 4.1 and 5.1.

4 Development of techniques for modelling and simulation

Chapter 2 and Chapter 3 have introduced and discussed what is known about AON and the basis of computational plant modelling. In this chapter, two major challenges for using computational modelling to study AON at the technical level are addressed: the reconstruction of legume root architectural development with nodulation and the coordination of signalling and developmental processes. Section 4.2 firstly describes how the soybean root structure is mapped (using a group of 3D schematic figures) from real plant and then presents an elongation algorithm for simulation of the root development in computer space. Section 4.3 firstly demonstrates how signal transport through plant structure can be supported by context-sensitive L-systems and then uses two sample signalling/developmental events to illustrate the synchronisation algorithm developed in this study.

The content of this chapter has been published in Annals of Botany as "A Functional-structural Modelling Approach to Autoregulation of Nodulation" (Han et al., 2010a). The figures have been redrawn by the publisher in the published version. Updated from the published content, Section 4.2 is divided into two sub-sections: one for description of root mapping and the other for root simulation. Two footnotes – one for explanation of the branching probabilities in Section 4.2 and the other for further illustration of the synchronisation algorithm in Section 4.3 – are not included in the published version but added here to make these points clearer. Since the main purpose of Chapter 4 is only to introduce the new methods for meeting the technical challenges, the empirical data and computational flow charts for root reconstruction are not included in this chapter but described with full details in Appendix A.3 to support the virtual experiments used in Chapter 5.

4.1 Introduction

Nodulation is a developmental process that forms root nodules, resulting from the symbiosis between legume plants and a group of soil-living bacteria commonly called "rhizobia" (Carroll et al., 1985b; Oldroyd and Downie, 2004; Kinkema et al., 2006). This process fixes (incorporates) nitrogen from air within soil into ammonium that the plant can use for synthesising amino acids and nucleotides, providing roughly 200 million tons of fixed nitrogen to the ecosystem each year and

representing an environmentally friendly alternative to the use of synthetic nitrogen fertilizers (Graham and Vance, 2003; Gresshoff, 2003). However, excessive nodulation may disturb the resource allocation available for legume growth (Oka-Kira and Kawaguchi, 2006). To maintain the balance of nodulation, seedling legumes have developed a signalling regulatory system known as "autoregulation of nodulation" or "feedback inhibition of nodulation" (Carroll et al., 1985b; Delves et al., 1986; Caetano-Anollés and Gresshoff, 1991b; Oka-Kira and Kawaguchi, 2006). It has been hypothesised that a signal (Q) is induced by lipo-oligosaccharide induction of nodule primordia, which then moves from the root to the leaf vascular parenchyma (Gresshoff, 2003; Hayashi et al., 2008). A leucine-rich repeat (LRR) receptor kinase located in the phloem parenchyma of leaf vascular tissue (Nontachaiyapoom et al., 2007) – referred to as *GmNARK* in soybean (Searle et al., 2003; Miyahara et al., 2008), *HAR1* in *Lotus japonicas* (Krusell et al., 2002), and *SUNN* in *Medicago truncatula* (Schnabel et al., 2005) – is activated by the function of the Q signal and triggers the production of a shoot-derived inhibitor (SDI), which is transported to the root to inhibit further nodulation by inhibiting proliferation of early nodule primordia. Detailed mechanisms involved in autoregulation of nodulation, including signal production, transport, perception and function, as well as the identity of Q and SDI, are just evolving (Okamoto et al., 2009; Lin et al., 2010; Mortier et al., 2010). The purpose of our research is to use computer modelling and simulation to help in investigating these complexities as well as dynamic interactions with lateral root development (Beveridge et al., 2007).

Since autoregulation of nodulation is essentially a long-distance inter-organ regulatory network, our modelling efforts focus on the organ-scale signalling mechanisms. Functional-structural plant models, which link 'spatialisation of processes in plant functioning and morphogenesis' (Godin and Sinoquet, 2005), provide an ideal method for this study. With functional-structural modelling, we can simulate the hypothesised signalling mechanisms that are initiated by and affect plant organ development, and then use the regulated plant architecture as a direct reporter to evaluate these hypotheses. At the technical level, there are two major challenges for the application of functional-structural modelling to investigate autoregulation of nodulation: reconstruction of the 3D architecture of legume roots and coordination of the multi-rate signalling-development processes.

Unlike the clearly ordered composition and growth of shoots, the root system is difficult to observe and has much more complex patterns. In the past two decades, plant modellers have developed various approaches for collecting root architectural data, handling 3D simulation of root topology and geometry as well as modelling interactions between root growth and environmental factors

(Diggle, 1988; Danjon and Reubens, 2008). Most previous efforts focused on woody roots, while legume roots, particularly with nodulation as a distinguishing feature, have not drawn much attention. Since the number of nodules and their distribution are the main phenotypic aspects for studying the underlying regulatory system (Carroll et al., 1985a, b; Delves et al., 1986; Caetano-Anollés and Gresshoff, 1990, 1991a; Gresshoff, 2003; van Noorden et al., 2006), collecting empirical data on these attributes and reflecting them in the architectural model are crucial to our work. Non-destructive or automatic technologies for root data collection, including those using Ground Penetrating Radar, CT imaging and 3D laser scanning (Danjon and Reubens, 2008), still have limitations in resolution and branch detection or have restrictions due to environmental conditions, which lowers their effectiveness or feasibility in application to this study. The semi-automatic scanning method described by Lira and Smith (2000) could help in counting nodules in high-resolution root images and could possibly be improved to recognise nodules automatically. But the reliance on 2D images taken from one angle means that the 3D lateral branching cannot be well-classified by this technology and some nodules hidden by primary or lateral roots would be ignored. On the other hand, 3D digitising approaches (Room et al., 1996; Sinoquet and Rivet, 1997) could position plant organs precisely. However, legume roots are usually highly flexible and therefore change their 3D position once removed from the soil, which makes it difficult for digitisation to capture their spatial patterns.

Recreating the architecture is not enough; an architectural model is more useful when it helps to reveal internal and environmental factors that influence its development, which requires the linking of plant structure and function. Most previous functional-structural models of root development (Danjon and Reubens, 2008) take environmental factors into account. However, the root architecture is ‘shaped’ not only by environmental factors but also by the influence of endogenous signals such as hormonal stimuli (Aloni et al., 2006). For functional-structural modelling of internal signalling control, the stage has been set by previous work (Janssen and Lindenmayer, 1987; Prusinkiewicz and Lindenmayer, 1990; Buck-Sorlin et al., 2005; Buck-Sorlin et al., 2008; Lucas et al., 2008) using discrete information transfer based on context-sensitive L-systems. We build on this work through the integration of internal signal regulation with root architectural details, the visualisation of signal allocation in the root system and, in particular, the synchronisation of signalling and developmental processes with various empirical or hypothetical rates.

In this paper, we choose soybean (*Glycine max* L. Merrill) as the target legume plant and use context-sensitive L-systems (Prusinkiewicz and Lindenmayer, 1990) as the modelling tool to show our current methods in meeting these challenges.

4.2 Root development reconstruction

4.2.1 Mapping root structure

For the investigated root system, we classify its components as primary root, first-order lateral roots and nodules. The second-order laterals were not considered or modelled. When observing the soybean root from above (Figure 4.1), we found a regular branching pattern of the lateral roots: the radial angles of lateral emission, as defined by Jourdan and Rey (1997), are usually around 90^{o}. This pattern is suggested to be the result of lateral formation opposite xylem poles (Bell and McCully, 1970; Mallory et al., 1970; Abadia-Fenoll et al., 1982; Rolfe and Gresshoff, 1988; Jourdan and Rey, 1997).

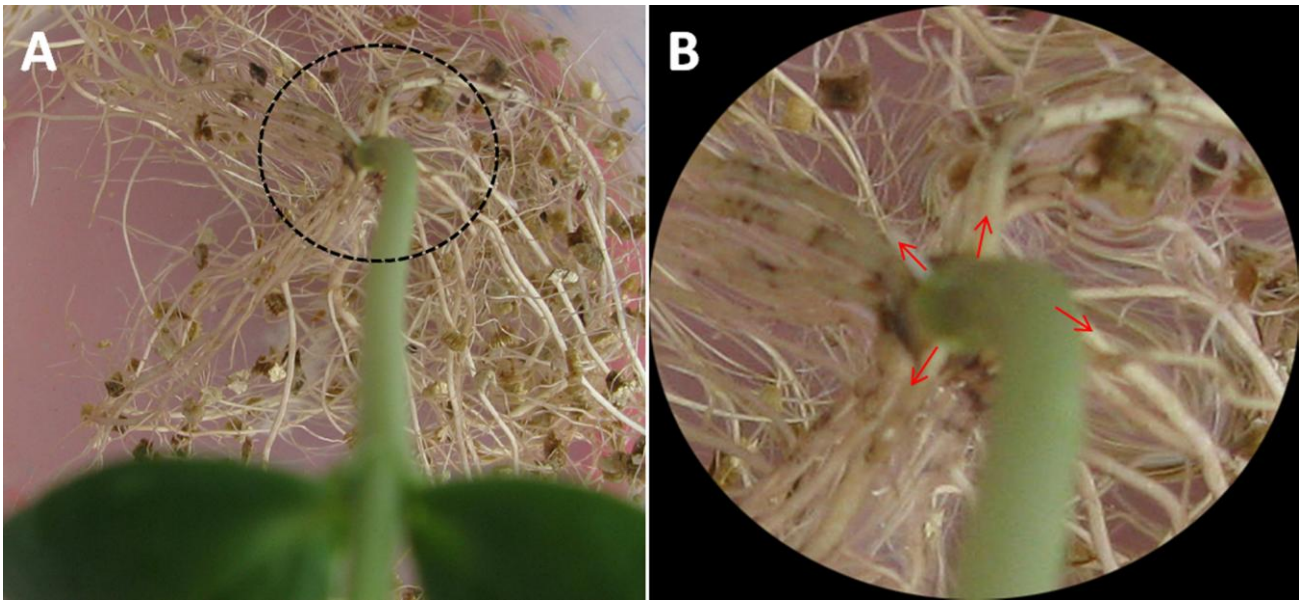


Figure 4.1. Observation of soybean root branching pattern. From an overhead view (A), the radial angles of lateral root emission are usually around 90^{o} (B), demonstrating a regular pattern.

To characterise lateral roots based on this emission pattern and collect their developmental data, we developed a "RULD" root mapping method (Han et al., 2007). With the "RULD" method, the first-order laterals are categorised into quadrants (right-R, up-U, left-L and down-D), according to the relative positions of their emission points to the obtuse angle composed by the two cotyledons

(Figure 4.2). Then the lateral roots are further classified and identified by “regions” (a 50mm-long section on the primary root), “segments” and “sites” (Figure 4.3).

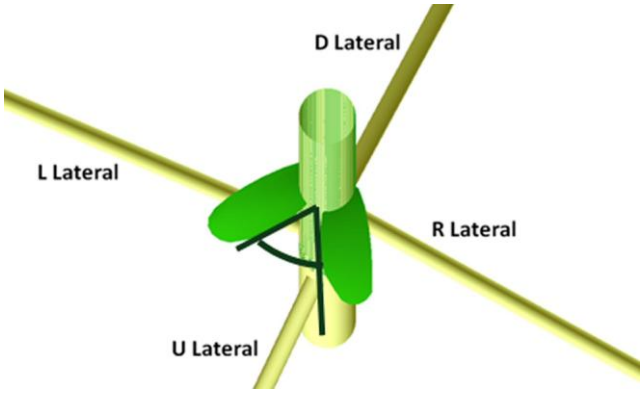


Figure 4.2. “RULD” mapping to characterise first-order lateral roots. For a growing soybean plant at early stages, its two cotyledons compose an obtuse angle in the horizontal plane. The lateral roots can be categorised as being in one of four quadrants – R (right), U (up), L (left) and D (down) – according to the position of their emission point relative to this obtuse angle.

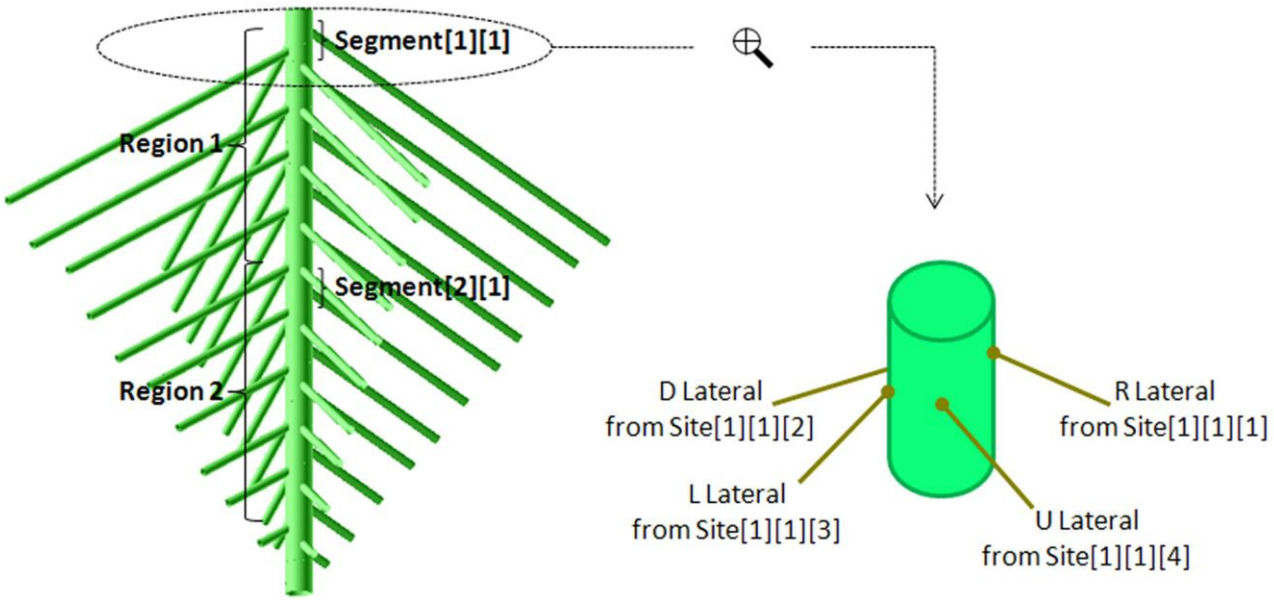


Figure 4.3. Identification of lateral root based on its formation sites. A region is defined as a 50mm-long primary root section. A segment is defined as a smaller area with four R-U-L-D quadrants in a region. Thus the formation site of each lateral root can be identified by the region and segment it is located in as well as its R-U-L-D side. Numbering in square brackets proceeds downwards, segments are numbered within regions, and sites within segment.

To capture nodule distribution information, we define a “nodulation section” (Han et al., 2009) for each region of the primary root and for each first-order lateral root (Figure 4.4). The nodulation

section covers the distance between the first and the last nodules in a particular primary-root region or on a lateral root. The relative locations of the first and the last nodules to the starting point of the primary or the lateral root are measured to position each nodulation section. The length of a nodulation section and its number of nodules are used to calculate the nodule density for this section.

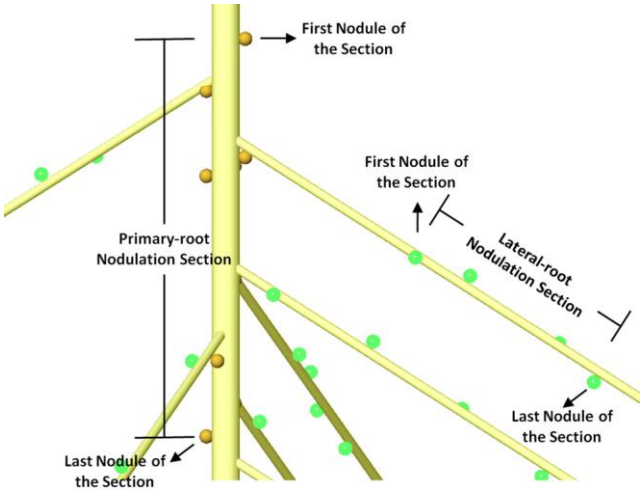


Figure 4.4. Definition of nodulation section. Each region of the primary root and each lateral root is defined to have its unique nodulation section. The positions of the first and the last nodules of a nodulation section determine its location in the root system. The length of a nodulation section and the number of nodules it includes are used to calculate the interval between two successive nodules in this section.

4.2.2 Simulation of root elongation

The empirical architectural data collected with the above measurement methods are then used to drive the root simulation. The primary root architecture is extended by the root tip, where potential positions for nodulation or lateral emission are also created. The root elongation has an appropriate rate to match the empirical length value, which defines the corresponding root tip location. There are three major rules for modelling the primary root elongation:

(1) When the root length indicates that the primary root has entered a section for future nodulation, the model starts checking whether the root tip should be marked as a potential nodule formation site. If the root tip location matched the positions of the first or last nodules (of this section) or if the root elongation since the last marked position is longer than the interval between two successive nodules, a potential site for nodule formation is made available.

(2) If the primary root elongation has covered an interval between two successive laterals, a potential site for lateral root emission is made available.
(3) If the current position of primary root tip does not match the conditions in (1) and (2), it has neither potential for nodulation nor for lateral emission.

The potential sites pre-set by Rules (1) and (2) are only spatial markers; whether and when nodules or lateral roots will be formed there depend on other conditions derived from empirical data. The maximum number of nodules formed in each primary root region during each day is used to restrict further nodulation. At a potential nodulation site, a nodule will be formed if and only if the relevant maximum number of nodules is not exceeded. Whether a lateral root will be emitted from a potential formation site is determined by branching probabilities[1], while how soon the lateral will appear and how far it can elongate are controlled by the empirical daily growth rate. The apical zone around the root tip in the architectural model has no nodules attached and remains unbranched, as the maximum nodule number and the lateral growth rate for this area are both 0.

The architectural model was built with an L-system-based language "*cpfg*" in L-studio (Prusinkiewicz and Lindenmayer, 1990; Hanan, 1997; Prusinkiewicz, 2004a). In L-system plant models, a shoot organ is usually represented by a single module that not only contains its developmental information but also simulates its structure graphically. A number of such modules are then assembled into a string to represent the entire shoot architecture composed of different organs. The dynamics of plant development is supported by step-by-step application of a set of rules called "productions". The productions are usually written as

predecessor : (condition) → *successor*

where "predecessor" is an original module while "successor" is a module or a string of modules to replace the predecessor as long as the "condition" is matched. At each simulation step, the modules in the current string are matched against the productions and only those with matched predecessor and condition are applied. If multiple productions have the same predecessor, they are checked sequentially and only the first one with matched condition is allowed to produce its successor. After all modules have been processed, the resulting string represents the structure of the plant at the end of the time step.

[1] The branching probabilities are calculated using the number of laterals in a region from one side (R, U, L or D) divided by the number of segments in this region. For example, if the number of R laterals in a region i is NR_i and the number of segments is NS_i, then the branching probability for R laterals in this region is NR_i / NS_i.

In this root study, the elongation information (such as overall root length already created) is recorded within a module representing the root tip, while the graphical role is played by a set of "sub-modules" each with the same length (UNIT). The standard sub-modules here are similar to the elementary units used to model oil-palm root system (Jourdan and Rey, 1997). During root elongation, a sub-module is added behind the root tip module at each simulation step, forming a string of sub-modules to make up root structure. For example, if we use "RT" to represent root tip, "R" to represent a sub-module, the initial structural string is "RT" and the production is

$$RT \rightarrow R\ RT$$

then the string will become "R RT" after one step and "R R RT" after two steps. The sub-module length – UNIT – is definable by the modellers or users depending on specific requirements in specific cases. In this study, to allow the increment of root length to match conditions for setting potential nodulation or lateral formation sites, the value of UNIT should be smaller than the minimum interval between two successive nodules or laterals. For details of the time scale, see the next section. The algorithm for primary root elongation with sub-modules is illustrated in Figure 4.5. Its implementation using the "*cpfg*" language is given in Text S1 (Appendix A.1). Although all the sub-modules have the same length, their widths vary and increase with radial growth, capturing appropriate diameters at different stages of root development. The lateral root elongation is simulated in a similar way except that the second-order laterals are not taken into account. The simulation of root heading behaviours is supported by methods from the ROOTMAP model (Diggle, 1988) and is representative only. Since the plants used for studying autoregulation of nodulation in this research are at their early developmental stages, the mortality of lateral root tips is not considered. A sample visualisation of soybean root architecture with nodulation is given in Figure 4.6 and Video S6 (Appendix A.6).

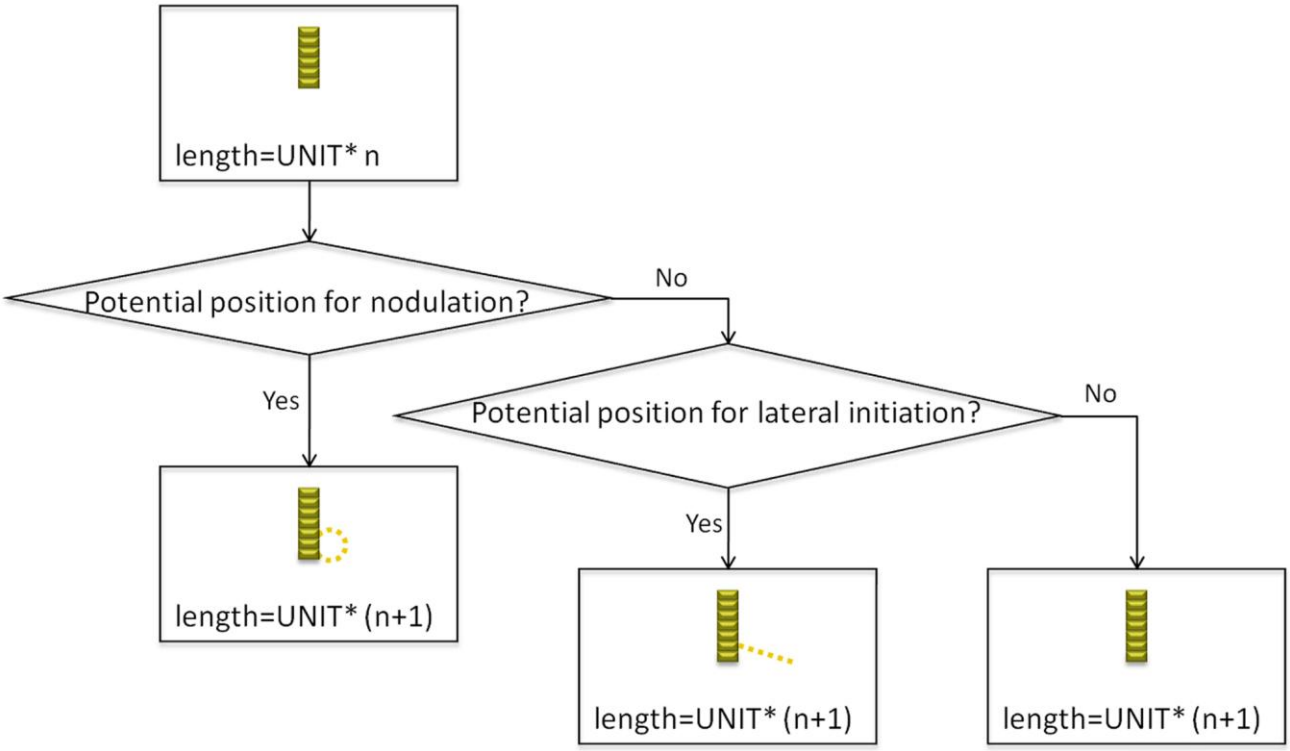


Figure 4.5. Root elongation algorithm. At the beginning of each L-system time step during root elongation, the root tip module checks whether its current location has potential for future nodulation or lateral initiation. If it matches potential positions for nodule or lateral formation, such positions are made available and a sub-module is added to elongate the root structure with one UNIT.

