THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Functional characterisation of novel peptide hormones in legume nodulation and plant development

April Hannah Hastwell

BAppSc, BSc (Hons)

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2018

School of Agriculture and Food Science

Abstract

Creating agricultural sustainability is a critical step to producing more food, mitigating climate change and protecting the Earth's biodiversity. Legumes are key to this process, owed to their ability to form a beneficial symbiosis with nitrogen-fixing bacteria, resulting in fixed atmospheric nitrogen within specialised plant organs called nodules. However, to maximise the benefit of using legumes for their nutritional and environmental qualities, we require a better understanding of molecular mechanisms governing developmental pathways and, in particular, those that regulate nodulation. Gene family members that encode CLAVATA3/Endosperm Surrounding Region-related (CLE) peptides play vital roles in these developmental processes and are the focus of this thesis.

Few CLE peptide-encoding gene family members had previously been identified in legumes. Using bioinformatics-based approaches, genome-wide searches for candidate genes were conducted in soybean, common bean, *Medicago truncatula*, *Lotus japonicus* and pea. Subsequently, phylogenetic and transcriptomic meta-analyses were performed to elucidate potential roles for these genes in plant development. In total, 275 CLE peptide-encoding genes were identified; this includes orthologues of well-characterised CLE peptides (CLV3, TDIF and CLE40), ten multi-CLE domain encoding genes and new candidates for nodulation-suppressive CLE peptides. The findings from these studies provide a novel bioinformatic resource, which can now be used for further functional characterisation of these genes.

The characterisation of CLE peptides is reliant on genetic tools and synthetic chemistry, as methods to isolate significant quantities of the peptides for biological assays have not yet been achieved. However, CLE peptides are post-translationally modified with an arabinose moiety that cannot be synthesised using traditional and commercially available technologies. To overcome this, a stereoselective method was developed to synthesise a triarabinose 'building block' which could then be used to prepare triarabinosylated CLE peptides. The activity of the modification on a peptide was examined by synthesising GmCLE40a, identified in this thesis and orthologous to AtCLE40, which maintains proliferation and differentiation of the stem cell population in the root apical meristem. Using a novel peptide-feeding assay, the biological activity of GmCLE40a variants was compared. Application of triarabinosylated GmCLE40a suppressed root growth and showed greater root suppressive activity than hydroxylated GmCLE40a, indicating the

requirement for triarabinosylation. This gives further indication that other CLE peptides may also be modified in a similar manner.

Nodule number regulation is mediated by the Autoregulation of Nodulation signalling cascade and begins with the production of the CLE peptides GmRIC1 and GmRIC2; these have previously been predicted to be also triarabinosylated. Using the developed triarabinose building block, analogues of GmRIC1a and GmRIC2a were chemically synthesised to test their biological activity in a range of pea supernodulation mutants, some of which are unavailable in soybean. Using targeted petiole-feeding, only triarabinosylated variants of the synthesised peptides could suppress nodulation; this suppression was SYM29 dependent (encoding the CLE peptide receptor) and partially SYM28 dependent. Nodule number was also significantly reduced in *Psnod3*, which is mutated in an arabinosyltransferase proposed to be responsible for the *in vivo* glycosylation of nodulation-suppressive CLE peptides. This is an important advancement to the Autoregulation of Nodulation pathway.

In summary, this thesis presents the identification of CLE peptide-encoding genes in important model and crop legumes and provides insights into the biological role of individual candidate genes. Further functional characterisation of several CLE peptides was also performed using glycosylated variants and this affirmed that these modifications are required to achieve a biological response by the peptide. Together, using bioinformatic, molecular biology and chemical techniques, the findings from these studies have significantly increased our understanding of CLE-peptide encoding genes and provided numerous potential targets for crop improvement, including nodulation optimisation, fruit size and root morphology.

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, financial support 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 higher degree by research 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 policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis and have sought permission from co-authors for any jointly authored works included in the thesis.

Publications during candidature

Journal publications

Ferguson BJ, Mens C, **Hastwell AH**, Zhang MB, Su H, Jones CH, Chu XT, Gresshoff PM. **2018.** Legume nodulation: the host controls the party. *Plant, Cell & Environment*, https://doi.org/10.1111/pce.13348.

Hastwell AH*, Corcilius L*, Williams J, Gresshoff PM, Payne RJ Ferguson BJ. **2018**. Triarabinosylation is required for nodulation-suppressive CLE peptides to systemically inhibit nodulation in *Pisum sativum*. *Plant, Cell & Environment*, https://doi.org/10.1111/pce.13325. *Joint first author.

Corcilius L*, **Hastwell AH***, Zhang M, Williams J, MacKay JP, Gresshoff PM, Ferguson BJ, Payne RJ. **2017**. Arabinosylation modulates the growth-regulating activity of the peptide hormone CLE40a from soybean. *Cell Chemical Biology*, **24**(11): 1347-1355. *Joint first author.

Hastwell AH, de Bang TC, Gresshoff PM, Ferguson BJ. **2017**. CLE peptide-encoding gene families in *Medicago truncatula* and *Lotus japonicus*, compared with those of soybean, common bean and Arabidopsis. *Scientific Reports* **7**, 9384.

Hastwell AH, Gresshoff PM, Ferguson BJ. **2015**. Genome-wide annotation and characterization of CLAVATA/ESR (CLE) peptide hormones of soybean (*Glycine max*) and common bean (*Phaseolus vulgaris*), and their orthologues of *Arabidopsis thaliana*. *Journal of Experimental Botany*. **66** (17): 5271-5287.

Hastwell AH, Gresshoff PM, Ferguson BJ. **2015.** The structure and activity of nodulation-suppressing CLE peptide hormones of legumes. *Functional Plant Biology* **42**(3): 229-238.

Gresshoff PM, Hayashi S, Biswas B, Mirzaei S, Indrasumunar A, Reid D, Samuel S, Tollenaere A, van Hameren B, **Hastwell AH**, Scott P, Ferguson BJ. **2015**. The value of biodiversity in legume symbiotic nitrogen fixation and nodulation for biofuel and food production. *Journal of Plant Physiology* **172**: 128-136.

Ferguson BJ, Li D, **Hastwell AH**, Reid DE, Li Y, Jackson SA, Gresshoff PM. **2014**. The soybean (*Glycine max*) nodulation-suppressive CLE peptide, GmRIC1, functions

interspecifically in common white bean (*Phaseolus vulgaris*), but not in a supernodulating line mutated in the receptor PvNARK. *Plant Biotechnology Journal* **12**(8): 1085-1097.

Book chapters

Zhang M, Chu X, Su H, **Hastwell AH**, Gresshoff PM, Ferguson BJ. **2018**. Advances in understanding of soybean plant physiology and growth. *In: Achieving Sustainable Cultivation of Soybeans* (. Nguyen HT (ed.) Burleigh Dodds Science Publishing, Sawston, Cambridge, UK, in press.

Conference proceedings

Hastwell AH, Li D, Tollenaere A, Gresshoff PM, Ferguson BJ. **2014**. 'Negative regulation of legume nodules by inducible signal peptides' in the 17th Australian Nitrogen Fixation proceedings, ed. Gupta, VVSR, Unkovich, MJ and Kaiser, BN. ASNF, Adelaide, pp. 44-45.

Conference abstracts

(A) Oral presentations

Hastwell AH, Gresshoff PM, Ferguson BF. **2017**. Regulation of nodule numbers and root growth by CLE peptides in legumes. SUNFix, Sydney, Australia.

Hastwell AH, Corcilius L, Zhang M, Jones C, de Bang T, Payne R, Gresshoff PM, Ferguson BF. **2016**. CLAVATA3/EMBRYO surrounding region-related (CLE) peptides in legumes that control nodulation and root development, ComBio, Brisbane, Australia.

Hastwell AH, Corcilius L, Zhang M, Jones C, de Bang T, Payne R, Gresshoff PM, Ferguson BF. **2016**. The role of post-translational modification in CLAVATA/EMBRYO SURROUNDING REGION-related (CLE) peptides in root development. 4th European Workshop on Peptide Signalling in Plants, Bischoffsheim, France.

Hastwell AH, Li D, Tollenaere A, Jones C, Haaima L, Gresshoff PM, Ferguson BF. **2015**. Identification and functional characterisation of CLE peptides in legumes. JIC Legume Workshop. John Innes Centre, Norwich, United Kingdom.

Hastwell AH, Li D, Tollenaere A, Jones C, Haaima L, Gresshoff PM, Ferguson BF. **2015**. Identification and functional characterisation of CLE peptides in legumes. 3rd European Workshop on Peptide Signalling and Activity in Plants, Ghent, Belgium.

(B) Poster presentations

Hastwell AH, Corcilius L, Zhang M, Jones C, de Bang T, Payne R, Gresshoff PM, Ferguson BF. **2016**. Nodulation suppressive glycosylated CLE peptides in *Glycine max* and *Pisum sativum*. 12th European Nitrogen Fixation Conference, Budapest, Hungary.