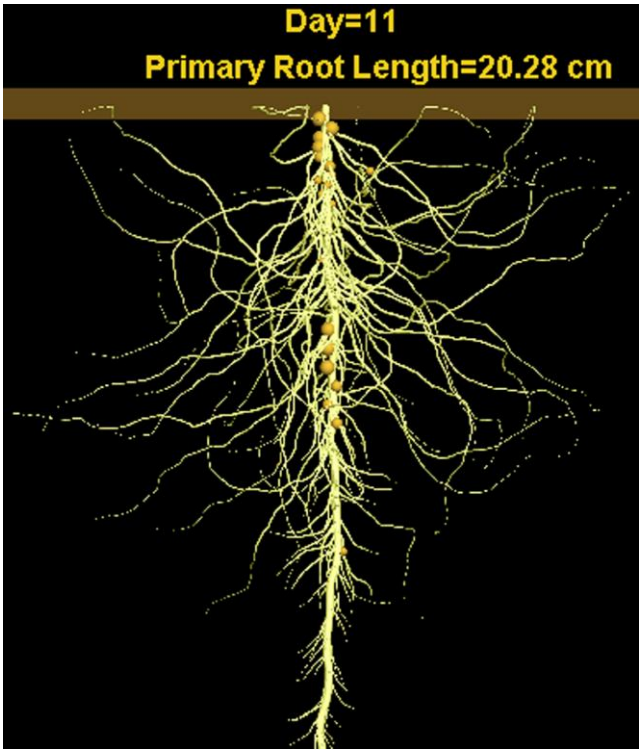


Figure 4.6. A sample visualisation of soybean root architecture with nodulation. The day-by-day developmental process leading to this root architecture is demonstrated in Video S6 (Appendix A.6).

To evaluate the root reconstruction, growth data of a soybean hypernodulation genotype *nts1116* (Carroll et al., 1985b; mutated at V837 of the *GmNARK* receptor kinase gene) are incorporated into the architectural model. *nts1116* was shown to have severely reduced *in vitro* kinase activity (Miyahara et al., 2008), consistent with a 3x elevated nodule number. The *nts1116* plants were cultured in glasshouse conditions over a 16-day period, inoculated on the second day, with five plants sampled for destructive measurements every two days starting on the third day. The comparison of real-plant *vs* virtual-plant root systems, on lateral root branching and nodule distribution (Han et al., 2009) as well as on root length, indicates a good fit between empirical data and simulation results (Figure 4.7).

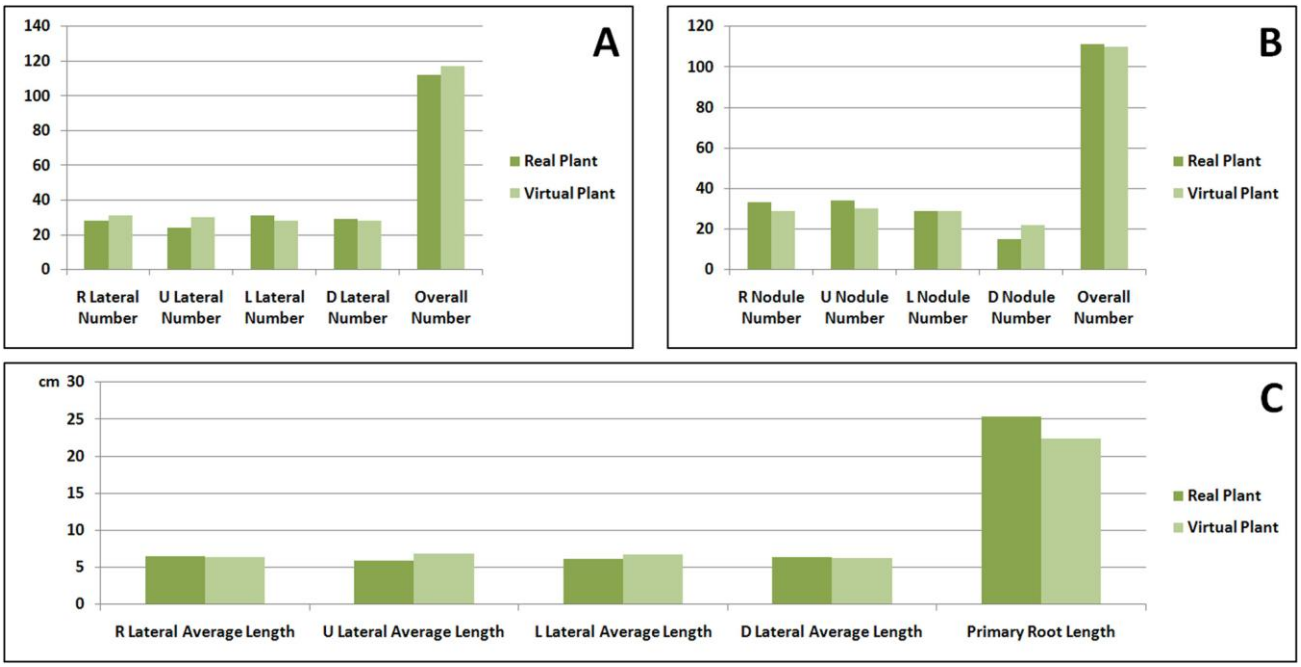


Figure 4.7. Comparison of empirical data with simulation results. The virtual-plant architecture, driven by empirical data, is compared with the real plant on the 16th day after sowing. For lateral root branching (A), the comparison classified by the R-U-L-D scheme as well as the overall number demonstrates a good fit (Han et al., 2009, Copyright 2009 IEEE). The simulated nodule distribution, represented by numbers of nodules located on R-U-L-D laterals (B), is close to the empirical data (Han et al., 2009, Copyright 2009 IEEE). For root elongation, the average lengths of R-U-L-D laterals as well as the primary root length (C) of virtual plant also demonstrates a good fit with the real plant.

4.3 Integration and synchronisation of signalling and development

Using information transfer in context-sensitive L-systems (Prusinkiewicz and Lindenmayer, 1990), the signal movement from one sub-module to its neighbour can be incorporated into the plant architectural model to upgrade it to a functional-structural model. To do this, each sub-module is given a parameter to represent a particular signal's concentration. For example, when the concentration level of signal in a sub-module meets a certain threshold, the value of this signal's amount will be passed to the concentration parameter in the next sub-module (Figure 4.8).

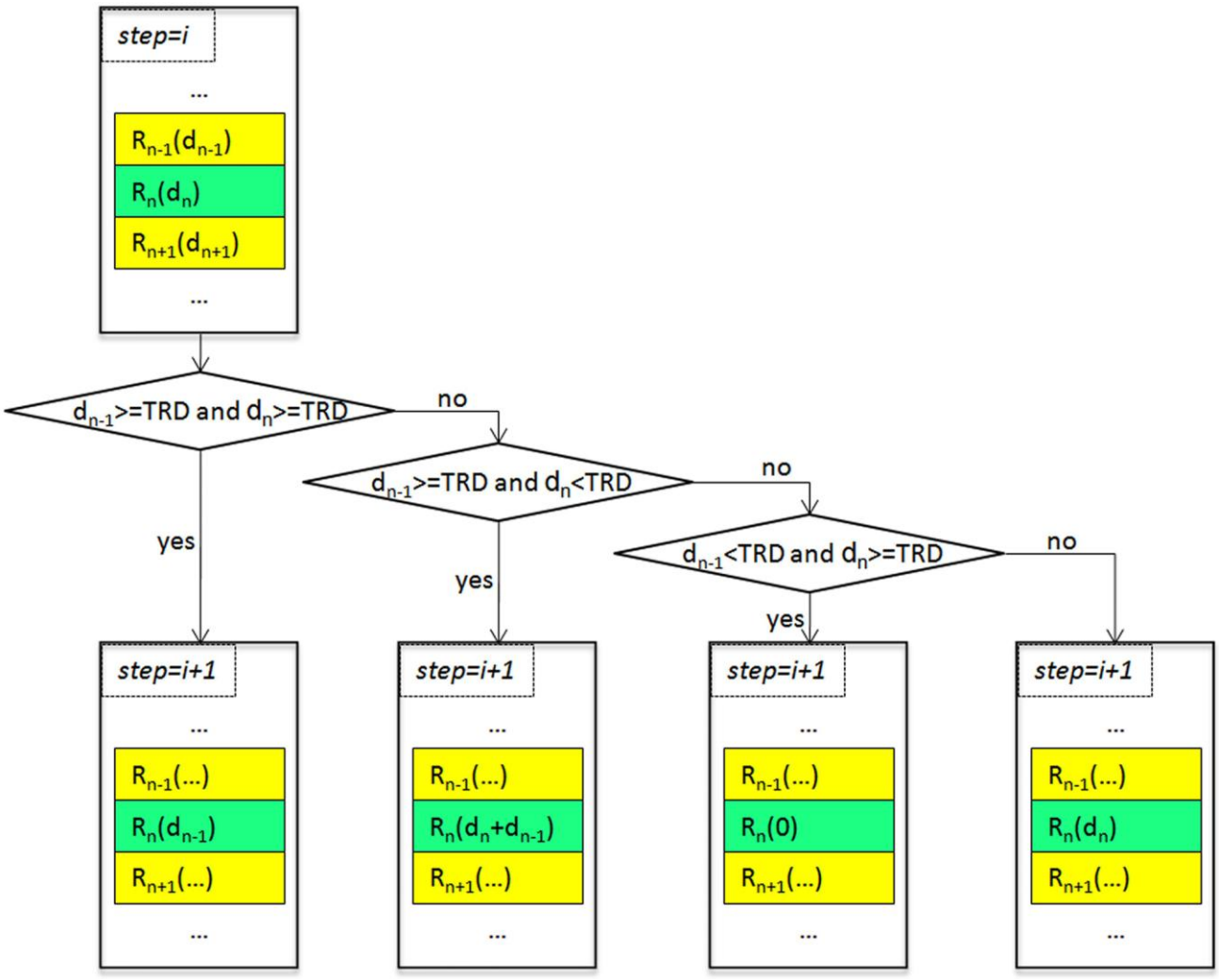


Figure 4.8. Signal transport supported by information transfer within context-sensitive L-systems. R_{n-1}, R_n, and R_{n+1} are three neighbouring sub-modules with signal concentration levels of d_{n-1}, d_n and d_{n+1} respectively. Assuming the direction of signal movement is from R_{n-1} to R_{n+1}, here we focus on changes in R_n to illustrate how the signal is transferred. The value of d_n is checked and compared with a threshold defined as "TRD" at every step during signal transport. If d_n is over TRD, its value will be passed by subtracting it from d_n and adding it to d_{n+1}; otherwise no signal will be moved out of R_n. The same rule is applied to R_{n-1} and R_{n+1} simultaneously.

Since there is more than one signal involved in autoregulation of nodulation and the signalling and developmental processes have various rates, a coordination mechanism to synchronise these multi-rate processes has been developed. Firstly, each signalling or developmental event has two states: "activated" and "stopped". During a single L-system time step, only one sub-module can be added to the current structure or be passed through for signal transport if the relevant elongation or signalling event is activated. When their states are marked as "stopped", those signalling or developmental events do not occur and will wait to become activated, independently. The difficulty is in switching

between these two states for each event so that the signalling-development processes can be dispatched in a synchronised way. In order to achieve this, we define a concept of "time division" that divides one day's time into lower-scale time sections (e.g. we can divide one day's time into 24 hours – each hour is represented by such a division). Each signalling or developmental event e_i is assigned a certain number of time steps, defined as c_i, during each time division. Assuming the rate of e_i is r_i and the number of time divisions per day is DIV, the value of c_i can be calculated using Equation 4.1:

$$c_i=\text{ceil}(r_i /(\text{DIV}*\text{UNIT})) \quad (4.1)$$

where ceil(x) is a function returning the smallest integer value not less than x. The number of time steps allocated to the fastest signalling or developmental event during a division, equivalent to $\max\{c_i|i=1,2,\ldots,n\}$, is the total number of time steps this division consists of. That is to say, the event with the fastest rate keeps occurring over all time steps during the division, while the events with slower rates are only activated over a smaller number of steps and are "stopped" during the remaining steps of a division. When a new division is initialised, all the stopped events will become activated again. Therefore all processes are synchronised at the beginning of the division. The synchronisation process is illustrated with two sample events in Figure 4.9. This strategy allows as many signalling-development events to be involved and synchronised as possible.

With the integrated and synchronised signalling and developmental processes, the functional-structural model of autoregulation of nodulation can produce different signal allocation and regulation patterns based on different parameter settings (Figure 4.10). For example, by colouring each sub-module according to its SDI concentration value, the allocation of SDI at different parts of the root can be visualised in detail (Figure 4.10 *A*, *B* and *C*), while the inhibited nodules (that are not apparent in nature) can either be visualised (Figure 4.10 *D*, *E* and *F*) or filtered (Figure 4.10 *G*, *H* and *I*) to help in analysing the nodulation pattern. These functionalities could be used as the basis of virtual experiments to investigate unknown or unclear attributes of autoregulation of nodulation.

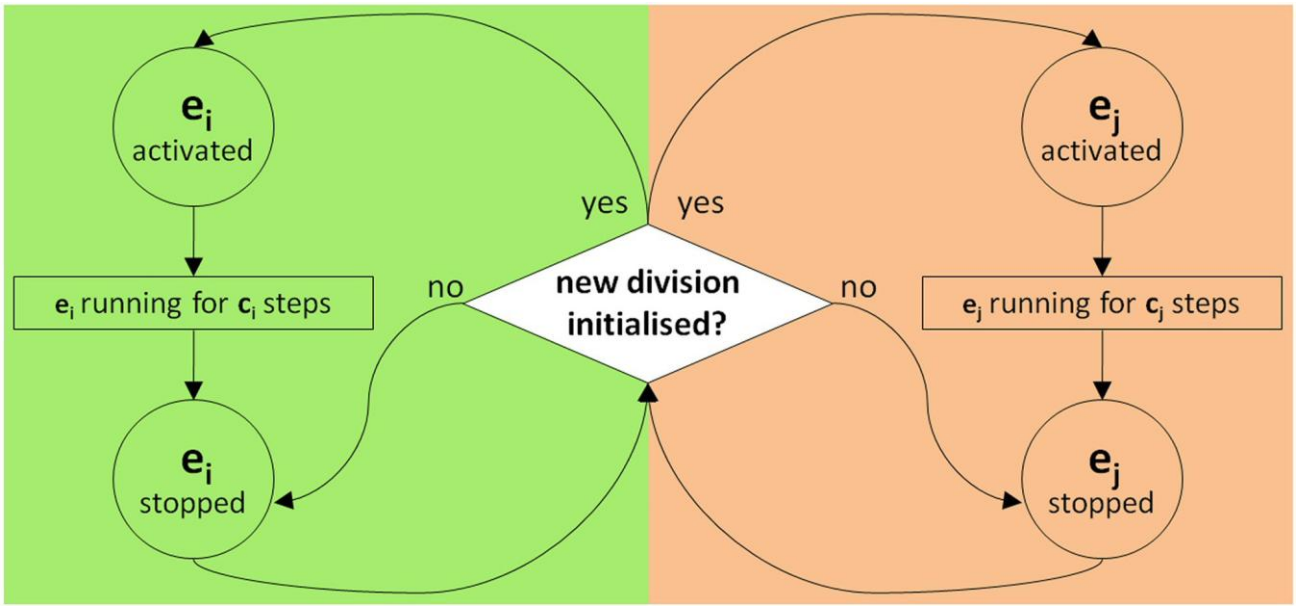


Figure 4.9. Synchronisation of signalling and developmental events. In this illustration, two independent events e_i (with green background, representing a developmental event) and e_j (with orange background, representing a signalling event) are used as an example to show how multiple processes with different rates are synchronised. The initial states of all the events are "activated". During a time division, e_i is switched from "activated" state to "stopped" state when it uses up the c_i steps allocated to it while e_j is switched after running for c_j steps. All events will be activated again when a new time division is initialised[2].

[2] Assuming a division has five steps, during which e_i is assigned with two step (c_i=2) and e_j with three steps (c_j=3), then both e_i and e_j are activated at the first two steps; e_i is stopped and e_j keeps running at the third step; and both of them are stopped during the remaining two steps of the current division until the next division is initialised.

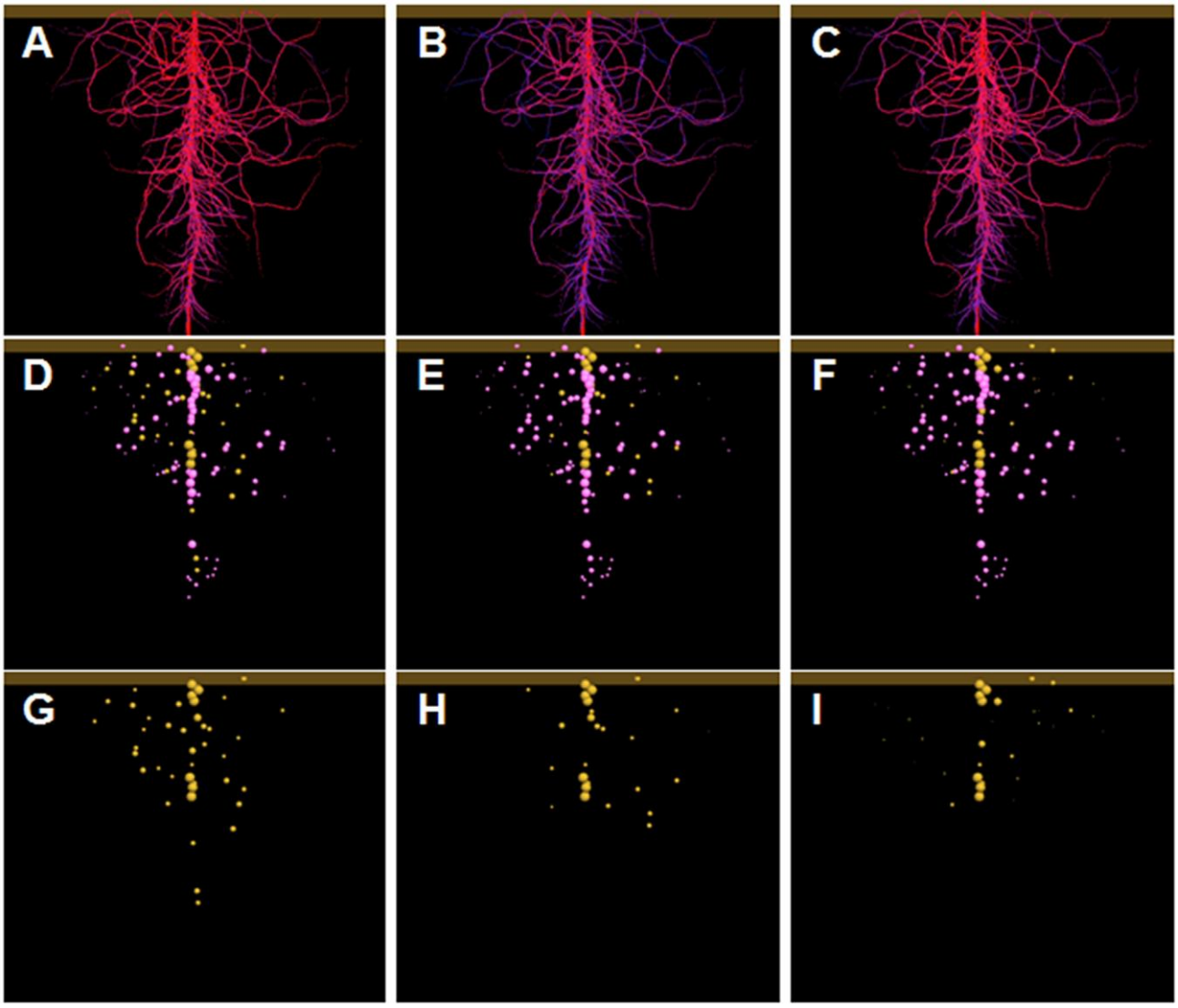


Figure 4.10. Illustration of virtual-experiment outputs. Assuming the transport rate of Q signal is 350 mm/day, different signalling allocation and regulation patterns were obtained by setting the transport rate of SDI with three different values: 100 mm/day (resulting in images A, D and G), 200 mm/day (resulting in images B, E and H) and 300 mm/day (resulting in images C, F and I). In A, B and C, the colours varying from blue to red represent lower to higher concentration of SDI. In D, E and F, the yellow nodules are developed nodules, while the purple ones are inhibited nodules due to autoregulation of nodulation. The inhibited nodules could also be filtered (G, H and I) to allow comparison with nodule distribution in real plants.

4.4 Discussion and conclusion

Technologies for root data collection and methods for root structure reconstruction have been widely developed in the past two decades. A major effort has been directed toward classifying and recreating lateral root branching patterns. As part of these pursuits, we have developed the "RULD"

root mapping method to characterise the patterns of soybean lateral root position. To simulate the topological development of soybean root system, we also developed a root elongation algorithm based on the use of standard sub-modules that are user-definable and capable of supporting internal signalling activities. This has been implemented using an L-system model and evaluated against the real-plant data.

Yet for legume plants, the root architecture is composed not only of primary and lateral roots, but also of nodules. The nodules play a critical role in supporting legume plants' growth and in providing fixed-nitrogen to the ecosystem, thus must not be ignored. In this paper, we have presented our current methods to capture nodulation patterns from the plants and to recreate them in computational models.

The reconstruction techniques developed in this study have some limitations. The data collection and model evaluation are based on young soybean roots, where the second-order laterals and mortality of root tips are not considered. These aspects of root reconstruction will need to be addressed if the study is expanded to older plants grown in field conditions. At this stage, the modelling focus has been on soybean. For other legume roots, the primary root elongation algorithm is also applicable, but modification of the data collection method may be required depending on whether the first-order laterals there can be characterised into quadrants (or whether the radial angles of lateral emission are around 90^{o}).

The observable plant structure serves to report the unobservable regulation mechanisms. Using functional-structural modelling, we have integrated long-distance signalling for autoregulation of nodulation with architectural development. In order to effectively coordinate the signalling and developmental processes with various rates, we also developed a synchronisation algorithm based on the use of sub-module structure and context-sensitive L-systems.

The functional-structural model, integrating root architectural details with signalling control, enables parameterisation of signalling hypotheses and produces regulation patterns within different forms (e.g. signal allocation and nodule distribution). This provides a basis for implementation of virtual experiments to analyse or evaluate mechanisms of autoregulation of nodulation that are unclear.

Looking into the future, the modelling methods developed here are not restricted to the regulation of nodules. We also see a potential for applying or combining these technologies in wider studies of

root systems, such as those on lateral root initiation (Dubrovsky et al., 2006) and other types of regulation based on signalling (Aloni et al., 2006; Lucas et al., 2008). Models of carbon allocation and water flow could also be integrated, allowing a system-level view of plant development and function.

4.5 Supplementary information

The implementation of soybean primary root elongation algorithm within "*cpfg*" syntax is described in Text S1 (Appendix A.1). A sample visualisation of recreated root architectural development with nodulation is given in Video S6 (Appendix A.6).

4.6 Acknowledgements

This study has been supported by the Australian Research Council Centre of Excellence for Integrative Legume Research (CILR), the Australian Research Council Centre for Complex Systems (ACCS), the School of Information Technology and Electrical Engineering (ITEE) based at the University of Queensland (UQ), and a UQIRTA award with a UQRS scholarship awarded to Liqi Han by the UQ Graduate School. We also thank Kuang-Hsu Wu from University of Queensland for his participation in collecting soybean architectural data.

5 Approach: computational complementation

The modelling and simulation methods presented in Chapter 4 provide the technical basis underlying the use of computational modelling to study AON. At the strategic level, the challenge is how to use those techniques to investigate AON mechanisms. In this chapter, a new modelling strategy developed in this study – computational complementation – is introduced. Section 5.2 describes the general methodology of computational complementation, and then Section 5.3 evaluates the feasibility of this approach through its first application to test whether or not wild-type soybean cotyledons provide the SDI signal to regulate nodule progression.

The content of this chapter has been published in PLoS Computational Biology as "Computational Complementation: A Modelling Approach to Study Signalling Mechanisms during Legume Autoregulation of Nodulation" (Han et al., 2010b). The footnote in Section 5.2 was not included in the published version but added here to clarify the relationship as well as the difference between the complementation model and the mutant empirical model. Neither was the footnote in Section 5.3 included in the published version, which is used here to elucidate why homografts were not used in this study. Another footnote is added to Section 5.4 to better explain the role of using computational complementation in identification of Q and SDI signals. The materials and methods for glasshouse experiments mentioned in this chapter are described in Appendix A.2. The details of empirical data and computational flow charts to establish the virtual experiments are explained in Appendix A.3. And the assumptions as well as conditions for those virtual experiments are listed in Appendix A.4.

5.1 Introduction

Legumes are one of the largest families of flowering plants that occupy about 15% of Earth's arable surface; yet they provide 27% of the world's primary crop production and more than 35% of the world's processed vegetable oil (Graham and Vance, 2003), signifying their cropping potential. Legumes are also the major natural nitrogen-provider to the ecosystem, contributing roughly 200 million tons of nitrogen each year (Kinkema et al., 2006) equivalent to over 200 billion dollars worth of fertiliser replacement value. Underlying this powerful fixation capability is a plant developmental process termed "nodulation", which results from the symbiosis of legume roots and soil-living

bacteria broadly called rhizobia. Yet for a legume plant itself, excessive nodulation may cause over-consumption of metabolic resources and disproportional distribution of internal growth regulators (Oka-Kira and Kawaguchi, 2006), and may interfere with developmentally related lateral root inception and function.

Legume plants have evolved a long-distance systemic signalling regulatory system, known as autoregulation of nodulation (AON), to maintain the balance of nodule formation (Carroll et al., 1985a, b; Delves et al., 1986; Gresshoff, 2003; Oka-Kira and Kawaguchi, 2006). It has been hypothesised that the induction of the nodule primordium produces a translocatable signal Q, which moves through a root-shoot xylem pathway to the leaves. This Q signal, or an intermediate, is detected in the phloem parenchyma of leaf vascular tissue by a transmembrane leucine-rich repeat (LRR) receptor kinase (Nontachaiyapoom et al., 2007) related in structure to CLAVATA1 in Arabidopsis. This kinase is referred to as *GmNARK* in soybean (Searle et al., 2003; Miyahara et al., 2008), *HAR1* in Lotus (Krusell et al., 2002), and *SUNN* in Medicago (Schnabel et al., 2005). Q is presumed to be a CLV3/ESR-related (CLE) peptide (Gresshoff et al., 2009; Okamoto et al., 2009). The perception of the Q signal by the LRR receptor kinase triggers production of a hypothetical shoot-derived inhibitor (SDI) that is transported to the root to inhibit further nodule initiation. SDI can be extracted from wild-type leaves, re-fed via petiole feeding into loss-of-function mutants, resulting in restoration of the wild-type phenotypes (Lin et al., 2010). It is a small, water-soluble, heat-stable and inoculation-dependent molecule**.** However, other mechanisms involved in AON signalling remain largely unknown, though the pre-NARK events (those setting up the signal transmission and then Q signal transduction) as well as the post-NARK events (firstly KAPP phosphorylation, ensuing transcriptional changes, and then SDI production) are being investigated (Kinkema and Gresshoff, 2008; Miyahara et al., 2008; Lin et al., 2010).

To help understand such biological complexities, system modelling has been broadly applied (Kitano, 2002; Minorsky, 2003; Hammer et al., 2004). From a systematic view, behind the signalling mechanisms is a network of components connected by intricate interfaces, with activities such as "assembly, translocation, degradation, and channelling of chemical reactions" occurring simultaneously (Weng et al., 1999). These components and their interactions – also responding to the temporally and spatially changing environment – frame dynamic and complex systems at multiple scales to orchestrate plant development and behaviour. As a full understanding of system properties emerging from component interactions cannot be achieved only by "drawing diagrams of

their interconnections" (Kitano, 2002), computational techniques become indispensable for processing massive datasets and simulating complex mechanisms (Neves and Iyengar, 2002).

Although computational approaches have been progressing rapidly for modelling plant signalling, such as for signal transport (Jönsson et al., 2006; Berleth et al., 2007), canalization (Rolland-Lagan and Prusinkiewicz, 2005) and signalling network (de Reuille et al., 2006), most efforts have focused on cellular or tissue levels. Since AON is in essence a long-distance inter-organ regulatory network, our investigation required modelling at the whole-plant scale. Functional-structural plant models (Godin and Sinoquet, 2005), such as those developed for resource allocation (Bidel et al., 2000; Allen et al., 2005; Drouet and Pagès, 2007) and shoot signalling (Janssen and Lindenmayer, 1987; Prusinkiewicz and Lindenmayer, 1990; Buck-Sorlin et al., 2005; Prusinkiewicz et al., 2009), can take inter-organ communication into account and use plant architecture as a direct reporter of underlying processes. Functional-structural modelling allowed us to simulate the hypothesised AON signalling and integrate it with nodulation. Yet the major difficulty was not how to model the hypotheses but how to test them through modelling. To meet this challenge, we have developed a new approach – Computational Complementation – for AON study.

Following description of the computational complementation method, we will present its first application in investigating whether wild-type cotyledons participate as an SDI producer in the AON system. Previous studies have indicated that mRNA for *GmNARK*, which, if translated, is responsible for perceiving the Q signal and triggering the SDI signal, exists in wild-type unifoliate and trifoliate leaves. It is expressed in all vascular tissue (Nontachaiyapoom et al., 2007) of the plant (including the root), but its product is functional only as a nodulation control receptor in the leaf (Delves et al., 1992). Thus the RNA expression pattern does not match biological function in AON. Relevant to the investigation here, the vasculature of the cotyledon also expresses RNA for *GmNARK*; whether this is functional in AON signalling was unclear. Therefore we used computational complementation to test two opposing hypotheses: (a) cotyledons function as part of the root, incapable of perceiving Q and producing SDI; or (b) cotyledons function as part of the shoot, involved in regulating root nodules.

5.2 Methods

Genetic complementation (Kahl, 1995) is a classical approach to define genetic cause-and-effect relations. For example, assuming two mutant organisms exhibit the same phenotype caused by loss-

of-function (recessive) mutations, then their hybrid will be wild-type, if the mutations are in different genes (called cistrons); conversely the hybrid will be mutant if the mutations are in the same cistron. In other words, the wild-type (functional) allele complements the deficiencies of the mutant. Genetic complementation is also used in transgenic analysis of organisms, as a loss-of function mutation in a candidate wild-type gene is deemed causal for a mutant phenotype if that mutant is effectively complemented by the transfer of a dominant wild-type allele. The complementation approach introduced here does not cross one genotype with another, but will use computational modelling to complement the deficiency (in an empirical model) of a mutant to determine if this recovers the virtual wild-type phenotype.

We use two well-characterized soybean (*Glycine max* L. Merrill) genotypes: the wild-type soybean Bragg and its loss-of-function mutant *nts1116* (Hansen et al., 1989). Wild-type soybean Bragg performs AON to keep its nodulation balance well-maintained (Figure 5.1 *A* and *C*), leading to characteristic crown nodulation in upper root portions. In its near-isogenic mutant *nts1116*, the Q signal generated from early nodule proliferation cannot induce SDI due to the lack of *GmNARK* activity in leaves (Figure 5.1 *B*). Reduced SDI in *GmNARK*-deficient plants leads to a phenotype with many more nodules than wild-type, called “supernodulation” or “hypernodulation” (Figure 5.1 *D*) (Carroll et al., 1985b). Compared with Bragg, the only deficiency of *nts1116* plants is the significantly reduced capacity of producing SDI.

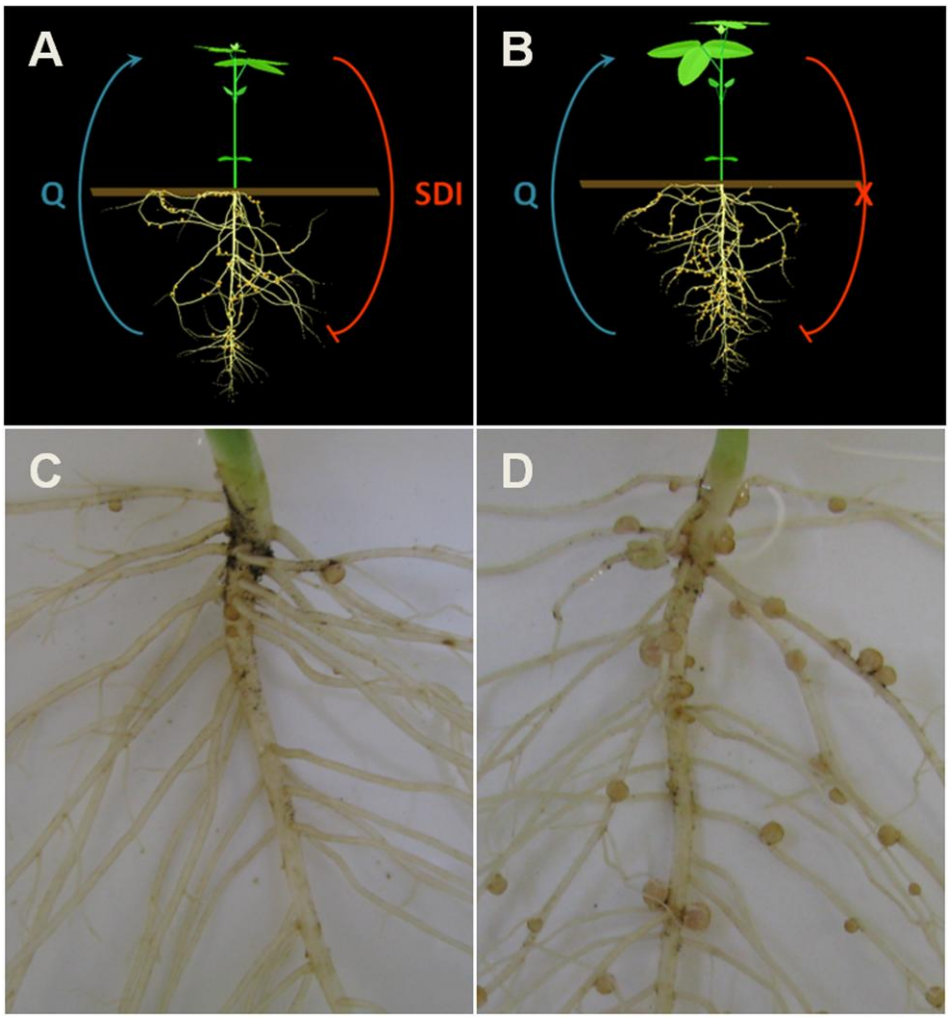


Figure 5.1. Wild-type soybean Bragg (left) and its supernodulation mutant *nts1116* (right). In the wild-type soybean Bragg, AON is well-established and the balance of nodulation is well maintained (A). This results in a phenotype with a normal number of nodules (C). In the mutant *nts1116*, GmNARK is not functional in the leaves, leading to the lack of SDI production (B) and consequently a supernodulation phenotype (D) with many more nodules than the wild-type.

The key idea of our complementation approach comes from this point. We "add" hypothetical components of AON signalling, including those of signal production, transport, perception and function (see also Text S4 in Appendix A.4), into the empirical model that depicts the growth behaviours of *nts1116* plants to see if a wild-type phenotype can be restored. The flowchart of methodology for this approach is given in Figure 5.2, including the following steps:

(i) Build empirical models to simulate architectural development of Bragg and *nts1116* plants based on biometric growth data collected from cultivation of the two genotypes under the same conditions. The empirical data include architectural information such as internode length and diameter, petiole length and diameter, leaf length and width, lateral root branching patterns,

and nodule number and distribution. Based on detailed organ-scale data, the architectural model can output realistic and dynamic visualisations as well as statistics of phenotypic development at a whole-plant scale. We call these outputs "system behaviours".

(ii) Extend the architectural model of a *nts1116* plant to a functional-structural model where simulation of inter-organ signalling activities is enabled and integrated with the signalling-development processes.

(iii) Parameterise the functional-structural model built in step (ii) based on the confirmed and the hypothesised mechanisms about AON signalling. After this parameterisation, we call the functional-structural model *"nts1116*+AON" [3]. The *nts1116*+AON model complements the deficiency of *nts1116*, and the resulting system behaviours represent a new nodulation phenotype.

(iv) Compare the new phenotype generated by the *nts1116*+AON model in step (iii) and the nodulation pattern produced by the Bragg architectural model in step (i). If they are same or similar, the hypotheses will be supported as reasonable. Otherwise, the hypotheses need to be modified and tested again from step (iii).

(v) If the hypotheses supported in step (iv) are testable by real-world experiments, the virtual-experiment process can suggest appropriate real-experiment methods to further evaluate them. The mechanisms further supported by real-plant experiments will then be used as "confirmed mechanisms" in step (iii) to serve the testing of remaining hypotheses.

(vi) If the hypotheses supported in step (iv) are not suitable or possible for evaluation through real-world experiments (due to limitation of current biological techniques), unknown attributes or characteristics about AON signalling can be predicted by virtual experiments.

[3] The potential architectural development, including nodule formation, in the *nts1116*+AON functional-structural model is also based on the empirical growth data of *nts1116*, as is the *nts1116* architectural model; the difference between them is that, in the *nts1116*+AON model, some nodules have the potential to occur (according to the empirical data) but do not occur because of the signalling regulation triggered by the successful formation of other nodules. From a model structure point of view, the code for the decision to produce a module in the *nts1116* model is extended in the *nts1116*+AON model to include a test of whether the inhibitory signal is present or not. This requires the addition of the synchronised signalling model as discussed in Section 4.3.

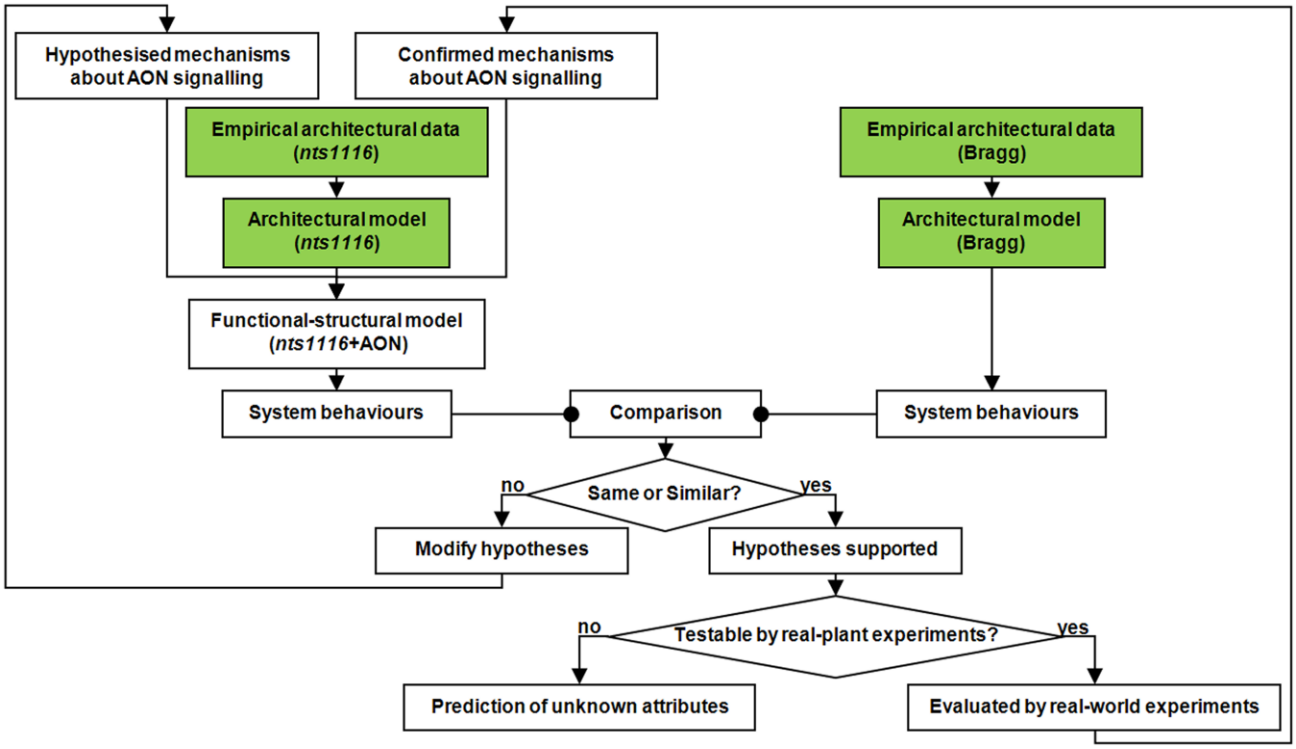


Figure 5.2. Flowchart of general computational modelling methodology. The first step (coloured in green) is to build empirical architectural models of parent cultivar Bragg and its derived mutant *nts1116*. The second step is to extend the *nts1116* architectural model to a functional-structural model enabling AON signalling. The confirmed and hypothesised mechanisms of AON signalling are then incorporated into the functional-structural model, to regulate the nodulation that cannot be regulated in real *nts1116* plants. The process is iterated until satisfactory comparison of system behaviours.

The architectural and functional-structural models mentioned in steps (i) and (ii) have been built with context-sensitive L-systems (Prusinkiewicz and Lindenmayer, 1990). The empirical data used for building architectural models of Bragg and *nts1116* plants were collected every second day from growth experiment under the same conditions until the 16th day post-sowing (all plants were inoculated on the 2nd day). Materials and methods for this glasshouse experiment are given in supporting Text S2 (Appendix A.2). The growth data, algorithms and techniques used for model construction are described in supporting Text S3 (Appendix A.3). The remaining steps of the flowchart, including (iii), (iv), (v) and (vi), are implemented for hypotheses testing and prediction.

5.3 Results

5.3.1 Application of the approach through virtual experiments

In this initial application of our computational complementation approach, two opposing hypotheses were tested: (a) cotyledons function as part of the root, incapable of perceiving Q and producing SDI (abbreviated as "cotyledon-root" hypothesis); (b) cotyledons function as part of the shoot, involved in regulating root nodules (abbreviated as "cotyledon-shoot" hypothesis). Since *GmNARK* is expressed in all organs (Nontachaiyapoom et al., 2007) (including cotyledons) and since cotyledons are short-term terminal organs (as they are degraded 7-14 days after germination), neither the cotyledon-root nor the cotyledon-shoot hypothesis was favoured a priori.

Theoretically speaking, if all other AON mechanisms (such as signal production, transport, perception and function) had been confirmed and used as basis for this application, the tested hypothesis leading to a wild-type nodulation pattern could be the correct one. However, the actions of many other signalling components also remain unclear. One or two virtual experiments are obviously insufficient to allow conclusions. Implementing too many experiments (to test all mechanisms together), however, would miss the emphasis and undermine efficiency. With these concerns, our strategy was to adjust parameters for signal production, transport, perception and function within a limited range, and use them as different conditions for different virtual experiments. Among all these experiments, if the complementation results (*nts1116*+AON) based on the cotyledon-root hypothesis are always or in most cases closer to Bragg than those based on the cotyledon-shoot hypothesis, then the cotyledon-root hypothesis would be considered plausible; otherwise, the cotyledons are more likely to function as general-sense leaves to regulate root nodulation.

According to this specific strategy, 27 virtual experiments (varying three rates of transport for both Q and SDI and three levels of nodulation inhibitory threshold) were designed for each of the two hypotheses: CRH_1~CRH_27 for cotyledon-root testing and CSH_1~CSH_27 for cotyledon-shoot testing. The only difference between CRH_i and CSH_j, if i=j, is whether cotyledons can function for AON signalling or not. Details of the virtual-experiment assumptions and conditions are described in the supporting Text S4 (Appendix A.4).

To quantify the comparison between complementation results and Bragg phenotype, we define their similarity degree S_{cp} as

$$S_{cp} = \frac{N_{nt} - N_{cp}}{N_{nt} - N_{br}} \tag{5.1}$$

where N_{nt}, N_{br} and N_{cp} are the nodule numbers generated respectively by the architectural model of *nts1116* plants, the architectural model of Bragg, and the functional-structural model of *nts1116*+AON. This can be understood as the ratio of the number of nodules inhibited by the virtual experiment to the number of nodules inhibited by a real Bragg plant. The similarity degrees of overall nodule number produced by virtual experiments on the 10th and the 16th day after sowing are listed in Figure 5.3 and Figure 5.4, where R_q and R_{sdi} represent the transport rates of Q and SDI signals (mm/day). These data indicated that the similarity degrees resulting from cotyledon-shoot hypothesis were generally much higher than those from cotyledon-root hypothesis, supporting the former hypothesis. Considering that values of S_{cp} greater than 100% may mean over-regulation and might not be optimal, the criterion for further evaluating S_{cp} is defined in Figure 5.5. According to this criterion, the virtual experiments based on cotyledon-root hypothesis produced unsatisfactory results on the 10th day (Figure 5.3, left-hand column), in sharp contrast to the cotyledon-shoot experiments (Figure 5.3, right-hand column). Although there were good results derived from virtual experiments CRH_1, CRH_2, CRH_11 and CRH_13 on the 16th day (Figure 5.4, left-hand column) in terms of nodule number, the nodule size and density from these experiments were all far from similar with the Bragg pattern (Figure 5.6). In comparison, the nodule distribution generated by CSH_1 (Figure 5.6 *D*) – the opposite of CRH_1 – was quite close to that of the Bragg architectural model.

We predicted from these complementation experiments that the cotyledons should be part of the shoot and participate as an SDI producer in wild-type soybean plants.

	cotyledon-root hypothesis									cotyledon-shoot hypothesis								
	CRH_1 ~ CRH_9			CRH_10 ~ CRH_18			CRH_19 ~ CRH_27			CSH_1 ~ CSH_9			CSH_10 ~ CSH_18			CSH_19 ~ CSH_27		
R_{sdi} / R_q	360	160	60	360	160	60	360	160	60	360	160	60	360	160	60	360	160	60
360	39% CRH_1	33% CRH_2	29% CRH_5	63% CRH_10	65% CRH_11	31% CRH_4	22% CRH_19	25% CRH_20	22% CRH_23	78% CSH_1	78% CSH_2	71% CSH_5	82% CSH_10	82% CSH_11	73% CSH_14	71% CSH_19	49% CSH_20	61% CSH_23
160	27% CRH_3	24% CRH_4	24% CRH_8	37% CRH_12	57% CRH_13	27% CRH_17	18% CRH_21	22% CRH_22	22% CRH_26	49% CSH_3	76% CSH_4	69% CSH_8	49% CSH_12	80% CSH_13	71% CSH_17	39% CSH_21	47% CSH_22	59% CSH_26
60	14% CRH_6	6% CRH_9	16% CRH_7	18% CRH_15	12% CRH_18	18% CRH_16	8% CRH_24	6% CRH_27	6% CRH_25	24% CSH_6	37% CSH_9	61% CSH_7	24% CSH_15	37% CSH_18	63% CSH_16	20% CSH_24	29% CSH_27	51% CSH_25

Figure 5.3. Complementation similarity degrees (10 days after sowing, 8 days after inoculation). The virtual-experiment results based on cotyledon-root hypothesis were all unsatisfactory on the 10th day, while there were good results produced by cotyledon-shoot experiments. The colours varying from red to blue represent lower to higher similarity degrees (*cf.* Figure 5.5).

	cotyledon-root hypothesis									cotyledon-shoot hypothesis								
	CRH_1 ~ CRH_9			CRH_10 ~ CRH_18			CRH_19 ~ CRH_27			CSH_1 ~ CSH_9			CSH_10 ~ CSH_18			CSH_19 ~ CSH_27		
R_{sdi} / R_q	360	160	60	360	160	60	360	160	60	360	160	60	360	160	60	360	160	60
360	95% CRH_1	84% CRH_2	52% CRH_5	126% CRH_10	117% CRH_11	62% CRH_4	73% CRH_19	66% CRH_20	30% CRH_23	116% CSH_1	112% CSH_2	98% CSH_5	126% CSH_10	128% CSH_11	111% CSH_14	96% CSH_19	92% CSH_20	68% CSH_23
160	50% CRH_3	64% CRH_4	38% CRH_8	72% CRH_12	112% CRH_13	53% CRH_17	39% CRH_21	60% CRH_22	28% CRH_26	74% CSH_3	110% CSH_4	100% CSH_8	84% CSH_12	134% CSH_13	110% CSH_17	52% CSH_21	89% CSH_22	59% CSH_26
60	21% CRH_6	22% CRH_9	15% CRH_7	27% CRH_15	40% CRH_18	33% CRH_16	9% CRH_24	9% CRH_27	10% CRH_25	29% CSH_6	55% CSH_9	74% CSH_7	30% CSH_15	60% CSH_18	100% CSH_16	20% CSH_24	28% CSH_27	38% CSH_25

Figure 5.4. Complementation similarity degrees (16 days after sowing, 14 days after inoculation). On the 16th day, four of the cotyledon-root experiments resulted in good similarity degrees, according to the criterion defined in Figure 5.5. In comparison, there were twelve cotyledon-shoot experiments with good results produced.