Hastwell AH, Corcilius L, Zhang M, Jones C, de Bang T, Payne R, Gresshoff PM, Ferguson BF. **2016**. The role of post-translational modification in CLAVATA/EMBRYO SURROUNDING REGION-related (CLE) peptides in root development. 4th European Workshop on Peptide Signalling in Plants, Bischoffsheim, France.

Hastwell AH, Li D, Tollenaere A, Jones C, Haaima L, Gresshoff PM, Ferguson BF. **2015**. Identification and functional characterisation of CLE peptides in legumes. 3rd European Workshop on Peptide Signalling and Activity in Plants, Ghent, Belgium.

Publications included in this thesis

Hastwell AH, Gresshoff PM, Ferguson BJ. **2015.** The structure and activity of nodulation-suppressing CLE peptide hormones of legumes. *Functional Plant Biology* **42**(3): 229-238.

Incorporated as Chapter 2.

Contributor	Statement of contribution
April H Hastwell (Candidate)	Conception and design (50%) Analysis and interpretation (70%) Drafting and production (60%)
Peter M Gresshoff	Conception and design (20%) Analysis and interpretation (0%) Drafting and production (5%)
Brett J Ferguson	Conception and design (30%) Analysis and interpretation (30%) Drafting and production (35%)

Hastwell AH, Gresshoff PM, Ferguson BJ. **2015**. Genome-wide annotation and characterization of CLAVATA/ESR (CLE) peptide hormones of soybean (*Glycine max*) and common bean (*Phaseolus vulgaris*), and their orthologues of *Arabidopsis thaliana*. *Journal of Experimental Botany*. **66** (17): 5271-5287.

Incorporated as Chapter 3

Contributor	Statement of contribution
April H Hastwell (Candidate)	Conception and design (60%) Analysis and interpretation (80%) Drafting and production (60%)
Peter M Gresshoff	Conception and design (10%) Analysis and interpretation (0%) Drafting and production (10%)
Brett J Ferguson	Conception and design (30%) Analysis and interpretation (20%) Drafting and production (30%)

Hastwell AH, de Bang TC, Gresshoff PM, Ferguson BJ. **2017**. CLE peptide-encoding gene families in *Medicago truncatula* and *Lotus japonicus*, compared with those of soybean, common bean and Arabidopsis. *Scientific Reports* **7**, 9384.

Incorporated as Chapter 4

Contributor	Statement of contribution
April H Hastwell (Candidate)	Conception and design (70%) Analysis and interpretation (70%) Drafting and production (60%)
Thomas Christian de Bang	Conception and design (5%) Analysis and interpretation (10%) Drafting and production (5%)
Peter M Gresshoff	Conception and design (5%) Analysis and interpretation (0%) Drafting and production (5%)
Brett J Ferguson	Conception and design (20%)

	Analysis and interpretation (20%) Drafting and production (30%)

Corcilius L*, **Hastwell AH***, Zhang M, Williams J, MacKay JP, Gresshoff PM, Ferguson BJ, Payne RJ. **2017**. Arabinosylation modulates the growth-regulating activity of the peptide hormone CLE40a from soybean. *Cell Chemical Biology*, **24**(11): 1347-1355. *Joint first author.

Incorporated as Chapter 5

Contributor	Statement of contribution
Leo Corcilius	Conception and design (0%) Analysis and interpretation (30%) Drafting and production (32.5%)
April H Hastwell (Candidate)	Conception and design (25%) Analysis and interpretation (40%) Drafting and production (32.5%)
Mengbai Zhang, James Williams, Joel P MacKay	Conception and design (5%) Analysis and interpretation (5%) Drafting and production (2%)
Peter M Gresshoff	Conception and design (10%) Analysis and interpretation (5%) Drafting and production (3%)
Brett J Ferguson	Conception and design (40%) Analysis and interpretation (5%) Drafting and production (15%)
Richard J Payne	Conception and design (20%) Analysis and interpretation (5%) Drafting and production (15%)

Hastwell AH*, Corcilius L*, Williams J, Gresshoff PM, Payne RJ Ferguson BJ. **2018**. Triarabinosylation is required for nodulation-suppressive CLE peptides to systemically inhibit nodulation in *Pisum sativum*. *Plant, Cell & Environment*, https://doi.org/10.1111/pce.13325. *Joint first author.

Incorporated as Chapter 6

Contributor	Statement of contribution
April H Hastwell (Candidate)	Conception and design (30%) Analysis and interpretation (45%) Drafting and production (35%)
Leo Corcilius	Conception and design (0%) Analysis and interpretation (15%) Drafting and production (10%)
Peter M Gresshoff	Conception and design (10%) Analysis and interpretation (5%) Drafting and production (5%)
Richard J Payne	Conception and design (20%) Analysis and interpretation (5%) Drafting and production (20%)
Brett J Ferguson	Conception and design (40%) Analysis and interpretation (30%) Drafting and production (30%)

Ferguson BJ, Li D, **Hastwell AH**, Reid DE, Li Y, Jackson SA, Gresshoff PM. **2014**. The soybean (*Glycine max*) nodulation-suppressive CLE peptide, GmRIC1, functions interspecifically in common white bean (*Phaseolus vulgaris*), but not in a supernodulating line mutated in the receptor PvNARK. *Plant Biotechnology Journal* **12**(8): 1085-1097.

Incorporated as Chapter 7

Contributor	Statement of contribution
Brett J Ferguson	Conception and design (70%) Analysis and interpretation (40%) Drafting and production (70%)
Dongxue Li	Conception and design (0%) Analysis and interpretation (30%) Drafting and production (0%)
April H Hastwell (Candidate)	Conception and design (15%)

	Analysis and interpretation (30%) Drafting and production (20%)
Dugald E Reid	Conception and design (5%) Analysis and interpretation (0%) Drafting and production (0%)
Yupeng Li, Scott A Jackson	Conception and design (5%) Analysis and interpretation (0%) Drafting and production (0%)
Peter M Gresshoff	Conception and design (5%) Analysis and interpretation (0%) Drafting and production (10%)

Contributions by others to the thesis

Advisors, Dr Brett J Ferguson and Prof. Peter M Gresshoff assisted in experimental design and editing of the writing in all chapters.

Professor Richard Payne, Dr Leo Corcilius, James Williams contributed to the experimental design, peptide synthesis and writing of the associated sections in Chapters 5 and 6. Also in Chapter 5, Mengbai Zhang constructed the promoter::GUS fusion and Joel MacKay performed NMR analyses.

Statement of parts of the thesis submitted to qualify for the award of another degree

None

Research involving human or animal subjects

No animal or human participants were involved in this research.

Acknowledgements

Firstly, I would like to give my sincerest gratitude to my advisors, Dr Brett Ferguson and Professor Peter Gresshoff from the Centre for Integrative Legume Research (CILR). From the beginning they have both been supportive and understanding, while being exceptional role models. I am grateful for the immense wisdom and knowledge passed to me by Professor Gresshoff, his attention to detail is insurmountable but should be strived for by everyone. I am indebted to Dr Brett Ferguson for the time and energy spent on every aspect of my PhD at the CILR. My knowledge of legumes, nodulation and everything Canadian is owed to him.

I would also like to thank the past and present members of the CILR for their assistance, knowledge and friendship. The environment that the members of CILR create is enjoyable and collaborative, which in turn made my PhD less stressful than I think it should have been. In particular, I would like to thank Dongxue Li for her assistance in the lab; Dr Satomi Hayashi for her guidance; and Candice Jones, for being an excellent officemate.

Other people who have helped me along the way and require many thanks: Professor Peter Dart, gave me great presentation advice and helped me get to my first conference; my PhD committee; Professors Elizabeth Aitken and Peer Schenk gave me reassurance that I was going to finish my PhD; and Dr Annaliese Mason told me that there are only so many productive hours in a day and not to stress over people working more hours than me.

I give my greatest thanks to Christine, whose support and sacrifice was pivotal to this whole experiment that has been my PhD.

Finally, I would like to thank all the people in my life up until this very day. This is an accumulation of people and events, the good and the bad, and had any one small thing been different, I wouldn’t be where I am today.

Financial support

This scholarship for this PhD was gratefully supported by the Molly Budtz-Olsen Fellowship from the Fellowship Fund Inc. Travel support was also thankfully supported by the Australian Society of Nitrogen Fixation travel bursary, the AW Howard Memorial Trust Study Award, the Plant Nutrition Trust Travel Scholarship, the School of Agriculture and Food Science Research Higher Degree Travel Award and the Pulsford Student Travel Grant.

The work conducted at the Centre for Integrative Legume Research was funded by the Herman Slade Foundation and the Australian Research Council Discovery Project grants (DP130103084 and DP130102266).