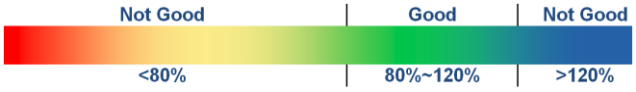


Figure 5.5. Criterion for evaluation of complementation similarity degree. If a similarity degree is between 80% and 120%, the complementation result it represents is viewed as "good"; otherwise the complementation result is viewed as "not good".

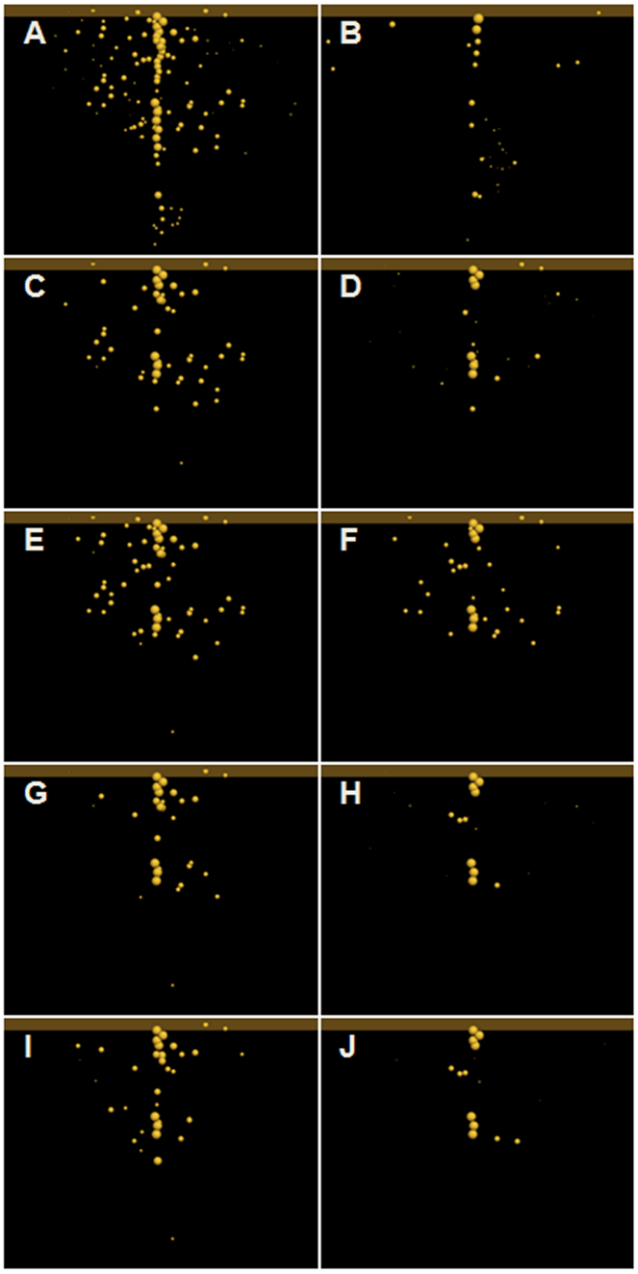


Figure 5.6. Visualisation of nodule distribution on the 16th day post-sowing. The primary and lateral roots were filtered in these visualisations to permit better observations of differences between nodule distribution patterns. As a guide, the pattern of yellow spots signifies essential AON characteristics in panel B, namely crown nodulation, restricted nodule number and small nodulation interval. (A) Nodule distribution generated by the *nts1116* architectural model. (B) Nodule distribution generated by the Bragg architectural model. The distribution patterns (C) (E) (G) and (I) resulted respectively from virtual experiments CRH_1, CRH_2, CRH_11 and CRH_13. The (D) (F) (H) and (J) were from CSH_1, CSH_2, CSH_11 and CSH_13. Potential nodules, which were not formed because of inhibition, can also be made visible, as shown by supporting Figure S5 (Appendix A.5).

5.3.2 Confirmation of the virtual-experiment result

To confirm the above prediction and also to evaluate the effectiveness of this approach, a "real-plant" grafting experiment was conducted. The critical experiment was to graft – between Bragg and *nts1116* plants [4] – the shoot of one genotype with cotyledons to the root of the other genotype without cotyledons, and also to graft the shoot of one genotype without cotyledons to the root of the other genotype with cotyledons, forming four graft combinations: Ns+Nc+Br, Ns+Bc+Br, Bs+Bc+Nr and Bs+Nc+Nr (Table 5.1). Materials and methods for this graft experiment are given in the supporting Text S2 (Appendix A.2). The collected empirical data for nodule number were not only classified by each graft type but also according to each plant's cotyledon retention status (Table 5.2).

Table 5.1. Real-plant graft types.

Ns+Nc+Br	*nts1116* shoot with cotyledons + Bragg root without cotyledons
Ns+Bc+Br	*nts1116* shoot without cotyledons + Bragg root with cotyledons
Bs+Bc+Nr	Bragg shoot with cotyledons + *nts1116* root without cotyledons
Bs+Nc+Nr	Bragg shoot without cotyledons + *nts1116* root with cotyledons

Table 5.2. Cotyledon retention status.

0_C	both cotyledons have fallen
1_YC	the plant only has one yellow cotyledon
2_YC	both cotyledons of the plant have turned yellow
2_GC	both cotyledons of the plant are green

According to the experimental results, the nodule number from the Ns+Nc+Br graft type was much higher than that from the Ns+Bc+Br (Figure 5.7 *A*). For the Ns+Bc+Br graft type alone, its plants with fallen cotyledons had more nodules than those with persisting cotyledons, and the plants with yellow cotyledons had more nodules than those with green cotyledons (Figure 5.7 *C*). These differences suggest Bragg cotyledons were the only leaves to regulate nodulation in Ns+Bc+Br plants, because unifoliate and trifoliate leaves of *nts1116* plants were unable to do so.

[4] Previous work (Delves et al., 1986; Men et al., 2002) has indicated that the homografts and ungrafted plants had no significant difference in nodulation and that nodulation parameters of heterografts in the *nark* system had no significant difference from the relevant homografts. Therefore homografts were not used in this experiment to measure the impact of grafting itself.

Data of another graft type with Bragg cotyledons – the Bs+Bc+Nr (Figure 5.7 *D*) also suggested that the Bragg cotyledons participated in providing SDI. However, more nodules were found in the Bs+Bc+Nr plants than in the Bs+Nc+Nr plants that had no Bragg cotyledons (Figure 5.7 *B*). An explanation for this observation is that the Bs+Nc+Nr allowed more nodules to be formed at early stages than the Bs+Bc+Nr, leading to more Q signal moving from root to shoot. As the cotyledon biomass declined greatly at later stages of seedling growth (resources are unloaded for plant growth and the "spent" cotyledon is eventually discarded), the difference in shoot between Bs+Bc+Nr and Bs+Nc+Nr became insignificant. Therefore larger amounts of Q triggered more SDI, which finally inhibited more nodules in Bs+Nc+Nr.

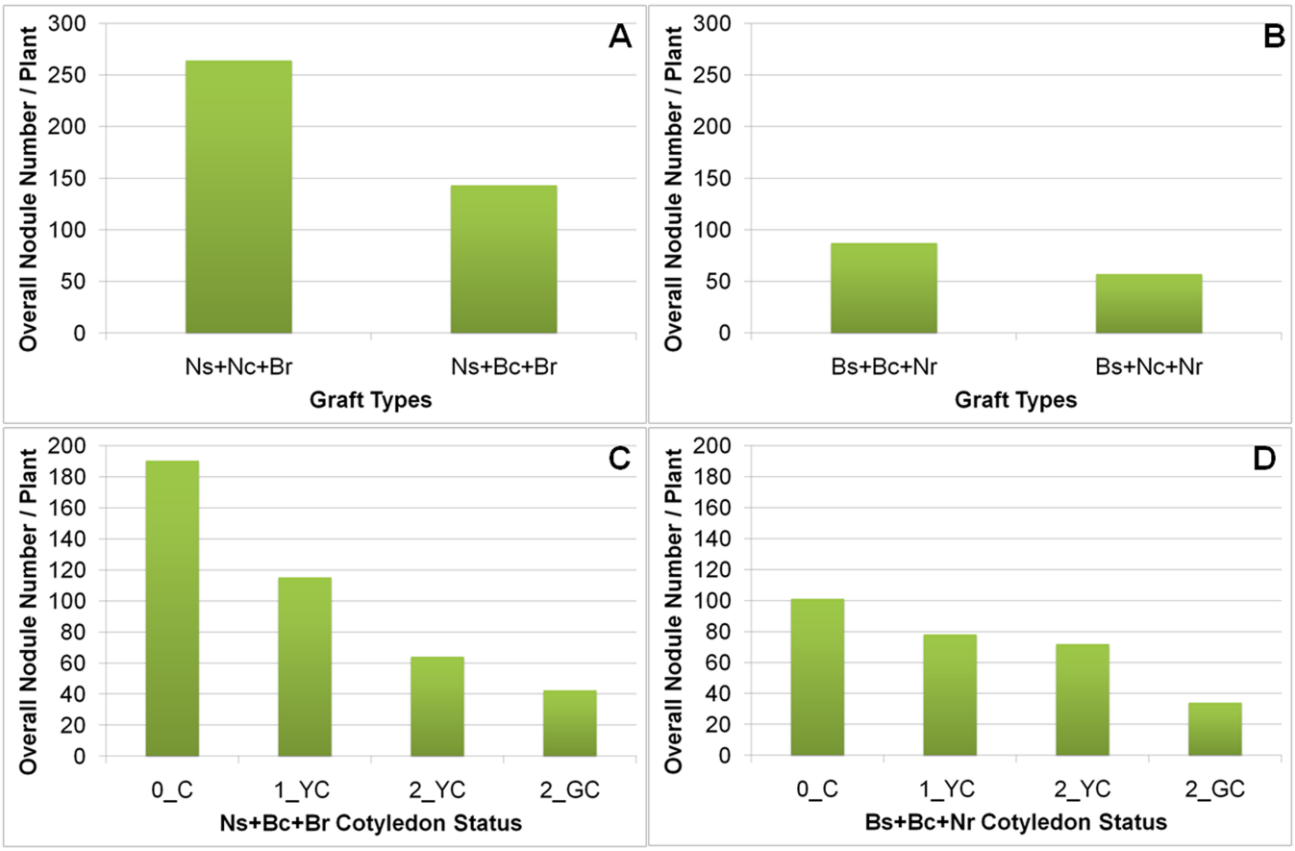


Figure 5.7. Nodulation of real-plant mutant-parent grafts. (A) Nodule numbers from Ns+Nc+Br (*nts1116* shoot with cotyledons + Bragg root without cotyledons) and Ns+Bc+Br (*nts1116* shoot without cotyledons + Bragg root with cotyledons) graft types. (B) Nodule numbers from Bs+Bc+Nr (Bragg shoot with cotyledons + *nts1116* root without cotyledons) and Bs+Nc+Nr (Bragg shoot without cotyledons + *nts1116* root with cotyledons) graft types. (C) Nodule numbers from Ns+Bc+Br plants classified by cotyledon retention status. (D) Nodule numbers from Bs+Bc+Nr plants classified by cotyledon retention status.

To better understand this nonlinear characteristic brought out by real-plant experiments, we returned to the virtual-experiment models and visualised the dynamic signal allocation during CRH_1 and CSH_1 (Figure 5.8). As demonstrated by the visualisation, the SDI concentration (in the root) of CRH_1 was lower than that of CSH_1 on the 5th day but became higher from the 10th day on, in agreement with the above analysis of the nodulation difference between Bs+Bc+Nr and Bs+Nc+Nr. Thus, we conclude that the testing result from our initial application of computational complementation is confirmed: the cotyledons "belong" to the shoot and function as a source of the nodulation regulator in wild-type soybeans.

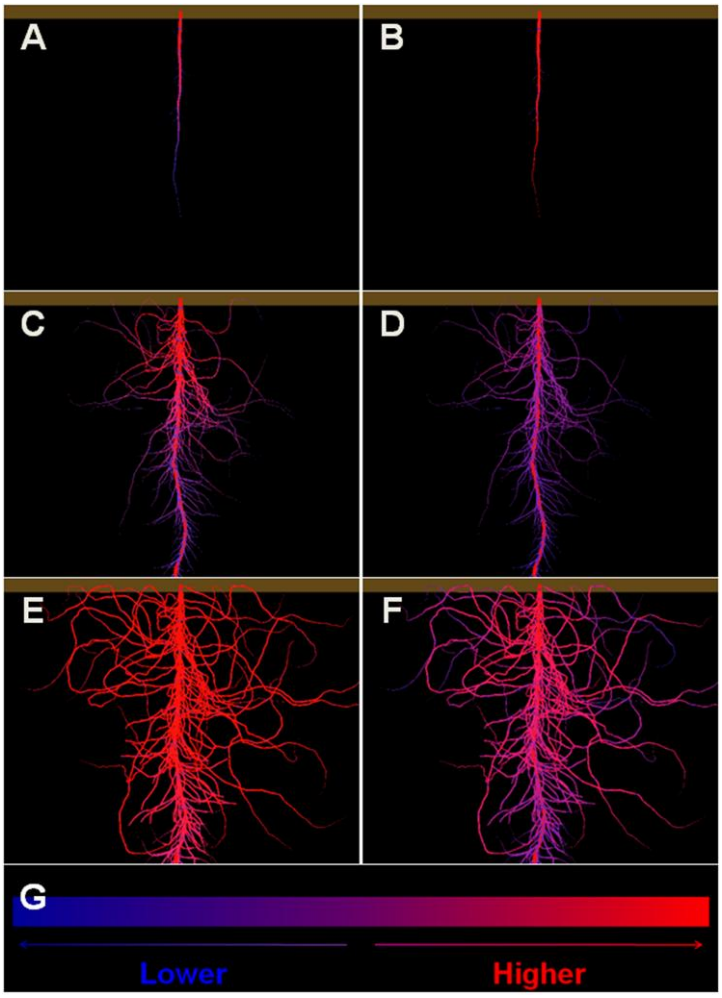


Figure 5.8. Allocation of the putative SDI signal during a virtual experiment. The allocation of SDI in root during CRH_1 was visualised on (A) the 5th, (C) the 10th, and (E) the 16th day post-sowing. Visualisations (B) (D) and (F) were from the functional-structural model used in CSH_1, respectively representing the 5th, 10th and 16th day. The colour scheme to represent signal concentration is given in (G).

5.4 Discussion

The computational complementation approach introduced here is an original contribution to the study of legume autoregulation of nodulation. Compared with conventional biological technologies with broader implications to plant development, one of the major advantages of this approach is its capability to complement the deficiency of a mutant plant at an organ scale with totally hypothetical and concept-derived physiological components. It is also able to make hypothetical signalling details manipulable and visible. For example, as demonstrated in the above case, signal transport rates can be modified as hypothesised and the allocation of signal can be dynamically visualised. These functionalities not only enable AON researchers to test hypotheses or make predictions using time- and resource-saving virtual experiments, but also bring out possible underlying details that are unobservable through real-plant experiments. Moreover, the application of this approach is not only limited to AON research, but also potential to other plant signalling studies such as those on branching regulation (*e.g.*, Dun et al., 2009), flowering control (*e.g.*, Wenden et al., 2009) and lateral root initiation (*e.g.*, Aloni et al., 2006).

This approach contributes a new idea to the domain of computational plant modelling – computational complementation. From a classic modelling point of view, one can formulate a model based on empirical data and then verify the model against the data, which has been used for development of crop (*e.g.*, Jones et al., 2001) and architectural (*e.g.*, Watanabe et al., 2005) models. However, what we investigate is a largely unclear internal signalling system – most of the detailed mechanisms remain unknown, which determines there is no direct parameterisation-and-verification data to evaluate the modelled signalling hypotheses. Using an indirect strategy, functional-structural modelling allows us to use the observable structure as a reporter for estimation of the unobservable function. But for this study, we have to link the structure of one genotype with the function of another genotype. The reason for this is: the wild-type Bragg nodulation has already been regulated, thus incorporating AON to Bragg architecture would double the regulation and have no reasonable comparison target for validation; in contrast, the *nts1116* is a non-AON plant and this is its only difference with Bragg, therefore activating AON in *nts1116* plant could result in system behaviours comparable with the wild type.

Another feature of this approach resides in the level of complexity for simulation of structural and signalling processes. We captured root details for studying shoot-root signalling rather than oversimplifying the root system. And the signalling pathways are constructed with sub-modules of which the size and number can be manipulated without limitation, which allows future modelling

work to be extended to lower-scale mechanisms (such as tissue and cellular scale). We also created a synchronisation algorithm for coordination of multi-rate procedures to enhance the precision of signalling-development interactions. A description of these modelling techniques is given in the supporting Text S3 (Appendix A.3).

The approach also has some limitations. For example, due to the nature of complementation, it can only be used for a single mutation at a time, though leaky mutants can be handled by parameter optimization. Another drawback is that it cannot distinguish between different mutations in the same pathway that result in the same phenotype in the first instance. In other words, if the hypothesised mechanisms used to complement the mutant are the same in both cases, and so is the phenotype of the two mutants, computational complementation cannot be used to say which gene component of the regulatory network has been mutated.

Our first application of this approach was to test whether wild-type soybean cotyledons are involved in production of SDI. Also but more importantly, we expected this application to evaluate whether the computational complementation idea is effective. The virtual-experiment results suggested the wild-type cotyledons can produce SDI, which was further confirmed by a graft experiment on real plants. This demonstrates the feasibility of computational complementation and shows its usefulness for future applications.

The next step is to apply this approach to support research for the identification of Q and SDI[5]. Candidate signals, such as CLE peptide for Q (Gresshoff et al., 2009; Okamoto et al., 2009) and auxin for SDI (Mathesius, 2008), will be tested to see if they play the roles in AON as hypothesised. In addition, environmental factors, such as soil nitrogen status, that have effects on the process could also be tested with this approach. Furthermore, the finding that wild-type soybean cotyledons act as an SDI producer in AON opens the door for testing physiological transgenerational effects, such as altered nodulation patterns influenced by the *Bradyrhizobium* infection status of mother plant through presence of SDI in cotyledons.

[5] Although the modelling approach can help in the identification of Q and SDI, its contribution is limited to virtual-experiment predictions or testings of ideas. The final identification of Q and SDI will be achieved through experiments based on real plants (*cf.* Lin et al., 2010; Reid et al., 2011).

5.5 Acknowledgements

The authors would like to thank Kuang-Hsu Wu from the University of Queensland for his participation in the collection of soybean growth data to support architectural models; and Dongxue Li from the ARC Centre of Excellence for Integrative Legume Research for carrying out the graft experiment.

6 General discussion

This thesis has contributed a group of computational modelling techniques as well as a computational complementation strategy that can be employed to investigate the complex signalling mechanisms occurring during legume autoregulation of nodulation (AON). Based on these approaches, a set of virtual experiments was implemented to test the hypotheses of whether wild-type soybean cotyledons can produce SDI signal to inhibit root nodulation. The affirmative result was then confirmed by a real-plant grafting experiment, demonstrating the feasibility, usefulness and advantages of these original contributions. Although there are still aspects in need of improvement, these modelling techniques and computational complementation strategy help pave the way for future investigations of other unknown mechanisms of AON and for applications to other root- or signalling- related studies.

6.1 Major contributions

As reviewed in Section 2.1 and 2.2, the autoregulation of nodulation has been studied since the 1980s. One major objective of these studies is to have a clearer understanding of how this signalling system works and what signals are involved, so that nodulation – the biggest natural source of fixed nitrogen – can be better used for agricultural, industrial, medical and environmental purposes. On one hand, many details about AON, from genetic sequences to physiological processes, have been revealed or hypothesised with traditional biological technologies; on the other hand, the subtlety and intricacy of its signalling mechanisms remain a mystery.

The mystery, however, provides a playground for computational models. As discussed in Section 2.3, behind the complexity of AON is a complex system: the leaves and the nodules are connected by a long-distance shoot-root network. This system keeps changing, not only as a result of plant growth but also due to regulation by the SDI signal. At the same time, the strengths of Q and SDI flows are modified in turn by the topological and dimensional changes to this network. To model such a complex system, with its demonstrated characteristics including dynamics, non-linearity, self-organisation and emergence, a diagram or a set of mathematical equations are far from enough. Computational modelling, taking advantage of its capabilities to handle large amounts of

information and dealing with complex processes (as introduced in Section 1.2), is an ideal and even indispensable method in this situation. The ideology of this thesis is to use computational modelling to simulate the local signalling interactions of AON and then evaluate the results – the emerging system behaviours – against real-plant performance.

6.1.1 At the technical level

The emergent system behaviours of AON are well represented by the number and distribution of nodules. Thus collecting nodulation information from a real plant and then representing it in a computational model are required to support the simulation-and-nature comparisons. Plant architectural modelling that makes possible 3D visualisations for description of plant topology and geometry is an ideal solution to meet this requirement. However, at the technical level, there were three major challenges for representation of a legume root system: how to characterise the irregularity of lateral roots, how to characterise nodule distribution, and how to restrict root development with empirical data in an architectural model. To meet the first challenge, a "RULD" root mapping method was established in this research. When a soybean root was put into water and observed from an overhead view (Figure 4.1), a regular branching pattern of the first-order lateral roots was found: the radial angles of lateral emission are usually around 90^{o}. Based on this pattern, the first-order laterals were categorised into quadrants (right-R, up-U, left-L and down-D) according to the relative positions of their emission points to the obtuse angle composed by the two cotyledons (Figure 4.2). Then the primary root was divided into regions, segments and sites (Figure 4.3), allowing each first-order lateral to be identified. To capture the nodule distribution information, a "nodulation section" was defined for each primary-root region and for each first-order lateral root (Figure 4.4). The positions of the first and the last nodules of a nodulation section were used to determine its location in the root system. The length of such a section was divided by the number of nodules it includes to calculate the interval between two successive nodules. The "RULD" mapping method and the definition of "nodulation section" not only set the framework for collection of root architectural details, but also provided the basic data structure to drive the root elongation algorithm developed in this research. The general tactic of this elongation algorithm is to move the primary root tip downwards while creating the root structure and setting the potential lateral and nodule formation sites behind (Figure 4.5). Whether and when the laterals and nodules will be formed at those potential sites are further determined by the conditions derived from empirical growth data. L-studio, an L-system-based tool for plant modelling, was chosen to implement the algorithm for simulation of root development. In previous L-system models built for shoot architectures, an inter-

node section was usually treated as a single module. In this thesis, a root area between branching points, nodulation sites or root tip was represented by a set of sub-modules with the same length "UNIT" rather than a single module. This is more appropriate to visualise the twisted roots and to set the potential nodule or lateral formation sites wherever needed.

The reconstruction of root development only describes the phenomena or empirical data that are observable from real plants. To investigate how the architecture is shaped by internal signals, the interactions between architectural development and underlying signalling must be addressed. Functional-structural modelling, which takes inter-organ communication into account and use plant architecture as a direct reporter of the physiological mechanisms, sets the stage for linking these processes. At the technical level, the simulation of signal movement from one sub-module to its neighbour is well supported by context-sensitive L-systems (Figure 4.8). The challenge, nevertheless, is how to coordinate signalling and developmental processes that have various rates. In L-system models, the production rules for architectural development as well as for information transfer between neighbouring modules are checked and implemented step by step. In this study for example, during one L-system simulation step, only one sub-module is added for root elongation or passed through for signal transport. If these two processes were both performed at every single step, their rates in simulation would be the same. To deal with this issue, two states – "activated and stopped" – were assigned to each signalling and developmental event, thereby the events with lower rates can be stopped for a number of steps when those with higher rates are activated. This "activated and stopped" scheme is similar to "process scheduling" used in computer operating systems and is not something new. The difficulty is in switching between the two states for each event properly, according to its empirical or hypothetical rate (e.g. a value scaled by "mm/day"), so that all the events can be dispatched in a synchronised way. To meet this challenge, a synchronisation algorithm was developed, where the key role is played by the UNIT – the length of the standard sub-module. Given a time division at any time scale, the number of steps allocated to an event for its activated state is equivalent to the number of UNITs covered by its elongation or signal transport during such a time division in nature. Once the event with highest rate uses up its simulation steps, a time division is finished and a new division is initialised. In essence, this method treats time based on space, sharing the same philosophy as in defining calendars (a year is in accordance with an orbit of the Earth surrounding the sun; a month is counted after the moon finishes an orbit around the Earth; etc.). The synchronisation algorithm supports as many signalling and developmental events as required and allows its users to define whatever time scales are needed. Furthermore, this algorithm

is useful not only for studying AON and signalling, but also for synchronisation of other physiological processes such as water and carbon flows.

In addition to the flexibility provided for root reconstruction and signalling-development synchronisation as discussed above, the use of sub-modules has another important advantage: by colouring each sub-module according to a signal's local amount, it allows the variation of signal allocation throughout the whole plant to be visualised in detail. This can help plant signalling researchers bring out the possible internal mechanisms that are unobservable from real plants and use them to analyse their regulatory results with complex behaviours (Figure 4.10).

6.1.2 At the strategic level

The techniques developed for modelling and simulation bring details of the observed development and the hypothesised signalling into a computational space, where the emergent system behaviours resulting from local interactions can be reproduced. The next step was, then, to use these techniques for a better understanding of unknown AON mechanisms. As discussed earlier in Section 6.1, comparing the simulated system behaviours (represented by nodulation pattern) with real-plant performance is this thesis's general intention. In practice, there was an issue that must be addressed: whether the simulation result is really comparable with the real plant. Unlike an architectural model that can be evaluated against empirical data directly, the system behaviours of a functional-structural AON model are not only restricted by empirical data but also controlled by the modelled signalling mechanisms. If such a functional-structural model reconstructed a wild-type soybean structure based on its empirical data and then incorporated the hypothesised signalling mechanisms, it would double the AON regulation, as the empirical nodulation data had already been a regulation result from the wild. To avoid this, a computational complementation strategy was developed in this research. The main idea of this strategy was to complement the deficiency of an empirical model of a loss-of-function (non-AON in this case) mutant with hypothetical signalling mechanisms. This complementation updates the mutant empirical model into a functional-structural model. The potential architectural development in the functional-structural model is still based on empirical data, which are the same as in the mutant empirical model; the difference is that the architectural development in the functional-structural model triggers and enables the signalling system that is lost in the mutant plant and consequently the potential growth of some organs is affected by the feedback signals. A wild-type plant also has such interactions between potential growth and signal regulation in nature. Since the wild-type architectural development is representable by its empirical model

(which is built the same way as the mutant empirical model though the empirical data are different), comparing the result of computational complementation and the performance of the wild-type empirical model can be used to evaluate whether the hypothesised signalling mechanisms are close to their counterparts in nature. If computational complementation demonstrates a phenotype similar to the wild-type plant, the signalling hypotheses would be suggested as "reasonable", since the loss of function is considered to have been complemented by the hypothesised mechanisms. If the reasonable hypotheses are further testable by real-plant experiments, the complementation process can suggest appropriate experimental methods to further evaluate them; if the hypotheses cannot be evaluated through real-plant experiments, due to the limitation of current biological techniques, some unknown attributes or mechanisms of the studied signalling can be predicted by computational complementation.

Two well-characterised soybean genotypes: the wild-type soybean Bragg and its loss-of-function mutant *nts1116* were used as the basic plant materials for this computational complementation study. The only difference between these two genotypes is that Bragg is regulated by AON, leading to a well-balanced nodulation pattern, while *nts1116* is deficient in the *GmNARK* activity in leaves, thus cannot produce SDI and has a "supernodulation" or "hypernodulation" phenotype with many more nodules than Bragg. The detailed empirical growth data of both Bragg and *nts1116* were collected to reconstruct their shoot and root architecture. Then the hypothesised AON mechanisms can be incorporated to the *nts1116* architectural model, forming an "*nts1116*+AON" functional-structural model. In the *nts1116*+AON model, the shoot-root signalling activity starts when the first nodule is initialised. Therefore some of the other nodules that should have appeared might not appear as a result of the complementation. This makes the simulation of the *nts1116*+AON model comparable with that of the Bragg architectural model, allowing determination of whether the wild-type phenotype is restored.

The first application of this strategy was to investigate whether the wild-type soybean cotyledons are involved in the production of SDI, which remained unknown before this study. Also but more importantly, this application was used to evaluate whether the computational complementation is an effective approach. Two alternative hypotheses were formulated for this testing: (a) cotyledons function as part of the root, incapable of perceiving Q and producing SDI (abbreviated as "cotyledon-root" hypothesis); (b) cotyledons function as part of the shoot, involved in regulating root nodules (abbreviated as "cotyledon-shoot" hypothesis). These two hypothesised mechanisms were used with other AON processes to complement an *nts1116* empirical model. Theoretically

speaking, if all other mechanisms had been confirmed, the hypothesis that can produce a wild-type nodulation pattern could be considered as the correct one. However, the situation was that, many other signalling mechanisms of AON also remained unclear. As discussed in Section 5.3, one or two virtual experiments are insufficient to allow conclusions, while implementing too many experiments (to test all mechanisms together) would miss the emphasis and undermine efficiency. An achievable and proper solution in this situation was to vary parameters of the signalling mechanisms within an appropriate range, use them as different virtual-experiment conditions and see which of the two cotyledon hypotheses can always or in most cases lead to a nodulation pattern closer to Bragg. The virtual-experiment results indicated the cotyledon-shoot hypothesis was more likely to be correct, suggesting that the wild-type soybean cotyledons are involved in AON signalling and can produce SDI. This prediction was subsequently confirmed by a grafting experiment on real plants, demonstrating the feasibility of computational complementation.

Compared with conventional biological technologies, the computational complementation approach developed in this thesis has the advantage that it has the capability to complement the deficiency of a mutant plant with totally hypothetical and concept-derived physiological mechanisms. As an original contribution to AON study, it provides a time- and resource-saving strategy for using virtual experiments to test ideas about AON mechanisms. The foundation for its future applications has been laid by this thesis with development of the modelling and simulation techniques, collection of the necessary empirical data and the confirmed discovery through its first application.

6.2 Limitations

The techniques and approaches developed in this thesis also have some limitations. At the technical level, the data collection and root reconstruction are based on young soybean roots, where the second-order laterals and mortality of root tips are not considered. If these aspects are required in future applications using older plants, additional techniques for data collection will need to be developed. The "RULD" branching pattern of the first-order lateral roots were found through observation of soybean roots. If a different legume species, such as pea, is used as the target plant to study AON, whether its first-order laterals there can be characterised into quadrants (or whether the radial angles of lateral emission are around 90^{o}) would need to be further investigated. Compared with the automatic or semi-automatic techniques developed in other studies to collect root data (as reviewed in Section 4.1), the methods developed in this thesis, based on "RULD" mapping and the definition of "nodulation section", are more appropriate to capture architectural details of soybean

roots to address their 3D patterns. However, using these methods, one needs to measure lateral roots and count nodules one by one based on manual work, which is labour-intensive and exhausting.

At the strategic level of application of computational complementation, the prerequisite is that a wild-type plant and its loss-of-function mutant are both available, which is a major limitation to the generality of this strategy. In studies for some other plant signalling systems, if a wild-type plant does not have such a loss-of-function mutant, it will not be suitable for using computational complementation to investigate them. Or, if the phenotypic differences between a wild type and its mutant are not significant or clear, there will be little ground to evaluate the complementation result. Furthermore, due to the nature of complementation, it can only be used for a single mutated process at a time. And it cannot distinguish between different mutations in the same pathway that result in the same phenotype in the first instance. For example, if the physiological mechanisms used to complement the mutant are the same at the organ scale, the complementation cannot distinguish which gene component of the regulatory network has been mutated.

6.3 Future directions

6.3.1 Towards better efficiency

As presented in Section 4.3, the simulation of signal movement between neighbouring sub-modules was supported by context-sensitive L-systems. At any L-system simulation step, as long as a signalling event is activated, its concentration level in every sub-module needs to be checked and updated. On one hand, from the visualisation point of view, this helps in revealing spatial and temporal details of signal allocation; on the other hand, however, it is not very efficient in information transfer, as the signal has to be passed from sub-module to sub-module and step by step in a plant structure that has thousands of such sub-modules.

In the newly developed L-system-based programming language "L+C" (Prusinkiewicz et al., 2007), a "fast information transfer" functionality has been included. Given an L-system string with n sub-modules, using the traditional context-sensitive information transfer takes n-1 simulation steps to move a signal's information from the beginning of this string to its end; while it only costs one step to accomplish this by using the fast information transfer. For example, if the string is given as

$$R_1(d_1)\ R_2(d_2)\ R_3(d_3)\ R_4(d_4)$$

where R_n is a sub-module and d_n represents its signal information, and a context-sensitive rule changes each sub-module from $R_n(d_n)$ to $R_n(d_{n-1})$, then it will take three steps for using traditional information transfer to pass a signal from the start to the end of the string (Figure 6.1 *A*). Using fast signal transfer, the next state of parameter values in the context are used, rather than the current state, thus allowing the signal to be passed in one step no matter how many sub-modules this string has (Figure 6.1 *B*).

Nevertheless, there will be some new challenges in using fast information transfer as an alternative simulation technique, which may include:

- Only one direction is permitted for the fast information transfer during each single simulation step. The direction needs to be specified as "Forward" or "Backward" in advance of each simulation step. Using the above string of sub-modules for example, the information can only be transferred from R_1 to R_4 if the transfer direction is set as "Forward" beforehand. To move another signal's information from R_4 to R_1, the transfer direction needs to be changed to "Backward". An appropriate method must be addressed to synchronise signals moving in opposite directions.
- An extra parameter is needed for representation of signal concentration. The transfer process illustrated in Figure 6.1 is a very simple case that only copies d_{n-1} from R_{n-1} to replace d_n in R_n. If the signal transport is as complex as that indicated in Figure 4.8 and if d_1 is over the concentration threshold at the first step, the signal concentration level in R_1 needs to be reduced to 0 after a simulation step. This means the next state of R_1 that is sensed by the current R_2 is $R_1(0)$ rather than $R_1(d_1)$, thereby the next state of R_2 is $R_2(0)$ and the string will become "$R_1(0)$ $R_2(0)$ $R_3(0)$ $R_4(0)$" at the second step. To avoid this, a possible solution is using two parameters to represent a signal's concentration level: one for the pre-existing amount plus the newly moved-in amount that can be sensed by its neighbour, and the other for the left amount after the transfer.
- A new scheme to synchronise the signalling and developmental processes with various rates needs to be developed. It seems that one simulation step can be used as one time division, as the fast information transfer allows the signal to move through multiple sub-modules during a single step. However, what if there are new sub-modules added to the structure when the information is being transferred? Since the plant structure keeps growing and branching, the question of how signals should be passed into the newly produced sub-modules needs to be addressed.

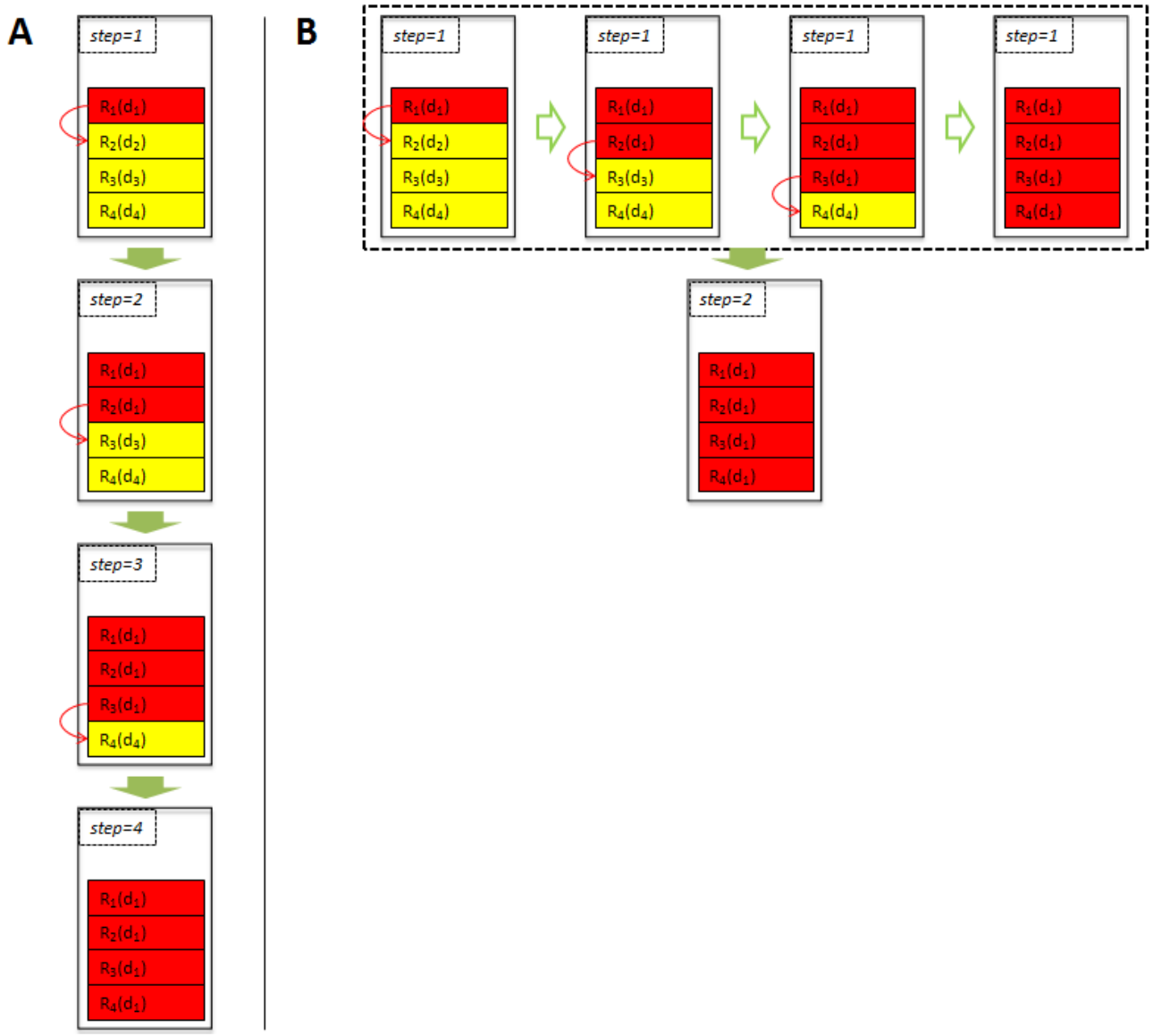


Figure 6.1. Comparison of traditional context-sensitive L-systems and fast information transfer. Using traditional context-sensitive L-systems, a sub-module R_n checks the current state of parameter values in its neighbouring R_{n-1}, therefore the information can only be passed through one sub-module at most after a simulation step. Using fast information transfer, the information can be passed through the whole string after a step regardless how many sub-modules the string has.

6.3.2 Towards identification of Q and SDI

Whether wild-type soybean cotyledons can produce SDI has been answered in this thesis. This can be used as a confirmed mechanism to support the testing of unconfirmed mechanisms in future. As reviewed in Chapter 2, more and more biochemical attributes of Q and SDI are being discovered and their candidate signals, such as CLE peptide for Q and auxin for SDI, have been hypothesised. Using these newly discovered and hypothesised signalling attributes to complement the supernodulation

mutant through the established virtual-experiment platform should be the next step of applying computational complementation. Even though this application might not be able to identify the AON signals, it could still provide some useful clues, such as what chemical substance is suggested unlikely to be Q or SDI based on the signalling rates tested by the model, to help in narrowing the possibilities to be tested.

6.3.3 Towards integration of lower-scale models

Although the modelling focus of this thesis has been on the organ-scale mechanisms of AON, the use of sub-modules provides the possibility for future integration of lower-scale signalling models. For example, each sub-module can be representative of a small cross-section through the root for studying local signalling activities such as diffusion of SDI within it. Such a cross-section could be represented by a sub-module in the functional-structural model while its local signalling processes can be simulated with a separate sub-model. This allows communications to be established between these two models at different scales: the signal flux moving into a sub-module can be used as an input to be processed by the relevant lower-scale model; and the output of the lower-scale model can be sent back to the sub-module and transferred to its neighbour. If this integration is successfully achieved in future, it can be used to examine more AON mechanisms with details at different scales and make the computational complementation more precise and reliable in hypothesis testing and prediction.

6.3.4 Beyond autoregulation of nodulation

The modelling and simulation techniques as well as the computational complementation approach developed in this thesis are not limited to autoregulation of nodulation. They have a potential to be applied for a wider range of studies on other root or signalling related systems. The method for characterisation of lateral roots might be useful for research on lateral root spatial patterning (Dubrovsky et al., 2006). The signalling-development synchronisation scheme and the computational complementation strategy may be amended to support other plant signalling studies such as those on branching regulation (Dun et al., 2009), flowering control (Wenden et al., 2009) and lateral root initiation (Aloni et al., 2006). In the branching regulation study for example, if the production of the regulatory signal is inhibited, which leads to a mutant with significantly increased branch number, then the computational complementation approach can be used to "complement" the deficiency of the mutant plant with hypothesised and confirmed signalling mechanisms to see if the wild-type branch number can be restored.

References

Abadia-Fenoll F, Lloret P, Vidal M, Caserc P (1982) Pattern distribution of lateral root primordia in *Allium cepa* L. Phyton, International Journal of Experimental Botany **42:** 175-177

Allen M, Prusinkiewicz P, DeJong T (2005) Using L-systems for modeling source–sink interactions, architecture and physiology of growing trees: the L-PEACH model. New Phytologist **166:** 869-880

Aloni R, Aloni E, Langhans M, Ullrich CI (2006) Role of cytokinin and auxin in shaping root architecture: regulating vascular differentiation, lateral root initiation, root apical dominance and root gravitropism. Annals of Botany **97:** 883-893

Barthélémy D, Caraglio Y (2007) Plant architecture: a dynamic, multilevel and comprehensive approach to plant form, structure and ontogeny. Annals of Botany **99:** 375-407

Bell JK, McCully ME (1970) A histological study of lateral root initiation and development in *Zea mays*. Protoplasma **70:** 179-205

Berleth T, Scarpella E, Prusinkiewicz P (2007) Towards the systems biology of auxin-transport-mediated patterning. Trends in Plant Science **12:** 151-159

Beveridge CA, Mathesius U, Rose RJ, Gresshoff PM (2007) Common regulatory themes in meristem development and whole-plant homeostasis. Current Opinion in Plant Biology **10:** 44-51

Bidel LPR, Pagès L, Rivière LM, Pelloux G, Lorendeau JY (2000) MassFlowDyn I: a carbon transport and partitioning model for root system architecture. Annals of Botany **85:** 869-886

Biswas B, Chan PK, Gresshoff PM (2009) A novel ABA insensitive mutant of *Lotus japonicus* with a wilty phenotype displays unaltered nodulation regulation. Molecular Plant **2:** 487-499

Broughton WJ, Dilworth MJ (1971) Control of leghaemoglobin synthesis in snake beans. Biochemical Journal **125:** 1075-1080

Buck-Sorlin G, Hemmerling R, Kniemeyer O, Burema B, Kurth W (2008) A rule-based model of barley morphogenesis, with special respect to shading and gibberellic acid signal transduction. Annals of Botany **101:** 1109-1123

Buck-Sorlin GH, Kniemeyer O, Kurth W (2005) Barley morphology, genetics and hormonal regulation of internode elongation modelled by a relational growth grammar. New Phytologist **166:** 859-867

Buzas DM, Gresshoff PM (2007) Short- and long-distance control of root development by *LjHAR1* during the juvenile stage of *Lotus japonicus*. Journal of Plant Physiology **164:** 452-459

Caetano-Anollés G, Gresshoff PM (1990) Early induction of feedback regulatory responses governing nodulation in soybean. Plant Science **71:** 69-81

Caetano-Anollés G, Gresshoff PM (1991a) Efficiency of nodule initiation and autoregulatory responses in a supernodulating soybean mutant. Applied and Environmental Microbiology **57:** 2205-2210

Caetano-Anollés G, Gresshoff PM (1991b) Plant genetic control of nodulation. Annual Review of Microbiology **45:** 345-382

Carroll BJ, McNeil DL, Gresshoff PM (1985a) Isolation and properties of soybean [*Glycine max* (L.) Merr.] mutants that nodulate in the presence of high nitrate concentrations. Proceedings of the National Academy of Sciences of the United States of America **82:** 4162-4166

Carroll BJ, McNeil DL, Gresshoff PM (1985b) A supernodulation and nitrate-tolerant symbiotic (*nts*) soybean mutant. Plant Physiology **78:** 34-40

Chelle M, Andrieu B (1999) Radiative models for architectural modeling. Agronomie **19:** 225-240

Chelle M, Andrieu B, Bouatouch K (1998) Nested radiosity for plant canopies. The Visual Computer **14:** 109-125

Chelle M, Evers JB, Combes D, Varlet-Grancher C, Vos J, Andrieu B (2007) Simulation of the three-dimensional distribution of the red:far-red ratio within crop canopies. New Phytologist **176:** 223-234

Chelle M, Hanan J, Autret H (2004) Lighting virtual crops: the CARIBU solution for open L-systems. *In* C Godin, J Hanan, W Kurth, A Lacointe, A Takenaka, P Prusinkiewicz, T DeJong, C Beveridge, B Andrieu, eds, 4th International Workshop on Functional–structural Plant Models, Montpellier, p 194

Cieslak M, Lemieux C, Hanan J, Prusinkiewicz P (2008) Quasi-Monte Carlo simulation of the light environment of plants. Functional Plant Biology **35:** 837-849

Danjon F, Reubens B (2008) Assessing and analyzing 3D architecture of woody root systems, a review of methods and applications in tree and soil stability, resource acquisition and allocation. Plant Soil **303:** 1-34

De Moraes CM, Schultz JC, Mescher MC, Tumlinson JH (2004) Induced plant signaling and its implications for environmental sensing. Journal of Toxicology and Environmental Health **67:** 819-834

de Reffye P, Edelin C, Francon J, Jaeger M, Puech C (1988) Plant models faithful to botanical structure and development. Computer Graphics **22:** 151-158

de Reuille PB, Bohn-Courseau I, Ljung K, Morin H, Carraro N, Godin C, Traas J (2006) Computer simulations reveal properties of the cell-cell signaling network at the shoot apex in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **103:** 1627-1632

Delves AC, Higgins A, Gresshoff P (1992) Shoot apex removal does not alter autoregulation of nodulation in soybean. Plant, Cell & Environment **15:** 249-254

Delves AC, Mathews A, Day DA, Carter AS, Carroll BJ, Gresshoff PM (1986) Regulation of the soybean-rhizobium nodule symbiosis by shoot and root factors. Plant Physiology **82:** 588-590

Diggle AJ (1988) ROOTMAP—a model in three-dimensional coordinates of the growth and structure of fibrous root systems. Plant and Soil **105:** 169-178

Doyle JJ, Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiology **131:** 900-910

Drouet J-L, Pagès L (2007) GRAAL-CN: a model of GRowth, Architecture and ALlocation for Carbon and Nitrogen dynamics within whole plants formalised at the organ level. Ecological Modelling **206:** 231-249

Drouet JL, Pagès L (2003) GRAAL: a model of GRowth, Architecture and carbon ALlocation during the vegetative phase of the whole maize plant: Model description and parameterisation. Ecological Modelling **165:** 147-173

Dubrovsky JG, Gambetta GA, Hernández-barrera A, Shishkova S, González I (2006) Lateral root initiation in arabidopsis: developmental window, spatial patterning, density and predictability. Annals of Botany **97:** 903-915

Dun EA, Hanan J, Beveridge CA (2009) Computational modeling and molecular physiology experiments reveal new insights into shoot branching in pea. The Plant Cell **21:** 3459-3472

Ferguson BJ, Indrasumunar A, Hayashi S, Lin M-H, Lin Y-H, Reid DE, Gresshoff PM (2010) Molecular analysis of legume nodule development and autoregulation. Journal of Integrative Plant Biology **52:** 61-76

Ferguson BJ, Mathesius U (2003) Signaling interactions during nodule development. Journal of Plant Growth Regulation **22:** 47-72

Godin C, Costes E, Sinoquet H (1999) A method for describing plant architecture which integrates topology and geometry. Annals of Botany **84:** 343-357

Godin C, Sinoquet H (2005) Functional-structural plant modelling. New Phytologist **166:** 705-708

Graham PH, Vance CP (2003) Legumes: importance and constraints to greater use. Plant Physiology **131:** 872-877