Keywords

CLE peptide, legume, glycopeptide, nodulation, peptide signalling, plant development, root growth

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 060703, Plant Developmental and Reproductive Biology, 50%
ANZSRC code: 060702, Plant Cell and Molecular Biology, 30%
ANZSRC code: 060705, Plant Physiology, 20%

Fields of Research (FoR) Classification

FoR code: 0607, Plant Biology, 100%

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AON	Autoregulation of Nodulation
Ara*f*$_3$Hyp	β1,2-linked tri-L-arabinofuranosyl
CLE	CLAVATA/ESR-related
CLE-RS	CLE-Root Signal
CLEL	CLE-Like
CLV	CLAVATA
ESR	Endosperm Surrounding Region
Fmoc-SPPS	Fmoc-strategy solid-phase peptide synthesis
GLV	GOLVEN
Hyp	Hydroxyproline
HAR1	Hypernodulation Aberrant Root Formation 1
HMM	Hidden Markov Models
IAD	Intramolecular aglycone delivery
KAPP	Kinase-associated protein phosphatases
LRR	Leucine-rich repeat
MYA	Million Years Ago
NARK	Nodulation Autoregulation Receptor Kinase
NF	Nod Factors
NIC	Nitrogen Induced CLE
Nod	Nodulation
Nts	Nitrate-tolerant symbiosis
OAC	Ontario Agricultural College
RAM	Root Apical Meristem
RDI	Root-derived Inhibitor

RIC	Rhizobia Induced CLE
RGF	Root Growth Factor
RK	Receptor Kinase
RLP1	Receptor-like protein-1
SAM	Shoot Apical Meristem
SDI	Shoot-derived Inhibitor
SUNN	Super Numeric Nodules
SYM	Symbiosis
TDIF	Treachery Element Inhibitory Factor
TrNARK	Truncated NARK
UBC9	Ubiquitin-conjugating enzyme E2
UFD1a	Ubiquitin Fusion Degradation protein 1a
YMB	Yeast Mannitol Broth

Chapter 1

General introduction: CLE peptides roles in plant development

Preface

CLE peptides are signalling molecules involved in numerous plant developmental processes. This general introduction describes the current knowledge on CLE peptides in nodulation and other plant signalling pathways. A further in-depth literature review on nodulation-suppressive CLE peptides is incorporated in this thesis in Chapter 2.

1.1 Legumes and nodulation

1.1.1 Nitrogen and plant development

Nitrogen is a limiting element for plant growth. It is required for nucleic acid, protein and chlorophyll production. However, the most abundant form of nitrogen (atmospheric) is unavailable for use by most organisms as it exists as an unreactive, triple bonded gas, which, to be of use, the nitrogen bond must be broken through fixation or reduction to form a reactive ion (Galloway and Cowling, 2002).

Since the early 1900's when ammonium was first synthesised from unreactive atmospheric N_2 gas, nitrogen-based fertilisers have been added to soil to increase the growth and yield of crop plants (Erisman et al. 2008). While this approach has been highly successful and has assisted in feeding the growing population, the adverse economical and ecological consequences are beginning to outweigh the benefits (Sutton et al. 2011b; Vicente and Dean 2017). Indeed, Sutton et al. (2011a) identified five major effects of the addition of reactive nitrogen-based fertilisers to soils: 1. Eutrophication and acidification of fresh water; 2. Increased NO_x emissions into the air, damaging the ozone and increasing incidence of respiratory disease states; 3. Climate change caused by reactive nitrogen; 4. Soil acidification; and 5. Damage to biodiversity and ecosystems caused by acidification. As a result of these negative consequences, alternative nitrogen inputs are required.

1.1.2 Biological nitrogen fixation

One such alternative nitrogen input is biological nitrogen fixation. Currently, it is responsible for fixing 50–70 Tg of nitrogen annually in agricultural systems (Herridge et al. 2008, Jensen et al. 2012).

Plant species of the family Fabaceae, commonly known as legumes, have the ability to enter into a symbiotic relationship with soil bacteria, collectively called rhizobia. Commonly grown legume crop varieties include soybean, pea, common bean, chickpea, lentil and peanut. Together, legumes and rhizobia provide a source of biological nitrogen fixation. The rhizobia use a specialised nitrogenase enzyme complex to "fix" atmospheric nitrogen and the plant host form an environment suitable for this nitrogen-fixation in a new organ that develops on the roots, called a nodule. The host plant is then provided with useable

forms of nitrogen (e.g. ammonium) and in return the rhizobia are provided with a carbon source derived from photosynthesis, predominately malate (Udvardi et al. 1988). In addition to increasing current crop yields, this process can be exploited in agriculture to increase the nitrogen content of the soil for future crops by using legumes as rotation crops (Gresshoff et al. 2014). Current efforts to commandeer this process include developing nitrogen-fixing cereals (Beatty and Good 2011), engineering the nitrogenase enzyme directly into alternative prokaryotes and eukaryotes (Yang et al. 2014) and more simply, optimising nitrogen-fixation and nodulation in legumes directly (Considine et al. 2017). Each approach requires complex cascades of molecular signalling pathways to occur before nitrogen-fixation can progress (Remigi et al. 2016; Vicente and Dean 2017). However, the biggest limitation is gaps in the current understanding of molecular processes controlling the symbiosis and increasing the knowledge of such pathways is critical to developing alternative and superior nitrogen-fixing crops.

1.1.3 Nodule development

The process of nitrogen fixation occurs in specialised organs called nodules that are located on the legume root. The development of the nodules is dependent on a number of internal and external signals (Ferguson and Mathesius 2003, 2014; Ferguson et al. 2010; Suzaki et al. 2015). Nodule morphology, however, is determined by the host plant. Indeterminate nodules, those formed by plants such as pea and alfalfa have a persistent meristem and are cylindrical. Determinate nodules in contrast lack a persistent meristem and form spherical nodules, such as those nodules formed by soybean and common bean (Newcomb et al. 1979; Ferguson et al. 2010).

The symbiosis is initiated by the release of flavonoids by the host plant into the rhizoshpere (Gibson and Harper, 1985; Redmond et al. 1986). Flavonoids attract compatible rhizobia to the sites of emerging root hairs where infection can occur (Bhuvaneswari et al. 1980; Gulash et al. 1984). Upon perception of flavonoids, rhizobia *nod* genes are expressed that synthesise nod factors (NF), which are signal molecules that are secreted and perceived by two plant LysM receptor-like kinases located on the epidermis and plasma membrane known as *GmNFR1* and *GmNFR5* in soybean; *LjNFR1 and LjNFR5* in *Lotus japonicus; and MtNFP and MtLYK3* in *Medicago truncatula* (Figure 1.1; Limpens et al. 2003; Madsen et al. 2003; Radutoiu et al. 2003; Smit et al. 2007; Indrasumunar et al. 2010; Indrasumunar et al. 2011). This triggers a downstream

signalling cascade to induce nodule development in the plant (Ferguson et al. 2010) and causes root hairs to deform and curl, encapsulating dividing bacteria (Yao and Vincent, 1969; Timmers et al. 1999).

Within the curling root hair, the increasing concentration of NF, along with enzymes for cell wall degradation trigger the formation of an infection thread, a structure allowing bacteria entry into the cortical cells of the plant. Cell divisions within the cortical cells then lead to nodule formation. Upon de-differentiation of the bacteria into bacteroids in the nodule, rhizobia nitrogenase activity begins to fix nitrogen for the plant. The complex molecular signalling pathway for nodule initiation is described in Figure 1.1 (Ferguson et al. 2010).

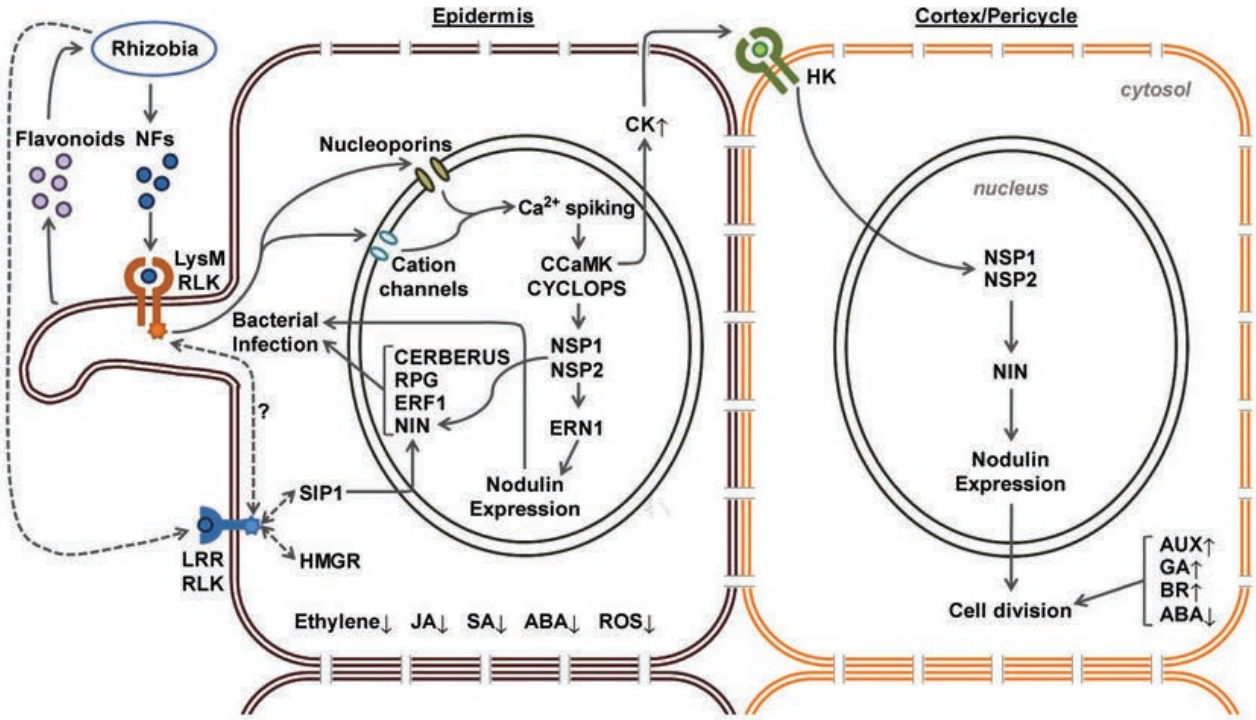


Figure 1.1 Nodulation molecular signalling cascade within the root taken from Ferguson et al. (2010). Flavanoid perception from the plant triggers rhizobia within the rhizoshpere to release NF, which is perceived by a leucine-rich repeat receptor like kinases. This leads to the downstream expression of Nodulin within the nucleus and the release of the plant hormone, cytokinin which initiates signalling in the cortex to induce cell divisions. Other plant hormones are also regulators of nodulation.

1.1.4 Autoregulation of nodulation

Nodule numbers are dependent on a number of internal and external signals and are controlled both locally and systemically by the plant (Reid et al. 2011b). Local controls, such as high soil nitrate levels directly prevent nodules from developing. The systemic

control of nodule formation acts v ia a process called Autoregulation Of Nodulation (AON; Kosslak and Bohlool 1984; Delves et al. 1986; Reid et al. 2011b). AON is closely associated with the nitrate regulatory pathway, but is exerted in the shoot, rather than locally in the root (Delves et al. 1986, Olsson et al. 1989).

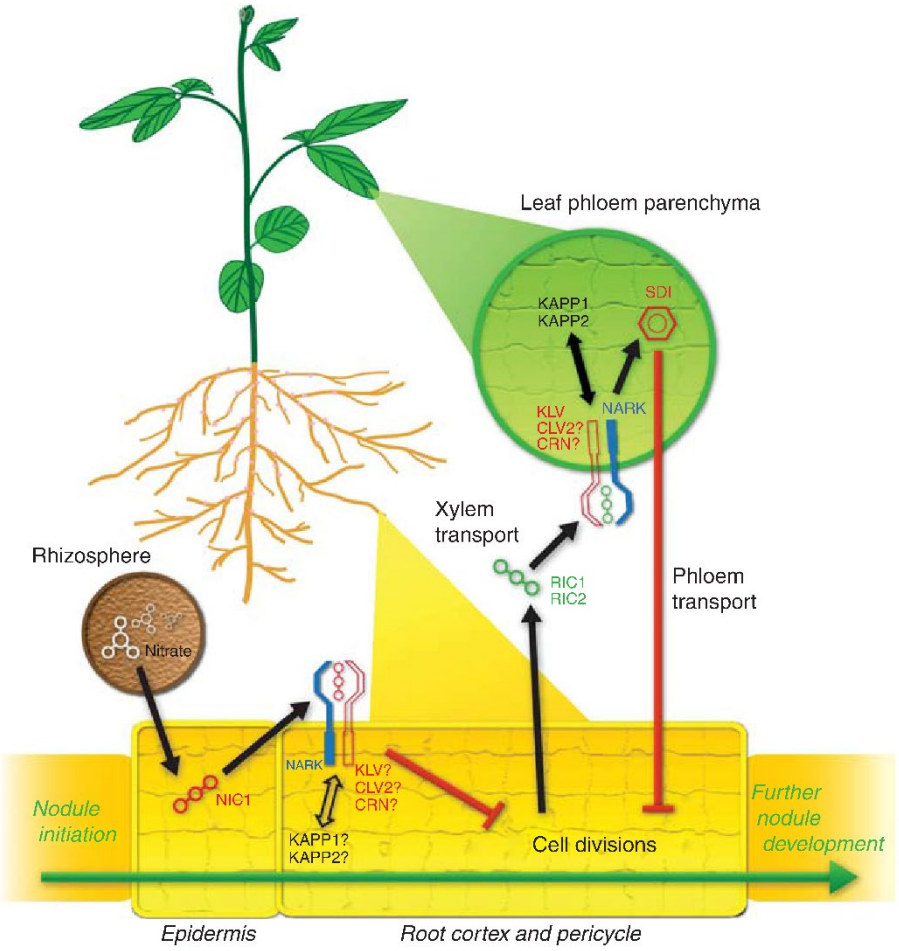


Figure 1.2 Nodulation molecular signalling cascade within the root taken from Ferguson et al. (2010). Perception of plant flavonoids by rhizobia triggers the release of Nod Factors, which then bind to plant LysM receptor like kinases. Subsequent calcium spiking then leads to the downstream expression of a cascade of nodulation genes within the nucleus eventually initiating cortical cell divisions. Plant hormones related to stress responses are reduced in the epidermis and cortex, while cytokinins promote cell division and thus, nodulation

AON begins with the production of a root-derived signal, which is a CLAVATA/ESR Related (CLE) peptide hormone (Figure 1.2; Okamoto et al, 2009; Mortier et al. 2010 Reid et al. 2011a; Ferguson et al. 2014; Hastwell et al. 2015b). The CLE signal is likely produced in the cortical cells in response to NF perception and initial cell divisions and is transported, via the xylem, to the leaf where it is perceived by Nodulation Autoregulation Receptor Kinase (GmNARK) in a heterodimer receptor complex with other factors

(Okamoto et al. 2013). *Phaseolus vulgaris* NARK, *Glycine soja* NARK; Lotus *japonicus* HAR1, *Medicago truncatula* SUNN, and *Pisum sativum* Sym29 are orthologous receptor kinases in other legumes (Krusell et al. 2002; Nishimura et al. 2002; Searle et al. 2003; Schnabel et al. 2005; Miyazawa et al. 2010; Krusell et al. 2011; Ferguson et al. 2014). Plants with mutations in *NARK* lack AON control and exhibit both a nitrate-tolerant symbiosis (*nts*) and a supernodulation phenotype (Day et al. 1987, Carroll et al. 1985). Subsequently, a downstream signal is produced and transported back to the roots via the phloem to inhibit further nodulation (Figure 1.2; Delves et al. 1986; Delves et al. 1992; Krusell et al. 2002; Searle et al. 2003; Lin et al. 2010). The NARK-dependent signal is unknown (Lin et al. 2010), but the most recent findings indicate a likely miRNA as the root-to-shoot signal (Tsikou et al. 2016; Zhang, Gresshoff and Ferguson, unpublished). Overall, the mechanism for nodulation control appears to be similar in most legumes, as has been shown through grafting studies (Lohar et al. 2005) and heterologous over-expression studies (Osipova et al. 2012; Ferguson et al. 2014).

1.2 CLE peptides in other signalling pathways

In addition to nodulation, CLE peptides have been found to play important roles in other plant developmental processes. The most widely described in the literature is the CLAVATA signalling pathway in the shoot apical meristem (Barton 2010; Sparks et al. 2013). Parallel meristematic control has also been shown in the root apical meristem and additionally in vascular development (Sparks et al. 2013; Furuta et al. 2014). Interestingly, in most CLE peptide pathways, the mechanistic action and downstream responses are only determined by the tissue in which the gene is expressed (Sparks et al. 2013).

1.2.1 Shoot apical meristem

CLE peptides identified in the Arabidopsis CLAVATA signalling pathway regulate shoot apical meristem function (reviewed in Barton 2010). The CLE peptide, AtCLV3, is the ligand for the receptor-like protein kinase AtCLV1 (Trotochaud et al. 1999; Ogawa et al. 2008) and acts in conjunction with the transcription factor, AtWUSCHEL (WUS) to maintain stem cell homeostasis in the shoot apical meristem (Figure 1.3A; Leyser and Furner 1992; Kondo et al. 2011; Barton 2010; Sparks et al. 2013). A non-functional mutation in CLV3 or CLV1 results in an enlarged shoot apical meristem and subsequent

fruit size in the case of tomato caused by over proliferation of the stem cells (Clark et al. 1993; Clark et al. 1995; Xu et al. 2015). The mature CLV3 peptide is 13 residues with three arabinose moieties attached to the hydroxyproline at residue seven (Ohyama et al. 2009). Interestingly, mutations in the arabinosyltransferase, *SlFIN*, also causes an increase in fruit size (Xu et al. 2015). The *L. japonicus* nodulation-suppressive CLE peptides are post–translationally modified in this exact manner (Okamoto et al. 2013).