Gresshoff PM (1990) The importance of biological nitrogen fixation to new crop development. *In* J Janick, JE Simon, eds, Advances in New Crops. Timber Press, Portland, pp 113-119

Gresshoff PM (2003) Post-genomic insights into plant nodulation symbioses. Genome Biology **4:** 201

Gresshoff PM, Lohar D, Chan P-K, Biswas B, Jiang Q, Reid D, Ferguson B, Stacey G (2009) Genetic analysis of ethylene regulation of legume nodulation. Plant Signaling & Behavior **4:** 1-6

Haefner JW (2005) Modeling Biological Systems: Principles and Applications. Springer, New York

Halle F, Oldeman RAA, Tomlinson PB (1978) Tropical Trees And Forests: An Architectural Analysis. Springer Verlag, Heidelberg

Hammer GL, Sinclair TR, Chapman SC, Oosterom Ev (2004) On systems thinking, systems biology, and the in silico plant Plant Physiology **134:** 909-911

Han L, Gresshoff PM, Hanan J (2007) Virtual soybean—a computational model for studying autoregulation of nodulation. *In* P Prusinkiewicz, J Hanan, B Lane, eds, The 5th International Workshop on Functional Structural Plant Models, Napier, New Zealand

Han L, Gresshoff PM, Hanan J (2009) Modelling root development with signalling control: a case study based on legume autoregulation of nodulation. *In* B Li, M Jaeger, Y Guo, eds, Plant Growth Modeling and Applications, Proceedings of PMA09. IEEE Computer Society, Los Alamitos, pp 134-141

Han L, Gresshoff PM, Hanan J (2010a) A functional-structural modelling approach to autoregulation of nodulation. Annals of Botany **DOI:** 10.1093/aob/mcq182

Han L, Hanan J, Gresshoff PM (2010b) Computational complementation: a modelling approach to study signalling mechanisms during legume autoregulation of nodulation. PLoS Computational Biology **6:** e1000685

Hanan J (1997) Virtual plants—integrating architectural and physiological models. Environmental Modelling & Software **12:** 35-42

Hanan JS, Room PM (1996) Practical aspects of virtual plant research. *In* MT Michalewicz, ed, Advances in computational life sciences. Kevin Jeans, Collingwood, pp 28-44

Hansen AP, Peoples MB, Gresshoff PM, Atkins CA, Pate JS, Carroll BJ (1989) Symbiotic performance of supernodulating soybean (*Glycine max* (L.) Merrill) mutants during development on different nitrogen regimes. Journal of Experimental Botany **40:** 715-724

Hayashi S, Gresshoff PM, Kinkema M (2008) Molecular analysis of lipoxygenases associated with nodule development in soybean. Molecular Plant-Microbe Interactions **21:** 843-853

Hedden P, Thomas SG (2006) Plant Hormone Signaling. Blackwell Publishing Ltd, Oxford

Hemmerling R, Kniemeyer O, Lanwert D, Kurth W, Buck-Sorlin G (2008) The rule-based language XL and the modelling environment GroIMP illustrated with simulated tree competition. Functional Plant Biology **35:** 739-750

Hill LL, Crosier SJ, Smith TR, Goodchild M (2001) A content standard for computational models. *In* The Magazine of Digital Library Research, Vol 7

Honda H, Tomlinson P, Fisher J (1981) Computer simulation of branch interaction and regulation by unequal flow rates in botanical trees. American Journal of Botany **68:** 569-585

Honda H, Tomlinson P, Fisher J (1982) Two geometrical models of branching of botanical trees. Annals of Botany **49:** 1-12

Jönsson H, Heisler MG, Shapiro BE, Meyerowitz EM, Mjolsness E (2006) An auxin-driven polarized transport model for phyllotaxis. Proceedings of the National Academy of Sciences of the United States of America **103:** 1633-1638

Jaeger M, de Reffye P (1992) Basic concepts of computer simulation of plant growth. Journal of Biosciences **17:** 275-291

Janssen JM, Lindenmayer A (1987) Models for the control of branch positions and flowering sequences of capitula in *Mycelis muralis* (L.) dumont (Compositae). New Phytologist **105:** 191-220

Jones JW, Keating BA, Porter CH (2001) Approaches to modular model development. Agricultural Systems **70:** 421-443

Jourdan C, Rey H (1997) Modelling and simulation of the architecture and development of the oil-palm (*Elaeis guineensis* Jacq.) root system. Plant and Soil **190:** 217-233

Kahl G (1995). *In* Dictionary of Gene Technology. VCH, Weinheim, p 92

Kinkema M, Gresshoff PM (2008) Investigation of downstream signals of the soybean autoregulation of nodulation receptor kinase GmNARK. Molecular Plant-Microbe Interactions **21:** 1337-1348

Kinkema M, Scott PT, Gresshoff PM (2006) Legume nodulation: successful symbiosis through short- and long-distance signalling. Functional Plant Biology **33:** 707-721

Kitano H (2002) Systems biology: a brief overview. Science **295:** 1662-1664

Krogh D (2009) Biology: A Guide To The Natural World Pearson/Benjamin Cummings, San Francisco

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

Li D, Kinkema M, Gresshoff PM (2009) Autoregulation of nodulation (AON) in *Pisum sativum* (pea) involves signalling events associated with both nodule primordia development and nitrogen fixation. Journal of Plant Physiology **166:** 955-967

Lin Y-H, Ferguson BJ, Kereszt A, Gresshoff PM (2010) Suppression of hypernodulation in soybean by a leaf-extracted, NARK- and Nod factor-dependent, low molecular mass fraction. New Phytologist **185:** 1074-1086

Lindenmayer A (1968) Mathematical models for cellular interaction in development, Parts I and II. Journal of Theoretical Biology **18:** 280-315

Lira MdA, Smith DL (2000) Use of a standard TWAIN scanner and software for nodule number determination on different legume species. Soil Biology & Biochemistry **32:** 1463-1467

Lopez G, Favreau RR, Smith C, Costes E, Prusinkiewicz P, DeJong TM (2008) Integrating simulation of architectural development and source–sink behaviour of peach trees by incorporating Markov chains and physiological organ function submodels into L-PEACH. Functional Plant Biology **35:** 761-771

Lucas M, Guédon Y, Jay-Allemand C, Godin C, Laplaze L (2008) An auxin transport-based model of root branching in *Arabidopsis thaliana*. PLoS ONE **3:** e3673

Mallory TE, Chiang S-H, Cutter EG, Gifford EM (1970) Sequence and pattern of lateral root formation in five selected species. American Journal of Botany **57:** 800-809

Mathesius U (2008) Auxin: at the root of nodule development? Functional Plant Biology **35:** 651-668

Men AE, Laniya TS, Searle IR, Iturbe-Ormaetxe I, Gresshoff I, Jiang Q, Carroll BJ, Gresshoff PM (2002) Fast neutron mutagenesis of soybean (*Glycine soja* L.) produces a supernodulating mutant containing a large deletion in linkage group H. Genome Letters **3:** 147-155

Minorsky PV (2003) Achieving the in silico plant. Systems biology and the future of plant biological research Plant Physiology **132:** 404-409

Mitchison G (1980) The dynamics of auxin transport. Proceedings of the Roy Society of London **209:** 489-511

Miyahara A, Hirani TA, Oakes M, Kereszt A, Kobe B, Djordjevic MA, Gresshoff PM (2008) Soybean nodule autoregulation receptor kinase phosphorylates two kinase-associated protein phosphatases *in vitro*. Journal of Biological Chemistry **283:** 25381-25391

Mortier V, Herder GD, Whitford R, de Velde WV, Rombauts S, D'haeseleer K, Holsters M, Goormachtig S (2010) CLE peptides control *Medicago truncatula* nodulation locally and systemically. Plant Physiology **153:** 222-237

Mulligan RM, Chory J, Ecker JR (1997) Signaling in plants. Proceedings of the National Academy of Sciences of the United States of America **94:** 2793-2795

Neves SR, Iyengar R (2002) Modeling of signaling networks. BioEssays **24:** 1110-1117

Nicolis G, Nicolis C (2007) Foundations of Complex Systems: Nonlinear Dynamics, Statistical Physics, Information and Prediction. World Scientific Publishing Co. Pte. Ltd, Hackensack

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

Nontachaiyapoom S, Scott PT, Men AE, Kinkema M, Schenk PM, Gresshoff PM (2007) Promoters of orthologous *Glycine max* and *Lotus japonicus* nodulation autoregulation genes interchangeably drive phloem-specific expression in transgenic plants. Molecular Plant-Microbe Interactions **20:** 769–780

Oka-Kira E, Kawaguchi M (2006) Long-distance signaling to control root nodule number. Current Opinion in Plant Biology **9:** 496-502

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

Oldroyd GED, Downie JA (2004) Calcium, kinases and nodulation signalling in legumes. Nature Reviews Molecular Cell Biology **5:** 566-576

Olsson JE, Nakao P, Bohlool BB, Gresshoff PM (1989) Lack of systemic suppression of nodulation in split root systems of supernodulating soybean (*Glycine max* [L.]Merr.) mutants. Plant Physiology **90:** 1347-1352

Perttunen J, Sievänen R, Nikinmaa E, Salminen H, Saarenmaa H, Väkevä J (1996) LIGNUM: a tree model based on simple structural units. Annals of Botany **77:** 87-98

Pradal C, Dufour-Kowalski S, Boudon F, Fournier C, Godin C (2008) OpenAlea: a visual programming and component-based software platform for plant modellin. Functional Plant Biology **35:** 751-760

Prusinkiewicz P (1998) Modeling of spatial structure and development of plants. Scientia Horticulturae **74:** 113-149

Prusinkiewicz P (2004a) Art and science for life: Designing and growing virtual plants with L-systems. Acta Horticulturae **630:** 5-28

Prusinkiewicz P (2004b) Modeling plant growth and development. Current Opinion in Plant Biology **7:** 79-83

Prusinkiewicz P, Crawford S, Smith RS, Ljung K, Bennett T, Ongaro V, Leyser O (2009) Control of bud activation by an auxin transport switch. Proceedings of the National Academy of Sciences of the United States of America **106:** 17431-17436

Prusinkiewicz P, Karwowski R, Lane B (2007) The L+C plant modelling language. *In* J Vos, LFM Marcelis, PHB de Visser, PC Struik, JB Evers, eds, Functional-Structural Plant Modelling in Crop Production. Springer, Dordrecht

Prusinkiewicz P, Lindenmayer A (1990) The Algorithmic Beauty of Plants. Springer-Verlag New York, Inc, New York

Prusinkiewicz P, Lindenmayer A, Hanan J (1988) Development models of herbaceous plants for computer imagery purposes. Computer Graphics **22:** 141-150

Raven PH, Evert RF, Eichhorn SE (1999) Biology of Plants. W. H. Freeman and Company Worth Publishers, New York

Ray LB, Chong LD, Gough NR (2002) Computational biology. Sci. STKE **2002:** eg10

Redmond JW, Batley M, Djordjevic MA, Innes RW, Kuempel PL, Rolfe BG (1986) Flavones induce expression of nodulation genes in *Rhizobium*. Nature **323:** 632-635

Reid DE, Ferguson BJ, Gresshoff PM (2011) Inoculation- and nitrate-induced CLE peptides of soybean control NARK-dependent nodule formation. Molecular Plant-Microbe Interactions**:** In press

Renton M, Kaitaniemi P, Hanan J (2005) Functional-structural plant modelling using a combination of architectural analysis, L-systems and a canonical model of function. Ecological Modelling **184:** 277-298

Rolfe BG, Gresshoff PM (1988) Genetic analysis of legume nodule initiation. Annual Review of Plant Physiology and Plant Molecular Biology **39:** 297-319

Rolland-Lagan A-G, Prusinkiewicz P (2005) Reviewing models of auxin canalization in the context of leaf vein pattern formation in Arabidopsis. The Plant Journal **44:** 854-865

Room P, Hanan J, Prusinkiewicz P (1996) Virtual plants: new perspectives for ecologists, pathologists and agricultural scientists. Trends in Plant Science **1:** 33-38

Room PM, Maillette L, Hanan JS (1994) Module and metamer dynamics and virtual plants. *In* M Begon, AH Fitter, eds, Advances in Ecological Research, Vol 25. Academic Press, pp 105-157

Schnabel E, Journet E-P, de Carvalho-Niebel F, Duc G, Frugoli J (2005) The *Medicago truncatula SUNN* gene encodes a *CLV1*-like leucine-rich repeat receptor kinase that regulates nodule number and root length. Plant Molecular Biology **58:** 809-822

Scott PT, Pregelj L, Chen N, Hadler JS, Djordjevic MA, Gresshoff PM (2008) *Pongamia pinnata*: an untapped resource for the biofuels industry of the future. BioEnergy Research **1:** 2-11

Searle IR, Men AE, Laniya TS, Buzas DM, Iturbe-Ormaetxe I, Carroll BJ, Gresshoff PM (2003) Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase. Science **299:** 109-112

Sinoquet H, Rivet P (1997) Measurement and visualization of the architecture of an adult tree based on a three-dimensional digitising device. Trees **11:** 265-270

Spaink HP (2000) Root nodulation and infection factors produced by rhizobial bacterial. Annual Review of Microbiology **54:** 257-288

Stamatopoulou I, Kefalas P, Gheorghe M (2007) Modelling the dynamic structure of biological state-based systems. BioSystems **87:** 142-149

Stelling J (2007) Understandable complexity. Sci. STKE **2007:** pe9

Trewavas A (2003) Aspects of plant intelligence. Annals of Botany **92:** 1-20

van Noorden GE, Ross JJ, Reid JB, Rolfe BG, Mathesius U (2006) Defective long-distance auxin transport regulation in the *Medicago truncatula super numeric nodules* mutant. Plant Physiology **140:** 1494–1506

van Riel NAW (2006) Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. Briefings in Bioinformatics **7:** 364-374

Vos J, Evers JB, Buck-Sorlin GH, Andrieu B, Chelle M, de Visser PHB (2010) Functional–structural plant modelling: a new versatile tool in crop science. Journal of Experimental Botany **61:** 2101-2115

Watanabe T, Hanan JS, Room PM, Hasegawa T, Nakagawa H, Takahashi W (2005) Rice morphogenesis and plant Architecture: measurement, specification and the reconstruction of structural development by 3D architectural modelling. Annals of Botany **95:** 1131-1143

Wenden B, Dun EA, Hanan J, Andrieu B, Weller JL, Beveridge CA, Rameau C (2009) Computational analysis of flowering in pea (*Pisum sativum*). New Phytologist **184:** 153 - 167

Weng G, Bhalla US, Iyengar R (1999) Complexity in biological signaling systems. Science **284:** 92-96

Wopereis J, Pajuelo E, Dazzo FB, Jiang Q, Gresshoff PM, de Bruijn FJ, Stougaard J, Szczyglowski K (2000) Short root mutant of *Lotus japonicus* with a dramatically altered symbiotic phenotype. The Plant Journal **23:** 97-114

Yan H-P, Kang MZ, de Reffye P, Dingkuhn M (2004) A dynamic, architectural plant model simulating resource-dependent growth. Annals of Botany **93:** 591-602

Appendices

A.1 Text S1: Description of primary root elongation algorithm in *cpfg*

The following describes the algorithm of the primary root elongation using "*cpfg*" in L-studio; Table A.1.1 lists the identifiers used. The module "RT" represents the root tip module and carries developmental information (such as the current root length). The root structure is composed of a set of sub-modules "R" (with standard length defined as "UNIT") added behind the leading RT module. The standard length UNIT can be set to any value, depending on the accuracy or level of detail that is required.

Table A.1.1. Identifiers defined in "*cpfg*".

Identifier	**Description**
RT	the module representing the primary root tip
R	the sub-module representing an elementary unit of root architecture
P_N	the module representing a potential nodulation site on primary root
P_L	the module representing a potential branching site on primary root
UNIT	the user-defined standard length for module R
len	the variable to calculate overall root length
nd_itvl	the counter to match successive nodule interval
ltr_itvl	the counter to match successive lateral interval
region	the marker of each primary-root region
PN_Fd[region]	the array providing data on the distance between the first nodule of a nodulation section and the starting point of the primary root
PN_Ld[region]	the array providing data on the distance between the last nodule of a nodulation section and the starting point of the primary root
Nodule_Interval[region]	the array providing data on the nodule density of a nodulation section
Lateral_Interval[region]	the array providing data on the interval between two successive lateral roots in a primary-root region

With some other identifiers described in Table A.1.1, the algorithm can be formulated in L-system syntax as:

/*Rule 1*/

/*When the primary root elongates into a section for future nodulation, a site with potential for nodule formation is made available if the current location of root tip matches the positions of the first or last nodules of this section or if the elongation has covered an interval between two successive nodules.*/

RT(len, nd_itvl, ltr_itvl):

(len>=PN_Fd[region]
&& len<=PN_Ld[region]
&& (len==PN_Fd[region] || len==PN_Ld[region] || nd_itv==Nodule_Interval[region])
)
{ *The statements bracketed by braces are executed if and only if the module is found in the string and the condition is matched*/
len=len+UNIT;
nd_itvl=0;
ltr_itvl=ltr_itvl+UNIT;
}
→ [P_N] R RT(len, nd_itv, ltr_itv)

/*Rule 2*/

/*If the primary root elongation has covered an interval between two successive laterals, a site potential for lateral root emission is made available.*/

RT(len, nd_itv, ltr_itv):

(ltr_itvl==Lateral_Interval[region])
{
len=len+UNIT;
nd_itvl= nd_itvl+UNIT;
ltr_itvl=0;
}
→ [P_L] R RT(len, nd_itv, ltr_itv)

/*Rule 3*/

/*If the current position of primary root tip does not match the conditions in Rule 1 and Rule 2, it is neither potential for nodulation nor for lateral emission.*/

RT(len, nd_itv, ltr_itv):

(1)
{
len=len+UNIT;
nd_itvl= nd_itvl+UNIT;
ltr_itvl= ltr_itvl+UNIT;
}
→ R RT(len, nd_itv, ltr_itv)

A.2 Text S2: Materials and methods for glasshouse experiments

Soybean growth experiment for architectural data collection

Soybean seeds of wild type parent Bragg and derived mutant *nts1116* (surface-sterilised with 3% H_2O_2 in 70% ethanol for 3 minutes) were sown into autoclaved pots (20 cm in diameter) filled with autoclaved grade 2 vermiculite. The pots were then watered until complete saturation. All plants were kept in an air-conditioned glasshouse with controlled temperature of 28 ºC during day and 23 ºC at night. On the day after sowing, the seedlings were inoculated with a commercial peat containing *Bradyrhizobium japonicum* strain CB1809 (130 g peat CB1809 mixed with 2,800 mL water). Five plants were kept in each pot after culling on the fifth day. The plants were fertilised twice during this experiment – on the third day and the eighth day after sowing - with B&D nutrient solution (Broughton and Dilworth, 1971) plus 2 mM KNO_3. This level of nitrate stimulates plant growth and has minimal effect on nodulation in soybean (Carroll et al., 1985a). The developmental data of shoot and root components, such as internode, cotyledon, unifoliate leaf, trifoliate leaf and leaflets, primary root, lateral root and nodules (including nodule number and nodule distribution) were collected every two days from the third day after sowing by destructive sampling of five plants for each genotype.

Soybean graft experiment

Soybean seeds of Bragg and *nts1116* were surface-sterilised, planted, and grown as described above. Grafting in the hypocotyl or epicotyl was carried out ten days post-sowing and the grafted plants were inoculated with a liquid YMB culture (2 g/L mannitol, 0.4 g/L yeast extract, 0.5 g/L K_2HPO_4, 0.2 g/L $MgSO_4 \cdot 7H_2O$, 0.1 g/L NaCl, pH 6.8) of CB1809 on the day after grafting. The plants were irrigated with B&D nutrient plus 1 mM KNO_3 solution. The nodule numbers were scored from the grafted plants two weeks after grafting.

A.3 Text S3: Growth data collection and model construction

Methods for growth data collection

To collect growth data, the shoot structure was mapped as sequences of symbols (Room et al., 1996) where each symbol represents a particular type of plant component (such as internode, cotyledon, petiole and leaflet). For the root system, we classify its components as primary root, first-order lateral roots and nodules. The second-order lateral roots were not taken into account in this work. To capture the branching pattern and help in identifying the first-order lateral roots, we developed an "RULD" root mapping method (Han et al., 2007) (Figure A.3.1). We also developed a nodule positioning method to help in recording the distribution information of nodulation (Figure A.3.2).

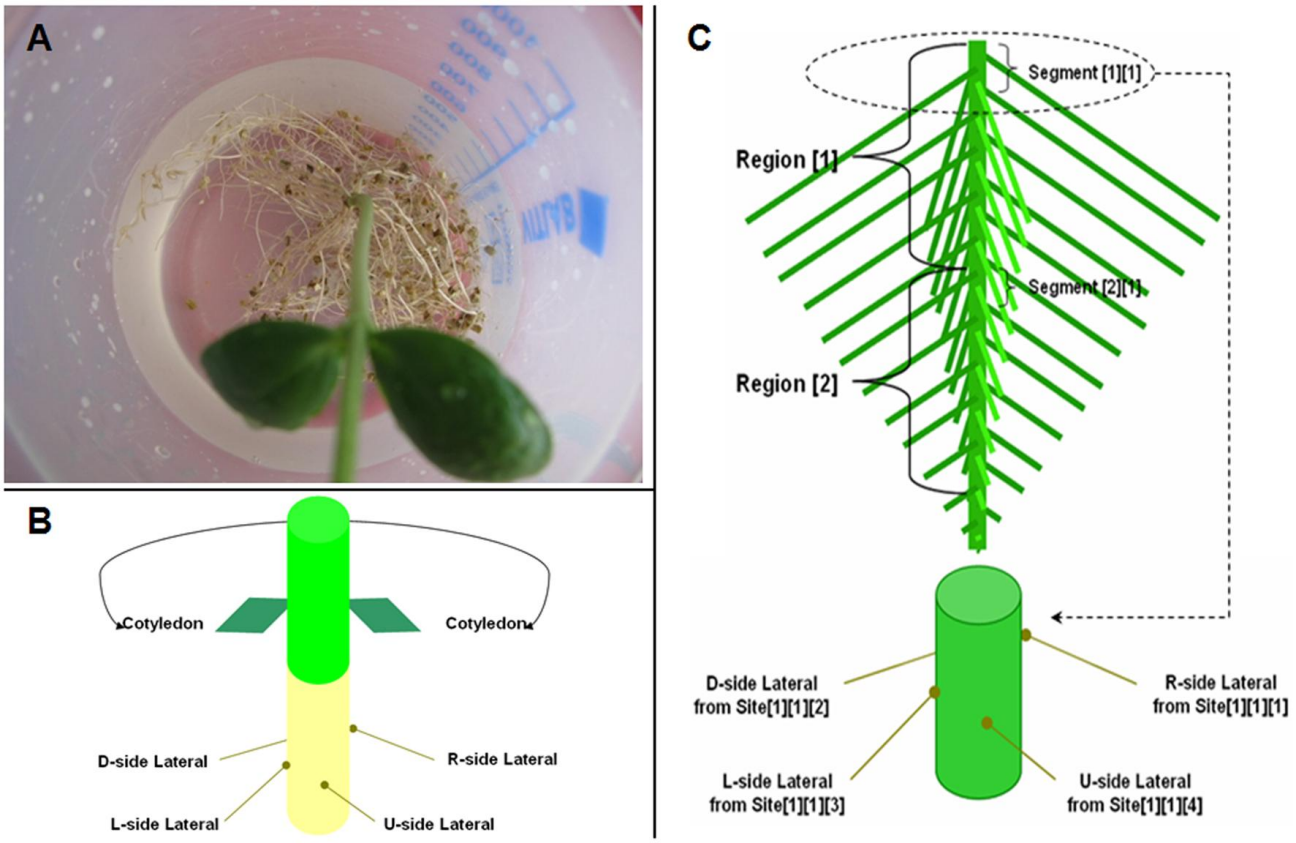


Figure A.3.1. The "RULD" root mapping method. Based on the observed lateral root branching pattern from soybean plants (as shown in image A), the first-order lateral roots were characterised into four sides (R, U, L and D) according to the relative positions of their starting points to the obtuse angle composed by cotyledons in the horizontal plane (as demonstrated in image B). Then the first-order lateral roots were further classified by regions (50 mm long), segments and sites (as illustrated in image C).

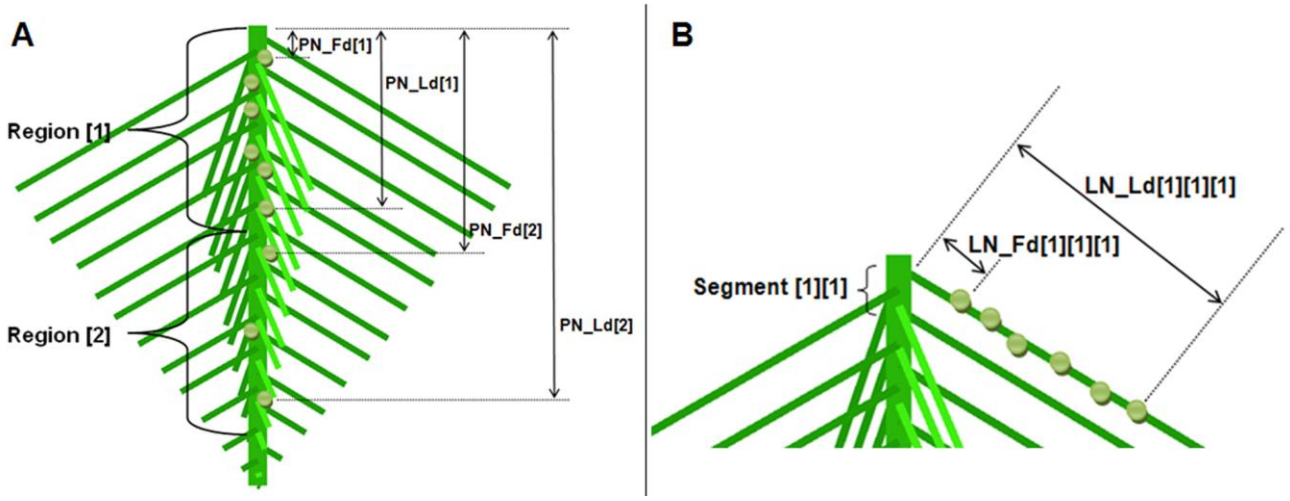


Figure A.3.2. Method for recording distribution information of nodulation. (*A*) For each region on the primary root (as defined in Figure A.3.1), the distance between the first nodule and the primary-root starting point was measured and recorded as "PN_Fd"; the distance between the last nodule and the primary-root starting point was measured and recorded as "PN_Ld". (*B*) For the lateral root nodules, the same method was used with the distances (LN_Fd and LN_Ld) measured from the lateral-root starting point. The positions of the first nodule and the last nodule determine the location of a nodulation section. The nodule density for a section was calculated based on its length and the number of nodules within the section.

Growth data

In analysing the raw data collected with the above methods, some mean values decreased with increasing day, due to the destructive sampling for measurement. The data (except for cotyledon dry weight) were processed to remove this anomaly by not allowing values to decrease, so that normal growth patterns could be simulated. For example, if l_6 and l_8 represent the length of an internode on the 6th day and the 8th day and if l_6 is bigger than l_8, then the value of l_8 is changed to be the same as l_6. The processed architectural data are given in Tables A.3.1 – A.3.12. The data on *nts1116* leaf dry weight, needed for building the functional-structural model, are given in Table A.3.13.

Table A.3.1. Dimension of shoot organs (Bragg).

Organ Dimension (mm) \ Day					4	6	8	10	12	14	16
Phytomer 1	Internode	Length			17.3	33.0	65.6	77.4	77.4	77.4	77.4
		Diameter			2.1	2.3	2.6	2.9	2.9	2.9	2.9
	Cotyledon 1	Length			16.6	20.1	29.6	30.9	31.2	31.2	31.2
		Width			8.8	12.2	17.2	18.6	18.6	18.6	18.6
	Cotyledon 2	Length			16.6	20.1	29.8	29.9	30.6	30.6	30.6
		Width			8.8	12.2	16.8	16.8	18.0	18.0	18.0
Phytomer 2	Internode	Length			0	0	20.7	54.7	69.2	85.8	95.2
		Diameter			0	0	1.8	1.9	2.0	2.1	2.4
	Unifoliate 1	Petiole	Length		0	0.8	5.2	9.8	10.6	11.8	15.5
			Diameter		0	0	1.1	1.1	1.1	1.3	1.3
		Surface	Length		0	10.5	12.0	18.9	23.2	26.0	27.0
			Width		0	4.2	8.2	19.0	26.0	30.6	34.8
	Unifoliate 2	Petiole	Length		0	0.8	5.1	8.2	12.2	15.2	18.8
			Diameter		0	0	1.1	1.1	1.1	1.2	1.5
		Surface	Length		0	10.5	10.5	19.0	27.6	31.2	33.5
			Width		0	4.2	6.3	18.8	31.2	37.2	44.6
Phytomer 3	Internode	Length			0	0	0	0	9.6	32.8	52.0
		Diameter			0	0	0	0	1.7	2.2	2.7
	Trifoliate	Petiole	Length		0	0	0	0	7.0	26.6	49.2
			Diameter		0	0	0	0	1.3	2.8	2.8
		Leaflet 1	Petiolule	Length	0	0	0	0	3.0	4.4	5.7
				Diameter	0	0	0	0	0.9	1.2	1.5
			Surface	Length	0	0	0	10.4	25.8	59.2	80.2
				Width	0	0	0	6.0	14.8	36.0	49.2
		Leaflet 2	Petiolule	Length	0	0	0	0	1.8	2.0	3.1
				Diameter	0	0	0	0	0.8	1.0	1.4
			Surface	Length	0	0	0	0	23.4	51.2	72.7
				Width	0	0	0	0	9.2	30.0	39.3
		Leaflet 3	Petiolule	Length	0	0	0	0	1.8	2.0	3.1
				Diameter	0	0	0	0	0.8	1.0	1.3
			Surface	Length	0	0	0	0	22.0	51.8	75.1
				Width	0	0	0	0	9.0	34.6	40.5
Phytomer 4	Internode	Length			0	0	0	0	0	3.4	13.2
		Diameter			0	0	0	0	0	1.3	2.1
	Trifoliate	Petiole		Length	0	0	0	0	0	2.8	15.5
				Diameter	0	0	0	0	0	1.0	1.5
		Leaflet 1	Petiolule	Length	0	0	0	0	0	2.2	5.7
				Diameter	0	0	0	0	0	0.8	1.3
			Surface	Length	0	0	0	0	0	19.0	46.6
				Width	0	0	0	0	0	8.8	27.6
		Leaflet 2	Petiolule	Length	0	0	0	0	0	1.0	3.0
				Diameter	0	0	0	0	0	0.7	1.2
			Surface	Length	0	0	0	0	0	15.6	34.5
				Width	0	0	0	0	0	4.6	20.2
		Leaflet 3	Petiolule	Length	0	0	0	0	0	1.0	2.7
				Diameter	0	0	0	0	0	0.7	1.2
			Surface	Length	0	0	0	0	0	16.6	33.9
				Width	0	0	0	0	0	4.4	20.2

Table A.3.2. Length of primary and lateral roots (Bragg).

Root Length (mm) \ Day		4	6	8	10	12	14	16
Primary root length		88.3	141.2	231.4	293.2	301.7	301.7	303.8
Average lateral root length	Region 1	0	10	73.6	135.9	135.9	135.9	151.7
	Region 2	0	0	18.4	32.5	47.2	53.4	53.4
	Region 3	0	0	7	18.4	37.9	37.9	37.9
	Region 4	0	0	0	16.3	27.3	27.3	34
	Region 5	0	0	0	0	8	20.3	24

Table A.3.3. Number of primary and lateral root nodules (Bragg).

Nodule Number \ Day		4	6	8	10	12	14	16
Region 1	Number of primary root nodules	0	0	2	2	5	5	5
	Number of lateral root nodules (overall)	0	0	0	0	2	19	25
Region 2	Number of primary root nodules	0	0	0	0	2	2	2
	Number of lateral root nodules (overall)	0	0	0	0	0	2	14
Region 3	Number of primary root nodules	0	0	0	0	1	1	1
	Number of lateral root nodules (overall)	0	0	0	0	0	0	2
Region 4	Number of primary root nodules	0	0	0	0	0	0	0
	Number of lateral root nodules (overall)	0	0	0	0	0	0	0
Region 5	Number of primary root nodules	0	0	0	0	0	0	0
	Number of lateral root nodules (overall)	0	0	0	0	0	0	0

Table A.3.4. Data for nodule distribution (Bragg).

Region	Root Type	Position of First Nodule (mm)	Position of Last Nodule (mm)	Nodule Interval (mm)
Region 1	Primary root	8.5	25.4	6.1
	Lateral roots (average)	29.5	87.3	16.5
Region 2	Primary root	67.3	77.1	12.6
	Lateral roots (average)	44.8	52.3	3.8
Region 3	Primary root	114.5	121.1	10.6
	Lateral roots (average)	N/A	N/A	N/A
Region 4	Primary root	N/A	N/A	N/A
	Lateral roots (average)	N/A	N/A	N/A
Region 5	Primary root	N/A	N/A	N/A
	Lateral roots (average)	N/A	N/A	N/A

Table A.3.5. Data for nodule growth (Bragg).

Days from nodule initialisation to appearance	Nodule diameter expansion rate (mm/day)	Maximum nodule diameter (mm)
2	0.19	2.6

Table A.3.6. Data for lateral root distribution (Bragg).

Region	Number of Segments	Lateral Interval (mm)	Probability of Lateral Root Generation			
			R Laterals	U Laterals	L Laterals	D Laterals
Region 1	17	0.7	0.7	0.7	0.5	0.6
Region 2	16	0.8	0.6	0.6	0.6	0.6
Region 3	14	0.9	0.6	0.5	0.5	0.4
Region 4	13	1	0.4	0.3	0.4	0.5
Region 5	12	1	0.4	0.4	0.5	0.2

Table A.3.7. Dimension of shoot organs (*nts1116*).

Organ Dimension (mm) \ Day					4	6	8	10	12	14	16
Phytomer 1	Internode	Length			20.4	42.2	77.8	77.8	77.8	77.8	79.6
		Diameter			1.7	1.7	2.3	2.3	2.3	2.4	2.4
	Cotyledon 1	Length			13.6	20.6	25.6	26.8	26.8	26.8	26.8
		Width			8.8	12.2	15.4	16	16	16.3	16.3
	Cotyledon 2	Length			13.6	20.2	25.4	26.2	26.2	26.2	26.8
		Width			8.8	12.6	15.8	16.2	16.2	16.2	16.2
Phytomer 2	Internode	Length			0	0	21.6	54.4	80.9	83.3	83.3
		Diameter			0	0	1.4	1.6	1.8	2	2.1
	Unifoliate 1	Petiole	Length		0	1.8	3.6	7	13	14.3	14.3
			Diameter		0	0	0.7	0.9	1.1	1.2	1.2
		Surface	Length		0	11	15.2	26.6	30.9	31	37.6
			Width		0	4.4	12.4	25	36.6	38	41.2
	Unifoliate 2	Petiole	Length		0	1.8	3.8	6.8	11.9	19.5	19.5
			Diameter		0	0	0.8	0.9	1.3	1.4	1.4
		Surface	Length		0	11	17.6	24.8	32.4	37.3	38.4
			Width		0	4.4	15	24.6	37.8	38.5	44.4
Phytomer 3	Internode	Length			0	0	0	0	7	21.7	28.6
		Diameter			0	0	0	0	1.4	1.8	2.1
	Trifoliate	Petiole	Length		0	0	0	0	3.9	16.8	23.4
			Diameter		0	0	0	0	1.1	1.3	1.6
		Leaflet 1	Petiolule	Length	0	0	0	0	2	5	6.2
				Diameter	0	0	0	0	0.6	1.2	1.3
			Surface	Length	0	0	0	8.2	16.8	42.5	55.4
				Width	0	0	0	2.4	4.9	26	34.4
		Leaflet 2	Petiolule	Length	0	0	0	0	1.6	2.3	2.3
				Diameter	0	0	0	0	0.6	1	1.1
			Surface	Length	0	0	0	0	14.5	37.2	50.6
				Width	0	0	0	0	4	27	28.4
		Leaflet 3	Petiolule	Length	0	0	0	0	1.5	2.5	2.5
				Diameter	0	0	0	0	0.6	1	1.1
			Surface	Length	0	0	0	0	14.3	40.7	51.4
				Width	0	0	0	0	3.9	20.7	29
Phytomer 4	Internode	Length			0	0	0	0	0	3	3.6
		Diameter			0	0	0	0	0	1.4	1.4
	Trifoliate	Petiole		Length	0	0	0	0	0	2.3	3.6
				Diameter	0	0	0	0	0	0.8	1
		Leaflet 1	Petiolule	Length	0	0	0	0	0	1.5	2.8
				Diameter	0	0	0	0	0	0.8	0.9
			Surface	Length	0	0	0	0	0	14.8	20.8
				Width	0	0	0	0	0	4	10.4
		Leaflet 2	Petiolule	Length	0	0	0	0	0	0.8	1.4
				Diameter	0	0	0	0	0	0.7	0.8
			Surface	Length	0	0	0	0	0	11.8	16
				Width	0	0	0	0	0	3	7.2
		Leaflet 3	Petiolule	Length	0	0	0	0	0	1	1.4
				Diameter	0	0	0	0	0	0.6	0.8
			Surface	Length	0	0	0	0	0	11.8	17.2
				Width	0	0	0	0	0	3.2	7

Table A.3.8. Length of primary and lateral roots (*nts1116*).

Root Length (mm) \ Day		4	6	8	10	12	14	16
Primary root length		80.8	164.4	214.8	231.5	243	243	253.8
Average lateral root length	Region 1	0	12.9	65.1	98.2	98.2	138.4	141.4
	Region 2	0	0	17.4	24.3	33.5	33.5	33.5
	Region 3	0	0	5.4	14.4	16.1	16.4	28.5
	Region 4	0	0	0	10	32.3	32.3	32.3
	Region 5	0	0	0	2.8	13.1	13.1	20.8

Table A.3.9. Number of primary and lateral root nodules (*nts1116*).

Nodule Number \ Day		4	6	8	10	12	14	16
Region 1	Number of primary root nodules	0	0	7	14	14	14	14
	Number of lateral root nodules (overall)	0	0	0	5	53	74	94
Region 2	Number of primary root nodules	0	0	6	6	8	8	8
	Number of lateral root nodules (overall)	0	0	0	0	2	9	13
Region 3	Number of primary root nodules	0	0	0	1	3	5	5
	Number of lateral root nodules (overall)	0	0	0	0	0	7	7
Region 4	Number of primary root nodules	0	0	0	0	0	0	0
	Number of lateral root nodules (overall)	0	0	0	0	0	0	0
Region 5	Number of primary root nodules	0	0	0	0	0	0	0
	Number of lateral root nodules (overall)	0	0	0	0	0	0	0

Table A.3.10. Data for nodule distribution (*nts1116*).

Region	Root Type	Position of First Nodule (mm)	Position of Last Nodule (mm)	Nodule Interval (mm)
Region 1	Primary root	6.3	43.9	2.5
	Lateral roots (average)	11.2	82.6	6.5
Region 2	Primary root	55.5	88.9	4.5
	Lateral roots (average)	15.5	31.8	3.3
Region 3	Primary root	107.8	124	7.4
	Lateral roots (average)	11.5	18.5	3.5
Region 4	Primary root	N/A	N/A	N/A
	Lateral roots (average)	N/A	N/A	N/A
Region 5	Primary root	N/A	N/A	N/A
	Lateral roots (average)	N/A	N/A	N/A

Table A.3.11. Data for nodule growth (*nts1116*).

Days from nodule initialisation to appearance	Nodule diameter expansion rate (mm/day)	Maximum nodule diameter (mm)
2	0.16	2.6

Table A.3.12. Data for lateral root distribution (*nts1116*).

Region	Number of Segments	Lateral Interval (mm)	Probability of Lateral Root Generation			
			R Laterals	U Laterals	L Laterals	D Laterals
Region 1	16	0.8	0.6	0.6	0.6	0.7
Region 2	13	0.9	0.6	0.5	0.6	0.5
Region 3	12	1.1	0.5	0.6	0.5	0.4
Region 4	13	1	0.4	0.3	0.4	0.4
Region 5	6	2.1	0.4	0.5	0.6	0.3

Table A.3.13. Leaf dry weight (*nts1116*).

Leaf \ Day		4	6	8	10	12	14	16
Phytomer 1	Cotyledon 1	63.9 mg	46.8 mg	41.4 mg	33.9 mg	25.8 mg	24.9 mg	22.8 mg
	Cotyledon 2	63.9 mg	46.8 mg	41.4 mg	33.9 mg	25.8 mg	24.9 mg	22.8 mg
Phytomer 2	Unifoliate 1	0 mg	2.8 mg	7.0 mg	12.4 mg	17.6 mg	17.6 mg	17.6 mg
	Unifoliate 2	0 mg	2.8 mg	7.0 mg	12.4 mg	17.6 mg	17.6 mg	17.6 mg
Phytomer 3	Trifoliate	0 mg	0 mg	0 mg	7.0 mg	23.3 mg	39.7 mg	60.9 mg
Phytomer 4	Trifoliate	0 mg	0 mg	0 mg	0 mg	5.6 mg	11.2 mg	27.9 mg

Methods for model construction

The architectural and functional-structural models were built with the L-system-based language "*cpfg*" in L-studio (Prusinkiewicz and Lindenmayer, 1990; Prusinkiewicz, 2004a). In L-system-based architectural models (Room et al., 1996; Renton et al., 2005; Watanabe et al., 2005), an organ-scale component (such as in internode) is usually simulated with a single module that not only contains its growth information but also represents its structure graphically. In this study, to enable the synchronisation of signalling and developmental processes with different rates and to visualise signal distribution details, we eliminate the graphical role from such modules and only used them to produce a set of standard "sub-modules" for simulation of shoot and root elongation. All sub-modules have the same length, defined as "UNIT" in this work, for both shoot and root structures. For example, for an internode, a leading module "I" is defined to restrict growth (such as elongation, etc.) based on empirical developmental data, while a set of sub-modules "D" play the role of making up structure of the internode. The addition of "D" sub-modules is determined by the empirical elongation data for this internode through the "I" module (Figure A.3.3).

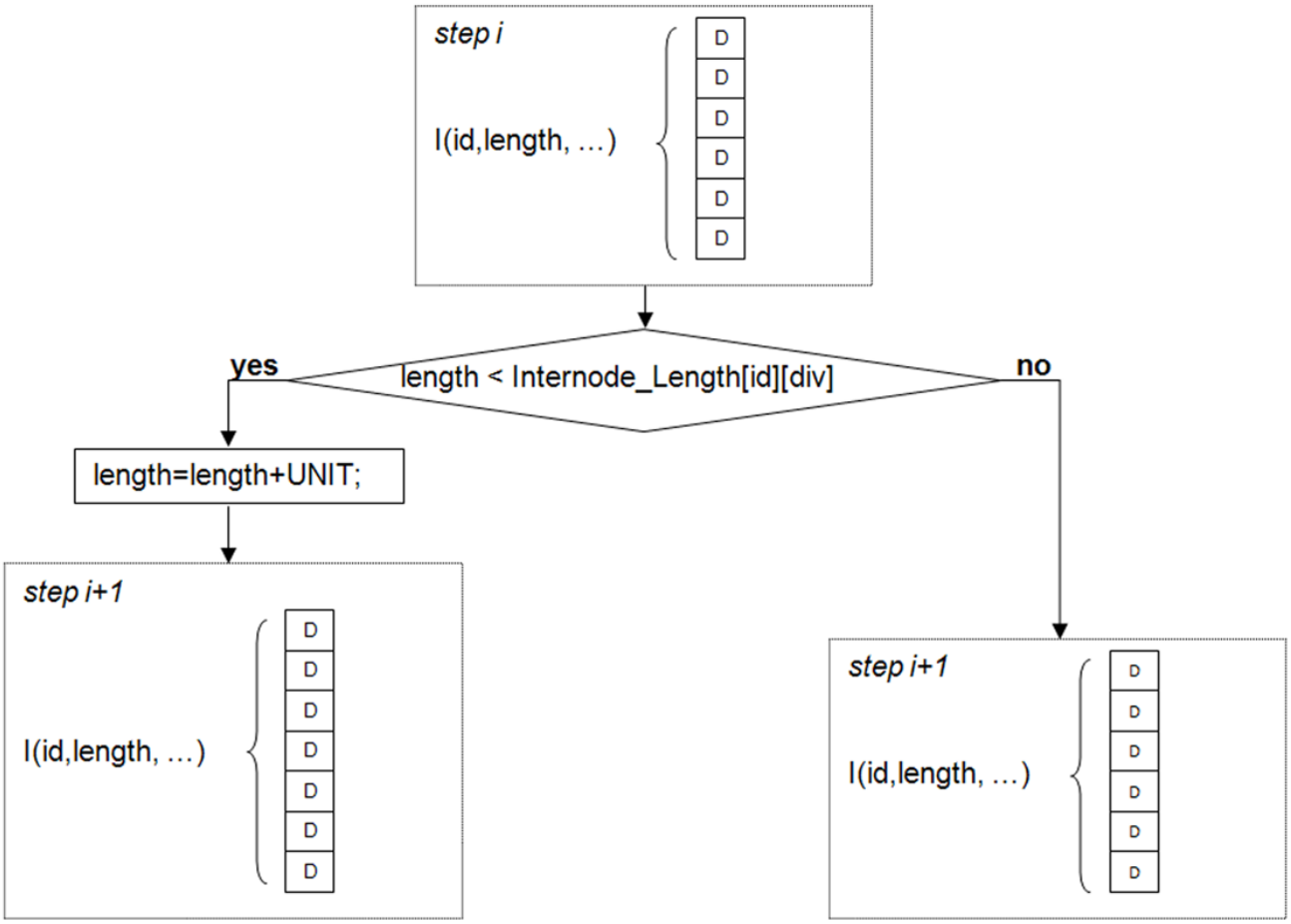


Figure A.3.3. Algorithm with sub-modules for internode elongation. The leading module I(id,length,…) carries the basic developmental information of an internode, where "id" is the identification number of this internode and "length" is the current length of this internode. The array "Internode_Length[id][div]" contains the length information of all internodes over time, which is derived from empirical growth data. The "div" is the most basic time unit that can be defined by the user (e.g. it could be an hour or could also be a day). A "div" contains multiple L-system time steps. At a certain step i, if the length of an internode has not reached the maximum length for the current div (say, the value of Internode_Length[id][div] that corresponds to this internode), it will be increased with a standard length "UNIT", meanwhile a sub-module "D" representing such an increment will be added to existing sub-modules to elongate the internode. Otherwise no addition will be made to the group of sub-modules and the internode length will remain at its previous value. The expressions in this flow chart are in C syntax.

The production of organ-scale components of the shoot is based on a set of developmental rules and relevant empirical data (Figure A.3.4). These production rules are similar to those used in previous L-system-based architectural models (Room et al., 1996; Renton et al., 2005; Watanabe et al., 2005).