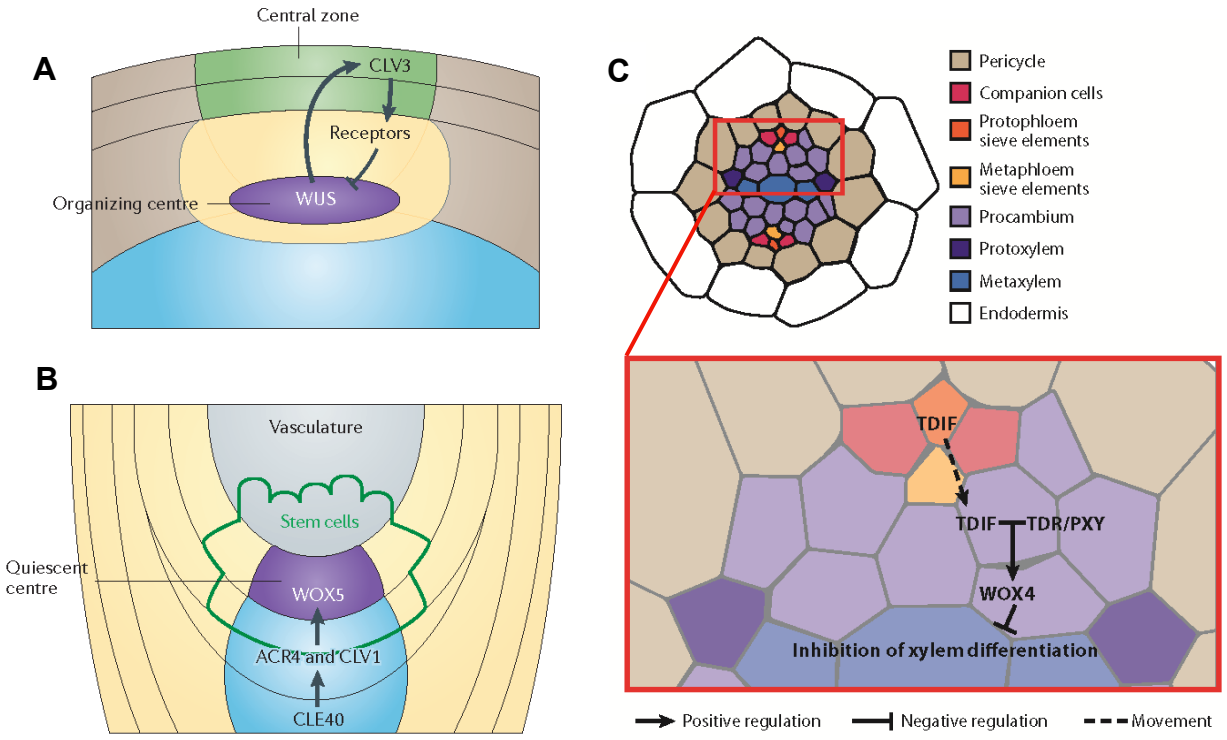


Figure 1.3 Schematic overview of Arabidopsis CLE peptide signalling **A** CLV3 signalling in the shoot apical meristem. CLV3 operates in a feedback loop with the receptors CLV1, CLV2, CORYNE (and other factors) and WUS (adapted from Sparks et al. 2013). **B** CLE40 signalling in the root apical meristem (adapted from Sparks et al. 2013). **C** Root cross-section and TDIF signalling pathway in plant vasculature (adapted from Furuta et al 2014).

1.2.2 Root apical meristem

In the root apical meristem, there is a highly similar mechanism for controlling cell proliferation and differentiation to that in the shoot (Figure 1.3B). Indeed, in Arabidopsis, CLE40 mirrors the structure and function of CLV3 and interacts with the receptor kinases CLV1 and ACR4 (Stahl et al. 2013). Excess or reduced levels of endogenous CLE40 triggered by overexpression or mutation reduces root growth (Hobe et al. 2003), as can the exogenous application of arabinosylated CLE40 (Corcilius et al. 2017).

A homologue of WUS, WUSHEL-RELATED HOMEOBOX 4 (WOX5), also maintains the stem cell population in the root (Sarker et al. 2007). Interestingly, WOX5 has also been shown to have a role in nodulation (Osipova et al. 2011). In contrast to CLAVATA signalling in the shoot apical meristem, signal transduction in the root only appears between adjacent cells and not the entire stem cell population (Sarker et al. 2007; Stahl et al. 2009; Stahl et al. 2013).

1.2.3 Vasculature

Signalling in the vasculature is also mediated by CLE peptides known as Tracheary Element Inhibitory Differentiation Factors (TDIF), AtCLE41, AtCLE42 and AtCLE44 in Arabidopsis (Figure 1.3C; Ito et al. 2006; Etchells et al. 2016). Orthologoes of these peptides appear to be the most highly conserved of all the CLE genes (Strabala et al. 2014; Hastwell et al. 2015a). Other known molecular components acting in vascular differentiation are typically homeologous to those found in the CLAVATA signalling pathway (Figure 1.3C; Furuta et al 2014). This includes the leucine rich repeat receptor kinase TDIF receptor/PHLOEM INTERCALATED WITH XYLEM (TDR/PXY) (Fisher and Turner 2007; Hirakawa et al. 2008) and the homeobox transcription factor, WOX4 (Hirakawa et al. 2010). Mutations disrupting correct vasculature development during this signalling pathway increases biomass, another critical agronomic trait of particular interest in biofuel production (Etchells et al. 2015; Etchells et al. 2016)

1.3 Significance

CLE peptides play diverse roles in plant development, including aspects that are agronomically important and can be targeted in molecular and traditional improvement programs. Manipulation of these processes first requires an increased understanding of the structure and function of CLE peptides and their associated pathways, which is the aim of this thesis.

This thesis identifies and bioinformatically characterises CLE peptides in several legume species in Chapters 3 and 4. Findings from this research can be used as a tool to identify novel functions for CLE peptide candidates (Hastwell et al. 2015a and 2017). Indeed, those techniques and results were employed to identify key CLE peptide candidates for studies conducted in Chapters 5, 6 and 7 (Ferguson et al. 2014; Corcilius et al. 2017;

Hastwell et al. 2017). Taken together, these findings increase our understanding of CLE peptides and their associated pathways by proving novel findings and insights into the structure-function relationship of these small signalling molecules.

1.4 References

Barton M. 2010. Twenty years on: the inner workings of the shoot apical meristem, a developmental dynamo. *Developmental Biology* **341**(1): 95-113.

Beatty PH, Good AG. 2011. Future prospects for cereals that fix nitrogen. *Science* **333**(6041): 416-417.

Considine MJ, Siddique KH, Foyer CH. 2017. Nature's pulse power: legumes, food security and climate change. *Journal of Experimental Botany* **68**(8): 1815-1818.

Bhuvaneswari T, Turgeon BG, Bauer WD. 1980. Early events in the infection of soybean (*Glycine max* L. Merr) by *Rhizobium japonicum* I. Localization of infectible root cells. *Plant Physiology* **66**(6): 1027-1031.

Clark SE, Running MP, Meyerowitz EM. 1993. CLAVATA1, a regulator of meristem and flower development in Arabidopsis. *Development* **119**(2): 397-418.

Clark SE, Running MP, Meyerowitz EM. 1995. CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. *Development* **121**(7): 2057-2067.

Corcilius L, Hastwell AH, Zhang M, Williams J, Mackay JP, Gresshoff PM, Ferguson BJ, Payne RJ. 2017. Arabinosylation modulates the growth-regulating activity of the peptide hormone CLE40a from soybean. *Cell Chemical Biology* **24**(11): 1347-1355.

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**(2): 588-590.

Erisman JW, Sutton MA, Galloway J, Klimont Z, Winiwarter W. 2008. How a century of ammonia synthesis changed the world. *Nature Geoscience* **1**(10): 636-639.

Etchells JP, Mishra LS, Kumar M, Campbell L, Turner SR. 2015. Wood formation in trees is increased by manipulating PXY-regulated cell division. *Current Biology* **25**(8): 1050-1055.

Etchells JP, Smit ME, Gaudinier A, Williams CJ, Brady SM. 2016. A brief history of the TDIF-PXY signalling module: balancing meristem identity and differentiation during vascular development. *New Phytologist* **209**(2): 474-484.

Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE, Gresshoff PM. 2010. Molecular analysis of legume nodule development and autoregulation. *Journal of Integrative Plant Biology* **52**(1): 61-76.

Ferguson BJ, Li D, Hastwell AH, Reid DE, Li Y, Jackson SA, Gresshoff PM. 2014. The soybean (*Glycine max*) nodulation-suppressive CLE peptide, GmRIC1, functions interspecifically in common white bean (*Phaseolus vulgaris*), but not in a supernodulating line mutated in the receptor PvNARK. *Plant Biotechnology Journal* **12**(8): 1085-1097.

Ferguson BJ, Mathesius U. 2014. Phytohormone regulation of legume-rhizobia interactions. *Journal of Chemical Ecology* **40**(7): 770-790.

Ferguson BJ, Mathesius U. 2003. Signaling interactions during nodule development. *Journal of Plant Growth Regulation* **22**(1): 47-72.

Fisher K, Turner S. 2007. PXY, a receptor-like kinase essential for maintaining polarity during plant vascular-tissue development. *Current Biology* **17**(12): 1061-1066.

Furuta KM, Hellmann E, Helariutta Y. 2014. Molecular control of cell specification and cell differentiation during procambial development. *Annual Review of Plant Biology* **65**: 607-638.

Galloway JN, Cowling EB. 2002. Reactive nitrogen and the world: 200 years of change. *AMBIO: A Journal of the Human Environment* **31**(2): 64-71.

Gibson A, Harper J. 1985. Nitrate effect on nodulation of soybean by *Bradyrhizobium japonicum*. C*rop Science* **25**(3): 497-501.

Gresshoff PM, Hayashi S, Biswas B, Mirzaei S, Indrasumunar A, Reid D, Samuel S, Tollenaere A, van Hameren B, Hastwell AH, Scott P, Ferguson BJ 2015. The value of biodiversity in legume symbiotic nitrogen fixation and nodulation for biofuel and food production. *Journal of Plant Physiology* **172**: 128-136.

Gulash M, Ames P, Larosiliere R, Bergman K. 1984. Rhizobia are attracted to localized sites on legume roots. *Applied and Environmental Microbiology* **48**(1): 149-152.

Hastwell AH, de Bang TC, Gresshoff PM, Ferguson BJ. 2017. CLE peptide-encoding gene families in *Medicago truncatula* and *Lotus japonicus*, compared with those of soybean, common bean and Arabidopsis. *Scientific Reports* **7**, 9384.

Hastwell AH, Gresshoff PM, Ferguson BJ. 2015a. Genome-wide annotation and characterization of CLAVATA/ESR (CLE) peptide hormones of soybean (*Glycine max*) and common bean (*Phaseolus vulgaris*), and their orthologues of *Arabidopsis thaliana*. *Journal of Experimental Botany* **66**(17): 5271-5287.

Hastwell AH, Gresshoff PM, Ferguson BJ. 2015b. The structure and activity of nodulation-suppressing CLE peptide hormones of legumes. *Functional Plant Biology* **42**(3): 229-238.

Herridge DF, Peoples MB, Boddey RM. 2008. Global inputs of biological nitrogen fixation in agricultural systems. *Plant and Soil* **311**(1-2): 1-18.

Hirakawa Y, Kondo Y, Fukuda H. 2010. TDIF peptide signaling regulates vascular stem cell proliferation via the *WOX4* homeobox gene in Arabidopsis. *The Plant Cell* **22**(8): 2618-2629.

Hirakawa Y, Shinohara H, Kondo Y, Inoue A, Nakanomyo I, Ogawa M, Sawa S, Ohashi-Ito K, Matsubayashi Y, Fukuda H. 2008. Non-cell-autonomous control of vascular stem cell fate by a CLE peptide/receptor system. *Proceedings of the National Academy of Sciences* **105**(39): 15208-15213.

Hobe M, Müller R, Grünewald M, Brand U, Simon R. 2003. Loss of CLE40, a protein functionally equivalent to the stem cell restricting signal CLV3, enhances root waving in Arabidopsis. *Development Genes and Evolution* **213**(8): 371-381.

Indrasumunar A, Kereszt A, Searle I, Miyagi M, Li D, Nguyen CD, Men A, Carroll BJ, Gresshoff PM. 2010. Inactivation of duplicated *nod factor receptor 5* (*NFR5*) genes in recessive loss-of-function non-nodulation mutants of allotetraploid soybean (*Glycine max* L. Merr.). *Plant and Cell Physiology* **51**(2): 201-214.

Indrasumunar A, Searle I, Lin M, Kereszt A, Men A, Carroll B, Gresshoff PM. 2011. Limitation of nodule organ number by a nodulation factor receptor kinase in soybean (*Glycine max* L. Merr.). *Plant Journal* **65**: 39-50.

Ito Y, Nakanomyo I, Motose H, Iwamoto K, Sawa S, Dohmae N, Fukuda H. 2006. Dodeca-CLE peptides as suppressors of plant stem cell differentiation. *Science* **313**(5788): 842-845.

Jensen ES, Peoples MB, Boddey RM, Gresshoff PM, Hauggaard-Nielsen H, Alves BJ, Morrison MJ. 2012. Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. A review. *Agronomy for Sustainable Development* **32**(2): 329-364.

Kondo T, Yokomine K, Nakagawa A, Sakagami Y. 2011. Analogs of the CLV3 peptide: synthesis and structure–activity relationships focused on proline residues. *Plant and Cell Physiology* **52**(1): 30-36.

Kosslak RM, Bohlool BB. 1984. Suppression of nodule development of one side of a split-root system of soybeans caused by prior inoculation of the other side. *Plant Physiology* **75**(1): 125-130.

Krusell L, Madsen LH, Sato S, Aubert G, Genua A, Szczyglowski K, Duc G, Kaneko T, Tabata S, de Bruijn F. 2002. Shoot control of root development and nodulation is mediated by a receptor-like kinase. *Nature* **420**(6914): 422-426.

Leyser HO, Furner I. 1992. Characterisation of three shoot apical meristem mutants of *Arabidopsis thaliana*. *Development* **116**(2): 397-403.

Limpens E, Franken C, Smit P, Willemse J, Bisseling T, Geurts R. 2003. LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. *Science* **302**(5645): 630-633.

Physiology **52**(9): 1613-1627.

Lin YH, 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**(4): 1074-1086.

Lohar DP, VandenBosch KA. 2005. Grafting between model legumes demonstrates roles for roots and shoots in determining nodule type and host/rhizobia specificity. *Journal of Experimental Botany* **56**(416): 1643-1650.

Madsen EB, Madsen LH, Radutoiu S, Olbryt M, Rakwalska M, Szczyglowski K, Sato S, Kaneko T, Tabata S, Sandal N. 2003. A receptor kinase gene of the LysM type is involved in legumeperception of rhizobial signals. *Nature* **425**(6958): 637-640.

Mortier V, Den Herder G, Whitford R, Van de Velde W, Rombauts S, D'haeseleer K, Holsters M, Goormachtig S. 2010. CLE peptides control *Medicago truncatula* nodulation locally and systemically. *Plant Physiology* **153**(1): 222-237.

Newcomb W, Sippell D, Peterson R. 1979. The early morphogenesis of *Glycine max* and *Pisum sativum* root nodules. *Canadian Journal of Botany* **57**(23): 2603-2616.

Nishimura R, Hayashi M, Wu G-J, Kouchi H, Imaizumi-Anraku H, Murakami Y, Kawasaki S, Akao S, Ohmori M, Nagasawa M. 2002. HAR1 mediates systemic regulation of symbiotic organ development. *Nature* **420**(6914): 426-429.

Ogawa M, Shinohara H, Sakagami Y, Matsubayashi Y. 2008. Arabidopsis CLV3 peptide directly binds CLV1 ectodomain. *Science* **319**(5861): 294-294.

Ohyama K, Shinohara H, Ogawa-Ohnishi M, Matsubayashi Y. 2009. A glycopeptide regulating stem cell fate in *Arabidopsis thaliana*. *Nature Chemical Biology* **5**(8): 578-580.

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**(1): 67-77.

Okamoto S, Shinohara H, Mori T, Matsubayashi Y, Kawaguchi M. 2013. Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. *Nature Communications* **4**: 2191.

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**(4): 1347-1352.

Osipova M, Dolgikh E, Lutova L. 2011. Peculiarities of meristem-specific *WOX5* gene expression during nodule organogenesis in legumes. *Russian Journal of Developmental Biology* **42**(4): 226-237.

Osipova MA, Mortier V, Demchenko KN, Tsyganov VE, Tikhonovich IA, Lutova LA, Dolgikh EA, Goormachtig S. 2012. *WUSCHEL-RELATED HOMEOBOX5* gene expression and interaction of CLE peptides with components of the systemic control add two pieces to the puzzle of autoregulation of nodulation. *Plant Physiology* **158**(3): 1329-1341.

Radutoiu S, Madsen LH, Madsen EB, Felle HH, Umehara Y, Grønlund M, Sato S, Nakamura Y, Tabata S, Sandal N. 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**(6958): 585-592.

Redmond JW, Batley M, Djordjevic MA, Innes RW, Kuempel PL, Rolfe BG. 1986. Flavones induce expression of nodulation genes in *Rhizobium*. *Nature* **323**(6089): 632-635.

Reid DE, Ferguson BJ, Gresshoff PM. 2011a. Inoculation-and nitrate-induced CLE peptides of soybean control NARK-dependent nodule formation. *Molecular Plant-Microbe Interactions* **24**(5): 606-618.

Reid DE, Ferguson BJ, Hayashi S, Lin Y-H, Gresshoff PM. 2011b. Molecular mechanisms controlling legume autoregulation of nodulation. *Annals of Botany* **108**(5): 789-795.

Remigi P, Zhu J, Young JPW, Masson-Boivin C. 2016. Symbiosis within symbiosis: evolving nitrogen-fixing legume symbionts. *Trends in Microbiology* **24**(1): 63-75.

Sarkar AK, Luijten M, Miyashima S, Lenhard M, Hashimoto T, Nakajima K, Scheres B, Heidstra R, Laux T. 2007. Conserved factors regulate signalling in *Arabidopsis thaliana* shoot and root stem cell organizers. *Nature* **446**(7137): 811-814.