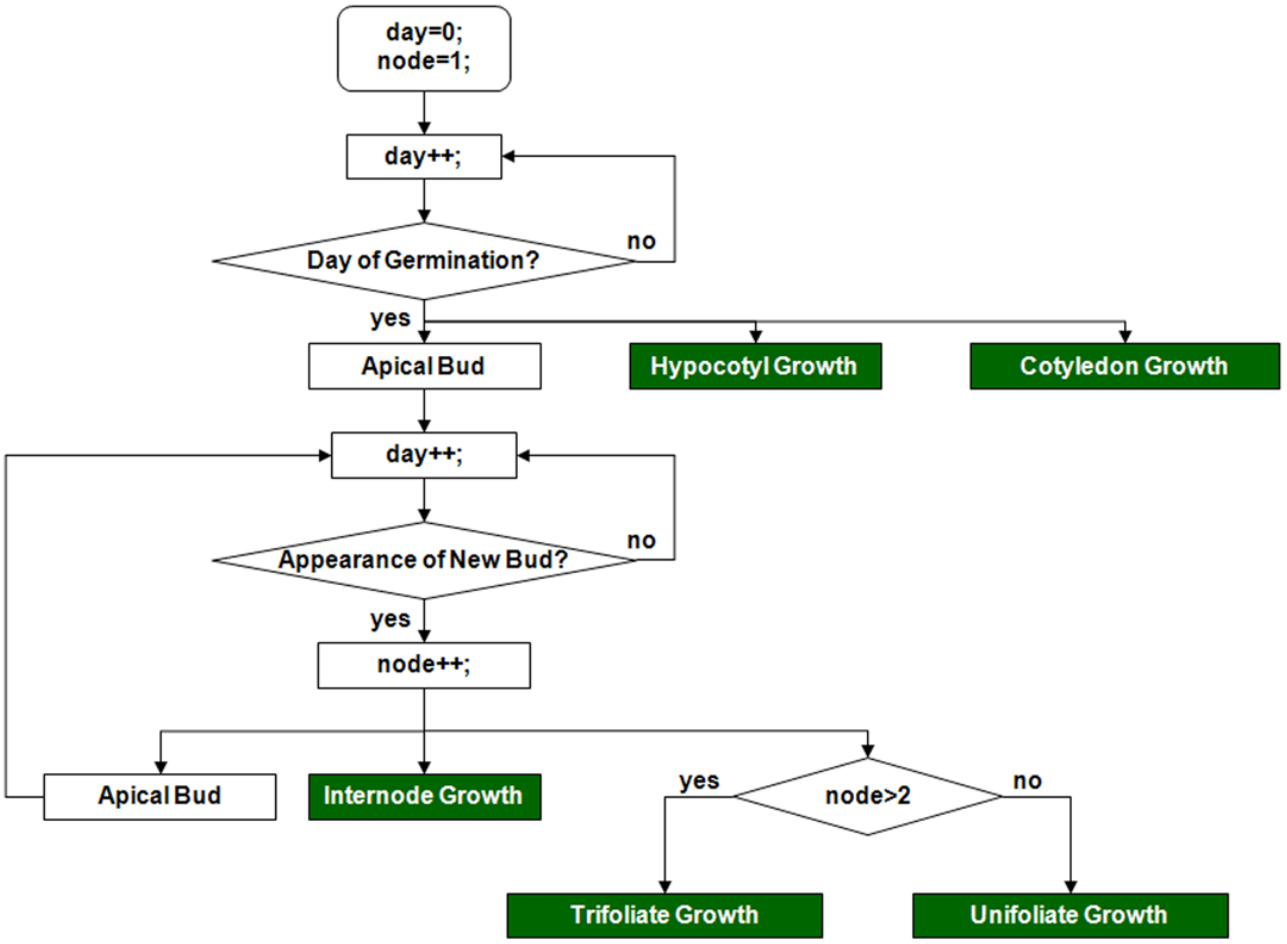


Figure A.3.4. Basic algorithm for production of shoot components. The growth of main stem is led by an apical bud. Once the day of new bud appearance is met, a new node will be added and relevant organ components such as internode and leaf will be produced. The growth of these components (coloured in dark green) behaves individually and simultaneously with the advancement of the main stem. The expressions in this flow chart are in C syntax.

Compared with the clearly "ordered" shoot composition and growth, the development of root system is more complex. To model the elongation of the primary root, we use a module "K" to represent the root tip and to carry the developmental information. When the root tip is heading downwards, a set of sub-modules (with standard length UNIT) is added behind the "K" module to create the primary root structure. Meanwhile, the potential nodulation positions and lateral formation sites are also made available (Figure A.3.5) according to empirical data collected with the root mapping (Figure A.3.1) and nodule recording (Figure A.3.2) methods.

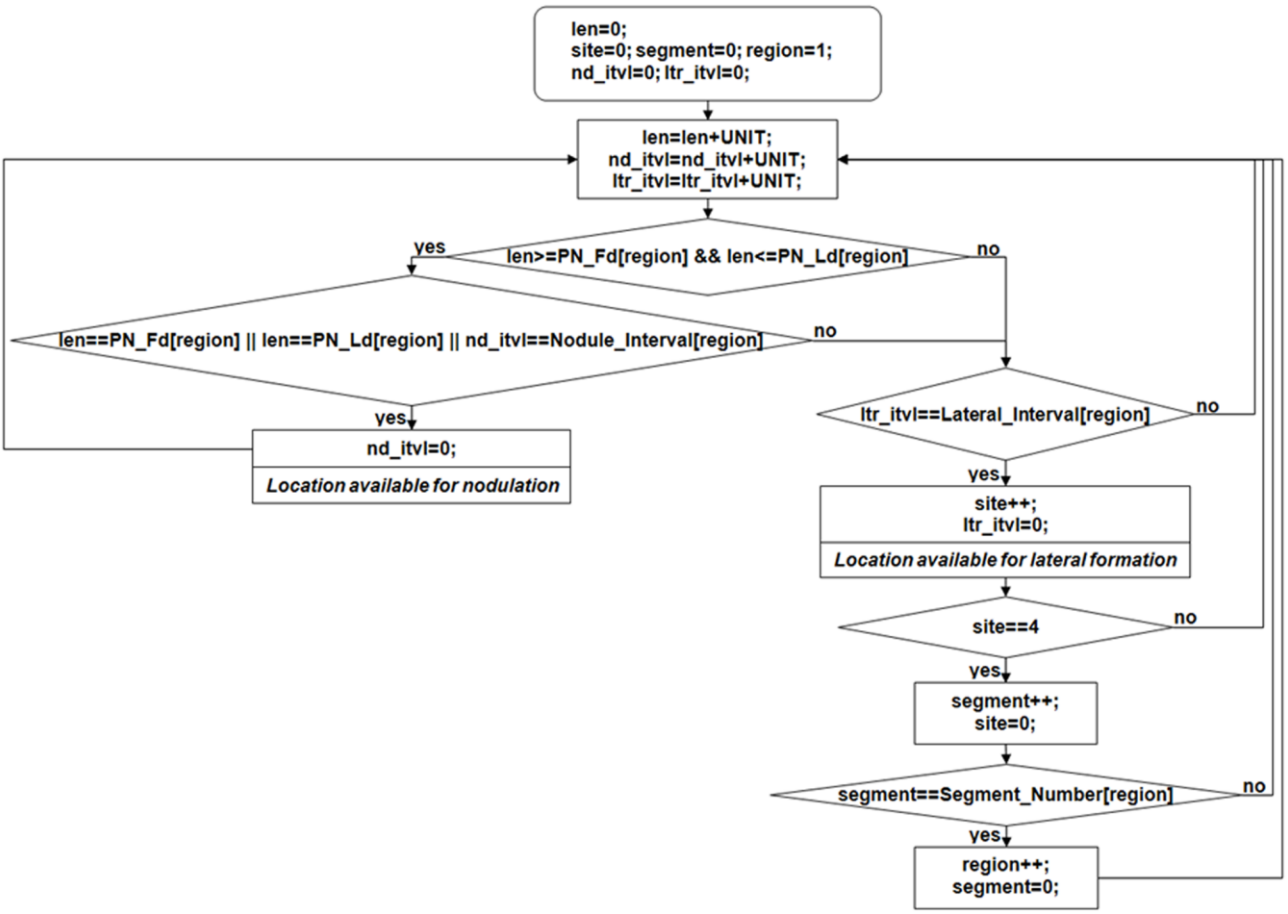


Figure A.3.5. Algorithm for primary root elongation. The variable "len" represents the current length of primary root. The "site", "segment" and "region" are counters to mark the site, segment and region (Figure A.3.1) where the root tip is currently located. The "nd_itvl" and the "ltr_itvl" are counters to match the interval between two nodules and the interval between two sites. The values of these intervals vary depending on different regions. When the root length reaches the values of PN_Fd[region] or PN_Ld[region], or when it is between PN_Fd[region] and PN_Ld[region] (as shown by Figure A.3.2) and the value of "nd_itvl" reaches the interval between two neighbouring nodules (represented by Nodule_Interval[region]), a location potential for nodule formation will be made available. When the root elongation covers a lateral site interval (represented by Lateral_Interval[region]), a site potential for lateral formation will be made available. When the site number reaches four, a new segment will be started. And when all segments in the current region (represented by Segment_Number[region]) are paved, a new region will be started. The expressions in this flow chart are in C syntax.

In the functional-structural model where the signalling activities are enabled, the signal transport between two neighbouring sub-modules is simulated with context-sensitive L-systems (Prusinkiewicz and Lindenmayer, 1990). Compared with traditional context-sensitive L-system

modelling where signals are passed from one organ (such as an internode) directly to another, the use of sub-modules (with standard length UNIT) here allows the signals to flow through and/or stop at a certain point within an organ. This improves the accuracy of signal transport, from a spatial point of view. And it also enables the visualisation of signal concentration variation at different locations of the same organ. However, since the developmental events (such as elongation) and the signalling events (such as signal transport) are based on various rates, the capability of signal exchange between two sub-modules was not enough and a coordination mechanism was required to synchronise these multi-rate events. We therefore developed a multi-rate synchronisation algorithm for the functional-structural model (Figure A.3.6). The key role of this algorithm is still played by the standard increment UNIT. In the functional-structural model, the UNIT is not only added to the growing structure as a spatial unit for elongation, but also used with L-system time steps as the basis of system clock (each single step corresponds with a single UNIT). A user-defined constant "DIV" is used in the algorithm to divide a whole day's time into lower-scale time sections (e.g. 24 hours). Other constants and variables involved in the algorithm can be described as $(E, R, T, C, \tau, \lambda, \theta, u, d, a, step, div, day, counter)$ by Equations A.3.1 - A.3.11.

$$E = \{e_i \mid i = 1,2,...n\} \tag{A.3.1}$$

E is a set of signalling and developmental events (such as signal transport, root elongation, internode elongation and so forth).

$$e_i = \begin{cases} 1\,(\text{activated}) \\ 0\ (\text{stopped}) \end{cases} \tag{A.3.2}$$

Each element e_i in E has two alternative values: 1 and 0. When e_i is equal to 1, its corresponding event (event *i*) is in an "activated" mode and keep going; otherwise the corresponding event is in a "stopped" mode and keep waiting.

$$R = \{r_i \mid i = 1,2,...n\} \tag{A.3.3}$$

Elements in R represent the transport or growth rates (mm/day) of elements in E.

$$\tau : E \to R \tag{A.3.4}$$

The τ means that a one-to-one mapping relationship exists between elements in E and elements in R.

$$T = \{t_i \mid i = 1,2,...n\} \quad \text{(A.3.5)}$$

$$u = UNIT \quad \text{(A.3.6)}$$

$$d = DIV \quad \text{(A.3.7)}$$

$$\lambda : t_i = ceil(r_i / (u * d)) \quad \text{(A.3.8)}$$

Each element t_i in T is the number of time steps that it takes to extend or move through a distance equal to the value of r_i in R; u is a constant equal to the value of UNIT; d is a constant equal to the value of DIV; the relationship between R and T is represented by λ.

$$a = \max(t_i) \quad \text{(A.3.9)}$$

The value of a is equal to the maximum value in T.

$$C = \{c_i \mid i = 1,2,...n\} \quad \text{(A.3.10)}$$

The set C defines the number of available time steps remaining for advancement of each signalling or developmental event during a time section (divided by DIV from a day's time).

$$\theta : E \to C \quad \text{(A.3.11)}$$

The definition of θ means that a one-to-one mapping relationship exists between elements in E and elements in C.

Variable *step* is used to count the number of L-system time steps conducted. From the start of system running, every group of a time steps conducted by the system is counted by a variable *div*. Variable *day* marks the current day. Variable *counter* counts how many L-system time steps are

still available for the current time section, initialised with the value of a. The flow chart of the algorithm is illustrated by Figure A.3.6.

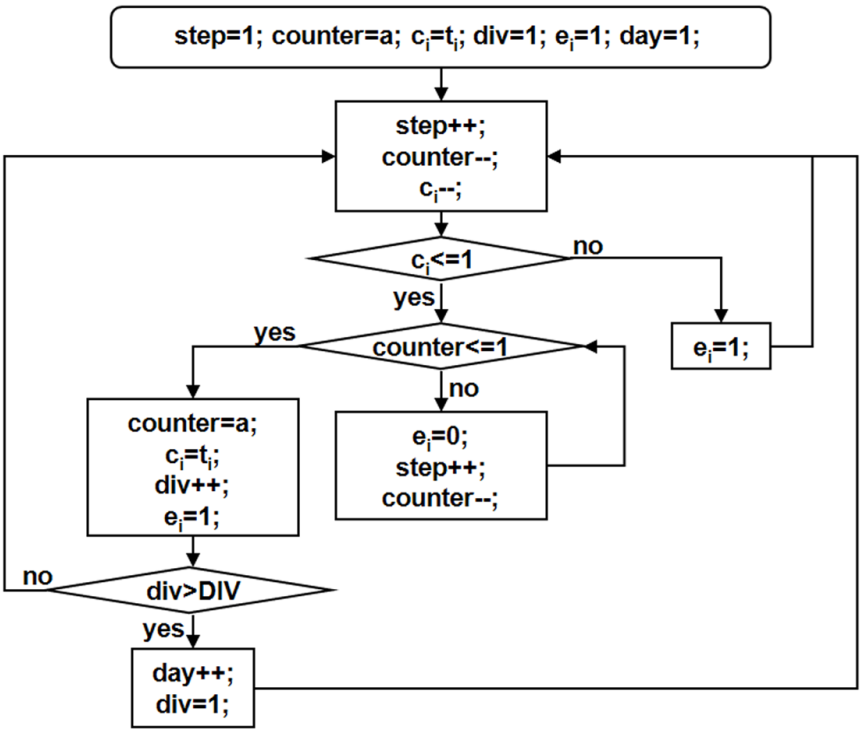


Figure A.3.6. Synchronisation algorithm for coordination of multi-rate developmental and signalling events. The values of "counter" and "c_i" are decremented by 1 after the completion of each L-system time step. When "c_i" is equal to or lower than 1, its corresponding signalling or development event "e_i" will be temporarily stopped until the current time section is finished; otherwise the event "e_i" keeps happening. When "counter" is equal to or lower than 1, a new time section will start and the values of "counter" and "c_i" will be re-initialised. When the number of time sections (counted by variable "div") exceeds the value of DIV, a new day will start and the value of "div" will be reinitialised. The expressions in this flow chart are in C syntax.

A.4 Text S4: Assumptions and conditions for virtual experiments

Since our testing target for this first application case is the cotyledon-root and the cotyledon-shoot hypotheses, all other signalling mechanisms, including signal production, transport, perception and function, play a supplementary role to support the running of the AON system. The unknown details of the other signalling mechanisms (excluding the testing target) could be temporarily assumed and manipulated according to biologists' understanding during virtual experiments but were not tested in this case.

Production of Q

During early stages of soybean nodulation the bacteria still actively produce nod factor, but expanded nodules with mature bacteroids show no stimulation of the NFR1/5 to CCamK cascade needed for nodulation and induction of SDI (Li et al., 2009). Thus, mature nodules have less SDI stimulating activity and we assume the production of Q signal from each nodulation site to be inversely proportional to nodule development stage, strongest at the stage of nodule initialisation but growing weaker as the nodule matures. Since quantitative knowledge about this process is unclear, we used nodule growth potential to represent this inverse relationship:

$$Q_{prdt} = Q_{ini} * (N_{final} - N_{current}) \quad \text{(A.4.1)}$$

where Q_{prdt} is the quantity of Q signal produced by a nodulation site at a certain moment, Q_{ini} is a parameter used to define the relationship with SDI inhibition threshold (see parameter setting below), N_{final} is the final size of this nodule, and $N_{current}$ is the current size of this nodule.

Signal transport

For the signal transport through roots, stem and petioles, multiple possible patterns of its movement from one tissue to the next are supported by our computational models. For example, the transport could be mass-flow and could also be restricted by certain concentration thresholds. In this case, we assumed mass flow as the transport pattern. The transport rates of Q and SDI were controlled respectively by parameters R_q and R_{sdi}.

Perception of Q and production of SDI

The quantitative pattern of Q perception in a leaf is also supported with multiple options. In this application case, we assumed all Q molecules arriving in the leaf could be fully perceived by *GmNARK*. As a consequent event of this perception, the triggered production of SDI at a certain moment was assumed to be proportional to the perceived Q quantity. Since mRNA expression of *GmNARK* apparently is uniform along the leaf vasculature and the vascular content per leaf is proportional to total leaf biomass (Nontachaiyapoom et al., 2007), we used leaf biomass as the coefficient for this proportion relationship:

$$SDI_{prdt} = Q_{pcpt} * B_{leaf} \quad \text{(A.4.2)}$$

Where SDI_{prdt} is the quantity of produced SDI signal at a certain moment, Q_{pcpt} is the perceived quantity right before this event and B_{leaf} represents the biomass of this leaf at this moment. At a certain moment, the relationship between SDI_{prdt} and Q_{pcpt} is proportional. However, due to the continuous change of leaf biomass, this process is actually nonlinear over time. For unifoliate and trifoliate leaves, their biomass keeps increasing with plant development, thus their capability to produce SDI also kept increasing in this series of virtual experiments. However, the cotyledon biomass declines as the plant grows, thus the production of SDI from the cotyledons during these virtual experiments kept being weakened.

Function of SDI

When the SDI signal arrives at a potential nodulation site in the root, a threshold is assumed for determining whether nodule initialisation from this site should be inhibited or not. This threshold is defined as SDI_{ihbt} in virtual experiments. If the quantity of SDI signal around a potential nodulation site is higher than SDI_{ihbt}, the potential nodule will be inhibited; otherwise, the potential nodule will be formed.

Parameter setting

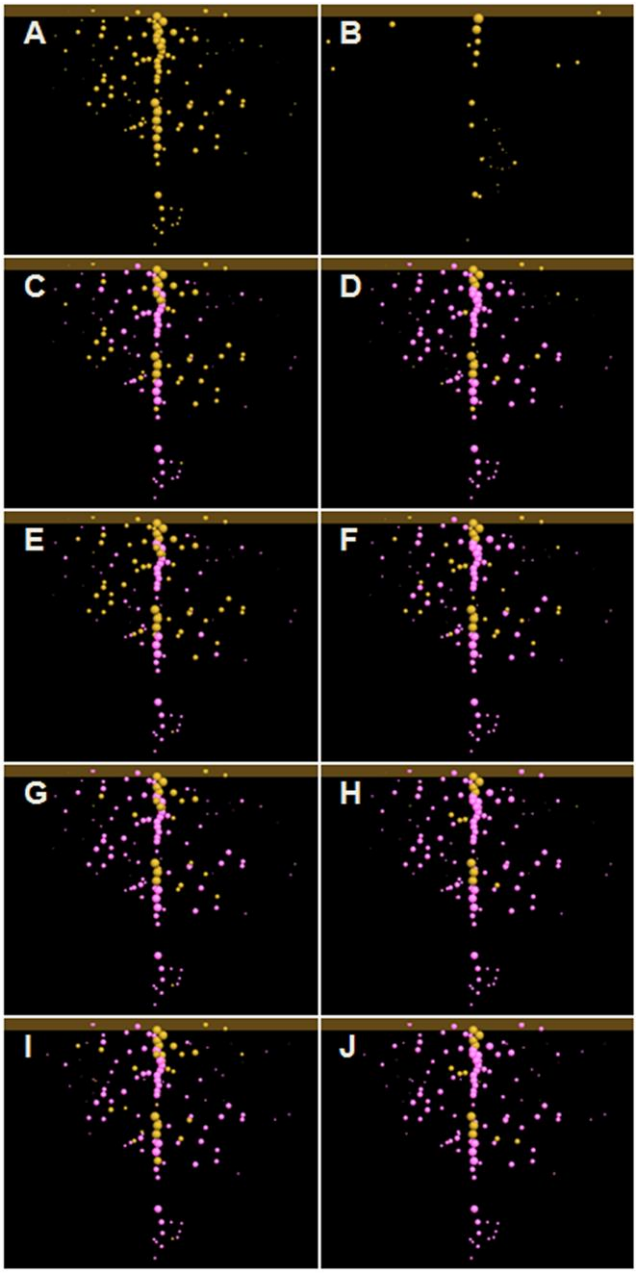
The strategy for this application was to adjust parameters for signal production, transport, perception and function within a physiologically appropriate range, as quantitative details about these mechanisms still remain largely unknown. Three qualitative relationships between Q_{ini} and SDI_{ihbt} – namely higher, equal and lower – were represented by setting Q_{ini} at 1, while varying SDI_{ihbt} through values of 0.5, 1 and 2. For signal transport rates, lab experiments demonstrated that auxin, which might play a role in AON, moves at rate of 60 mm/day in soybean (unpublished data), while

previous studies suggested that it could be transported "through many plant tissues" at a speed of 240-360 mm/day "in the general direction from the plant's apex to its roots" (Mitchison, 1980). Since SDI might also be other signals and transport rate for Q is unclear, we assumed three values for R_q and R_{sdi}: 60 mm/day, 160 mm/day and 360 mm/day. Combinations allowing different relative speeds for R_q and R_{sdi} gave 27 different conditions for cotyledon-root testing experiments CRH_1 to CRH_27 and cotyledon-shoot experiments CSH_1 to CSH_27, as shown in Table A.4.1.

Table A.4.1. Parameter setting for each virtual experiment.

Experiment ID	Cotyledon Hypothesis	Signal Transport Rates (mm/day)		Q_{ini} : SDI_{ihbt}
		Q	SDI	
CRH_1	cotyledon-root	360	360	1
CRH_2	cotyledon-root	360	160	1
CRH_3	cotyledon-root	160	360	1
CRH_4	cotyledon-root	160	160	1
CRH_5	cotyledon-root	360	60	1
CRH_6	cotyledon-root	60	360	1
CRH_7	cotyledon-root	60	60	1
CRH_8	cotyledon-root	160	60	1
CRH_9	cotyledon-root	60	160	1
CRH_10	cotyledon-root	360	360	2
CRH_11	cotyledon-root	360	160	2
CRH_12	cotyledon-root	160	360	2
CRH_13	cotyledon-root	160	160	2
CRH_14	cotyledon-root	360	60	2
CRH_15	cotyledon-root	60	360	2
CRH_16	cotyledon-root	60	60	2
CRH_17	cotyledon-root	160	60	2
CRH_18	cotyledon-root	60	160	2
CRH_19	cotyledon-root	360	360	0.5
CRH_20	cotyledon-root	360	160	0.5
CRH_21	cotyledon-root	160	360	0.5
CRH_22	cotyledon-root	160	160	0.5
CRH_23	cotyledon-root	360	60	0.5
CRH_24	cotyledon-root	60	360	0.5
CRH_25	cotyledon-root	60	60	0.5
CRH_26	cotyledon-root	160	60	0.5
CRH_27	cotyledon-root	60	160	0.5
CSH_1	cotyledon-shoot	360	360	1
CSH_2	cotyledon-shoot	360	160	1
CSH_3	cotyledon-shoot	160	360	1
CSH_4	cotyledon-shoot	160	160	1
CSH_5	cotyledon-shoot	360	60	1
CSH_6	cotyledon-shoot	60	360	1
CSH_7	cotyledon-shoot	60	60	1
CSH_8	cotyledon-shoot	160	60	1
CSH_9	cotyledon-shoot	60	160	1
CSH_10	cotyledon-shoot	360	360	2
CSH_11	cotyledon-shoot	360	160	2
CSH_12	cotyledon-shoot	160	360	2
CSH_13	cotyledon-shoot	160	160	2
CSH_14	cotyledon-shoot	360	60	2
CSH_15	cotyledon-shoot	60	360	2
CSH_16	cotyledon-shoot	60	60	2
CSH_17	cotyledon-shoot	160	60	2
CSH_18	cotyledon-shoot	60	160	2
CSH_19	cotyledon-shoot	360	360	0.5
CSH_20	cotyledon-shoot	360	160	0.5
CSH_21	cotyledon-shoot	160	360	0.5
CSH_22	cotyledon-shoot	160	160	0.5
CSH_23	cotyledon-shoot	360	60	0.5
CSH_24	cotyledon-shoot	60	360	0.5
CSH_25	cotyledon-shoot	60	60	0.5
CSH_26	cotyledon-shoot	160	60	0.5
CSH_27	cotyledon-shoot	60	160	0.5

A.5 Figure S5: Visualisation of nodule distribution with inhibited nodules on the 16th day post-sowing



A.6 Video S6: Sample visualisation of soybean root architecture with nodulation

This video is available in the attached “Video_S6.wmv” file.