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**(5603): 109-112.

Smit P, Limpens E, Geurts R, Fedorova E, Dolgikh E, Gough C, Bisseling T. 2007. Medicago LYK3, an entry receptor in rhizobial nodulation factor signaling. *Plant Physiology* **145**(1): 183-191.

Sparks E, Wachsman G, Benfey PN. 2013. Spatiotemporal signalling in plant development. *Nature Reviews Genetics* **14**(9): 631-644.

Stahl Y, Grabowski S, Bleckmann A, Kühnemuth R, Weidtkamp-Peters S, Pinto KG, Kirschner GK, Schmid JB, Wink RH, Hülsewede A. 2013. Moderation of Arabidopsis root stemness by CLAVATA1 and ARABIDOPSIS CRINKLY4 receptor kinase complexes. *Current Biology* **23**(5): 362-371.

Stahl Y, Wink RH, Ingram GC, Simon R. 2009. A signaling module controlling the stem cell niche in Arabidopsis root meristems. *Current Biology* **19**(11): 909-914.

Strabala TJ, Phillips L, West M, Stanbra L. 2014. Bioinformatic and phylogenetic analysis of the CLAVATA3/EMBRYO-SURROUNDING REGION (CLE) and the CLE-LIKE signal peptide genes in the *Pinophyta*. *BMC Plant Biology* **14**(1): 47.

Sutton MA, Howard CM, Erisman JW, Billen G, Bleeker A, Grennfelt P, Van Grinsven H, Grizzetti B. 2011a. *The European nitrogen assessment: sources, effects and policy perspectives*: Cambridge University Press.

Sutton MA, Oenema O, Erisman JW, Leip A, van Grinsven H, Winiwarter W. 2011b. Too much of a good thing. *Nature* **472**(7342): 159-161.

Suzaki T, Yoro E, Kawaguchi M. 2015. Chapter three-leguminous plants: inventors of root nodules to accommodate symbiotic bacteria. *International Review of Cell and Molecular Biology* **316**: 111-158.

Tsikou D, Zhe Y, Stougaard J, Markmann K. 2016. 'A Micro RNA Acts as a Signal in Systemic Control of Nodulation Symbiosis', in É Kondorosi, G Endre, P Kalo & A Kereszt (eds). 12 European Nitrogen Fixation Conference, Budapest, Hungary. 26 August 2016 pp 42.

Trotochaud AE, Hao T, Wu G, Yang Z, Clark SE. 1999. The CLAVATA1 receptor-like kinase requires CLAVATA3 for its assembly into a signaling complex that includes KAPP and a Rho-related protein. *The Plant Cell* **11**(3): 393-405.

Udvardi MK, Price GD, Gresshoff PM, Day DA. 1988. A dicarboxylate transporter on the peribacteroid membrane of soybean nodules. *FEBS Letters* **231**(1): 36-40.

Vicente EJ, Dean DR. 2017. Keeping the nitrogen-fixation dream alive. *Proceedings of the National Academy of Sciences* **114**(12): 3009-3011.

Xu C, Liberatore KL, MacAlister CA, Huang Z, Chu Y-H, Jiang K, Brooks C, Ogawa-Ohnishi M, Xiong G, Pauly M. 2015. A cascade of arabinosyltransferases controls shoot meristem size in tomato. *Nature Genetics* **47**(7): 784-792.

Yang J, Xie X, Wang X, Dixon R, Wang Y-P. 2014. Reconstruction and minimal gene requirements for the alternative iron-only nitrogenase in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **111**(35): E3718-E3725.

Chapter 2

The structure and activity of nodulation-suppressing CLE peptide hormones of legumes

Preface

CLE peptides are signalling molecules involved in numerous plant developmental processes as described in Chapter 1. Of particular relevance to legumes are those CLE peptides involved in nodulation regulation. This chapter is a comprehensive review describing in depth the current knowledge at the time of publication on nodulation-suppressive CLE peptides. This chapter has been published in Functional Plant Biology (2015, https://doi.org/10.1071/FP14222).

2.1 Abstract

Legumes form a highly-regulated symbiotic relationship with specific soil bacteria known as rhizobia. This interaction results in the *de novo* formation of root organs called nodules, in which the rhizobia fix atmospheric di-nitrogen (N2) for the plant. Molecular mechanisms that regulate the nodulation process include the systemic 'autoregulation of nodulation' and the local nitrogen-regulation of nodulation pathways. Both pathways are mediated by novel peptide hormones called CLAVATA/ESR-related (CLE) peptides that act to suppress nodulation via negative feedback loops. The mature peptides are 12–13 amino acids in length and are post-translationally modified from the C-terminus of tripartite-domain prepropeptides. Structural redundancy between the prepropeptides exists; however, variations in external stimuli, timing of expression, tissue specificity and presence or absence of key functional domains enables them to act in a specific manner. To date, nodulation-regulating CLE peptides have been identified in *Glycine max* (L.) Merr., *Medicago truncatula* Gaertn., *Lotus japonicus* (Regel) K. Larsen and *Phaseolus vulgaris* L. One of the *L. japonicus* peptides, called LjCLE-RS2, has been structurally characterised and found to be an arabinosylated glycopeptide. All of the known nodulation CLE peptides act via an orthologous leucine rich repeat (LRR) receptor kinase. Perception of the peptide results in the production of a novel, unidentified inhibitor signal that acts to suppress further nodulation events. Here, we contrast and compare the various nodulation-suppressing CLE peptides of legumes.

2.2 Introduction

The common agricultural practice of using nitrogen-based fertilisers to increase crop yields has been highly successful in generating sufficient food for the world's ever-growing population. It has been a major part of the 'green revolution' instigated more than 50 years ago. However, adverse economical and ecological consequences are beginning to outweigh the benefits of nitrogen fertiliser use (Erisman et al. 2008; Sutton et al. 2011; Jensen et al. 2012).

Symbiotic nitrogen fixation represents an alternative to chemical nitrogen fertiliser use. It involves a relationship mainly formed between plant species of the family Fabaceae, commonly known as legumes, and soil bacteria, collectively referred to as rhizobia. Major legume crop and pasture species include soybean, pea, common bean, clover, cowpea, medic, chickpea, lentil and peanut. Biological nitrogen fixation from this legume–rhizobia relationship currently results in ~50–70 Tg of nitrogen added into global agricultural systems each year (Herridge et al. 2008; Jensen et al. 2012).

The legume-rhizobia relationship is signified by the formation of a new plant organ, called the nodule. Nodule development is orchestrated by a complex signalling interaction (Ferguson and Mathesius 2003, 2014; Ferguson et al. 2010; Desbrosses and Stougaard 2011; Oldroyd 2013). Once formed, the nodule acts to house the rhizobia that provide the plant with a useable form of reduced nitrogen (namely ammonia) using a specialised enzyme complex to 'fix' un-reactive atmospheric di-nitrogen gas (N_2). In return, the rhizobia are provided with a carbon source derived from photosynthesis, predominately malate (Udvardi et al. 1988). In addition to increasing current crop yields, this process is exploited in agriculture to improve the nitrogen content and structure of soils by using legumes as rotation crops (Jensen et al. 2012).

2.3 Control of legume nodule numbers

Nodulation is costly to the host plant in terms of resources; as a result, the plant has developed both local and systemic mechanisms to control its nodule numbers (Delves et al. 1986; Gresshoff and Delves 1986; recently reviewed by Reid et al. 2011b). Local control mechanisms responding to high soil nitrate directly prevent or delay nodule development (Carroll et al. 1985a; Reid et al. 2011a). A systemic control mechanism, called the 'autoregulation of nodulation' (AON), is closely associated with the nitrate

regulatory pathway, but is induced by rhizobia, not nitrate, and acts systemically through the shoot, rather than locally in the root (Kosslak and Bohlool 1984; Delves et al. 1986; Gresshoff and Delves 1986; Reid et al. 2011a, 2011b).

The AON process begins with the production of a root-derived signal (Gresshoff and Delves 1986), which is expressed in response to a transcription factor, called NIN, involved in cortical cell division during early nodulation events (Soyano et al. 2014). This signal, formerly called 'Q' (Gresshoff and Delves 1986), is now known to be a CLAVATA/Embryo surrounding region (ESR) related (CLE) peptide. To date, CLE peptide-encoding genes having a role in nodulation have been identified in *Glycine max* (L.) Merr. (soybean), *Medicago truncatula* Gaertn., *Lotus japonicus* (Regel) K. Larsen and *Phaseolus vulgaris* L. (common bean) (Fig. 2.1; Okamoto et al. 2009, 2013; Mortier et al. 2010, 2012; Lim et al. 2011; Saur et al. 2011; Reid et al. 2011a, 2013; Ferguson et al. 2014). Recent biochemical advances have enabled the isolation and identification of one of the nodulation CLE peptides of *L. japonicus*, called LjCLE-RS2. The mature signal of this CLE peptide is 13 amino acids in length, is derived from a much larger prepropeptide and is post-translationally modified with three β 1-2 linked arabinose moieties at Hyp7 (Okamoto et al. 2013).

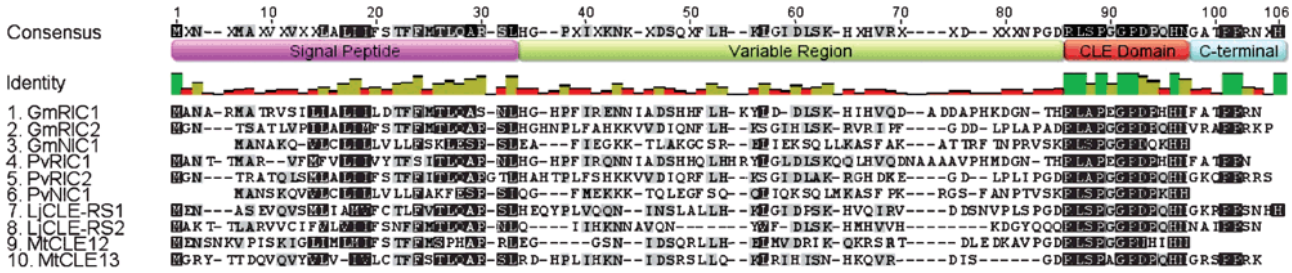


Fig. 2.1. Multiple sequence alignment and domain structure of the nodulation CLE prepropeptides. Shown are the amino acid sequences of the known nodulation-suppressing CLE peptides of *Glycine max* (soybean), *Phaseolus vulgaris* (common bean), *Lotus japonicus* and *Medicago truncatula*. The alignment was obtained using CLUSTALW multiple alignment (Larkin *et al*. 2007) in Geneious Pro 6.0. Shading of individual amino acids represents conservation amongst the prepropeptides, with the darker the shading the more highly conserved the residue. The CLE domain is highly conserved, with many other conserved residues found in the signal peptide and *C*-terminal extension domains. Conservation is particularly strong between orthologous genes of the different species. Not shown are the homeologous/duplicate copies of the soybean genes, which may have no-, reduced- or an alternative-function.

The nodulation-suppressing CLE peptide signal is exported from the root and transported via the xylem by an unknown mechanism (Okamoto et al. 2013) to the leaf phloem

parenchyma (Nontachaiyapoom et al. 2007) where it is perceived by a leucine-rich repeat serine-threonine receptor kinase (LRR RK), called GmNARK in soybean, LjHAR1 in *L. japonicus*, MtSUNN in *M. truncatula*, PsSYM29 in *Pisum sativum* L. (pea), GsNARK in *Glycine soja* Siebold & Zucc., and PvNARK in common bean (Krusell et al. 2002; Nishimura et al. 2002; Searle et al. 2003; Schnabel et al. 2005; Ferguson et al. 2014). These LRR RKs may act in a complex with other receptors to perceive the CLE peptide ligand. This includes factors such as LjCLAVATA2/PsCLAVATA2 and LjKLAVIER (Miyazawa et al. 2010; Krusell et al. 2011). Additional research has identified other factors that may interact with the LRR RK directly, or function downstream of it, to relay the perception of the signal and trigger downstream signalling events. This includes the kinase-associated protein phosphatases, GmKAPP1 and GmKAPP2 (Miyahara et al. 2008), the putative ubiquitin fusion degradation protein, GmUFD1a (Reid et al. 2012) and the root-acting F-box protein, TOO MUCH LOVE, LjTML (Magori et al. 2009; Takahara et al. 2013). Following the perception of the nodulation CLE peptide signal, a shoot-derived inhibitor (SDI) signal is produced and transported to the roots, likely via the phloem, to inhibit further nodulation development (Delves et al. 1986; Lin et al. 2010, 2011; Reid et al. 2011b; Sasaki et al. 2014). Recent studies in *L. japonicus* have indicated a role for cytokinin as a potential SDI-candidate in AON (Sasaki et al. 2014).

The gene encoding for the LRR RK is expressed in both shoot and root tissues (Krusell et al. 2002; Nishimura et al. 2002; Searle et al. 2003; Schnabel et al. 2005; Nontachaiyapoom et al. 2007) and plants having mutations in it exhibit both supernodulation (due to a lack of AON control) and nitrate-tolerant nodulation phenotypes (e.g. Carroll et al. 1985a, 1985b). Grafting studies using soybean have demonstrated that the *GmNARK* LRR RK is required for both AON in the shoot (Delves et al. 1986; Reid et al. 2011a) and nitrate-regulation of nodulation in the root (Reid et al. 2011a). Similar to AON, the nitrate regulation of nodulation mechanism in soybean begins with the production of a CLE peptide that is predicted to be perceived by GmNARK. However, unlike AON, this CLE peptide, called GmNIC1, responds to nitrate, not rhizobia, and acts locally in the root, not systemically in the shoot. These findings helped to confirm that there are two independent pathways controlling nodulation: the systemic rhizobia-induced AON pathway and the local nitrate-induced regulation of nodulation pathway (reviewed by Reid et al. 2011b). No candidates for the root-derived inhibitor (RDI) have been identified to date, but as it is produced downstream of GmNARK, it may be similar to, or even the same as, the SDI signal in AON.

Unlike soybean, *L. japonicus* and *M. truncatula* appear to have overlapping local and systemic molecular mechanisms that act to regulate nodulation in response to both rhizobia and nitrate (Okamoto et al. 2009; Mortier et al. 2010). Although the reason for this difference amongst species is unknown, it is likely that it relates to genomic duplication events undergone in soybean that have enabled genetic divergence and the development of new molecular signals and mechanisms through the process of neofunctionalisation (Schmutz et al. 2010). This may also explain why soybean has three functional CLE peptides that are known to regulate nodule numbers, in addition to three homeologous (duplicate) copies that may have no-, reduced- or diverged-function (Reid et al. 2011a), whereas *L. japonicus* and *M. truncatula* appear to have only two such peptides (Table 2.1). Interestingly common bean, which shared a duplication event with soybean 53 million years ago, has orthologous copies of the three soybean CLE peptide genes, but lacks the duplicate copies of each of these genes as a result of not undergoing the more recent genome duplication event approximately 13 million years ago (Ferguson et al. 2014).

The CLE peptides that act as a trigger for AON and nitrate-regulation of nodulation belong to a large group of heavily processed, cysteine-poor secreted plant peptides related to AtCLV3 in *Arabidopsis thaliana* (L. Heynh) (Matsubayashi 2014). AtCLV3 functions in the CLAVATA pathway to regulate the shoot apical meristem stem cell population. It acts as a ligand to a receptor-complex involving AtCLV1, AtCLV2 and AtCORYNE (Ogawa et al. 2008). The AtCLV1 receptor is a LRR RK that is highly similar in structure to the LRR RKs that are central to nodulation control. Other Arabidopsis CLE peptides of note that are similar to the nodulation CLE peptides of legumes include AtCLE1 to AtCLE7, which have roles in root architecture and development (Cock and McCormick 2001; Strabala et al. 2006; Oelkers et al. 2008; Araya et al. 2014).

The mechanisms controlling nodulation in legumes are highly conserved, as demonstrated by the interspecific function of AON CLE peptides from soybean in common bean and from *M. truncatula* in pea (Osipova et al. 2012; Ferguson et al. 2014). There are, however, many differences in the sequences, structures and inducing factors of the various nodulation CLE peptides that allow for specificity of function (Fig. 2.1; Tables 2.1, 2.2). These similarities and differences, and how they impact on nodule suppression, are reviewed here.

Table 2.1 Known nodulation CLE peptides and their key reported features

Species/peptide	Prepropeptide length (aa)	Receptor	Local/systemic	Mode of induction	Induction time	Signal peptide motif (TLQAR) conservation	*C*-terminal domain	Refererence
Glycine max								
GmRIC1	95	GmNARK	Systemic	Rhizobia	Early (<12 h)	Y (80%)	Y	Reid et al. (2011a)
GmRIC2	93	GmNARK	Systemic	Rhizobia	Late (48–72 h)	Y (100%)	Y	Reid et al. (2011a)
GmNIC1	80	GmNARK	Local	Nitrate		N (<40%)	N	Reid et al. (2011a)
Phaseolus vulgaris								
PvRIC1	97	PvNARK	ND	Rhizobia	Early (<24 h)	Y (100%)	Y	Ferguson et al. (2014)
PvRIC2	93	PvNARK	ND	Rhizobia	Late (<5 days)	Y (100%)	Y	Ferguson et al. (2014)
PvNIC1	80	PvNARK	ND	ND	ND	N (<40%)	N	Ferguson et al. (2014)
Lotus japonicus								
LjCLE-RS1	93	LjHAR1	Systemic	Rhizobia and nitrate	Early (<24 h)	Y (100%)	Y	Okamoto et al. (2009)
LjCLE-RS2	81	LjHAR1	Systemic	Rhizobia	Early (<24 h)	Y (100%)	Y	Okamoto et al. (2009)
Medicago truncatula								
MtCLE12	81	MtSUNN	Systemic	Rhizobia	Late (~4–6 days)	N (<40%)	N	Mortier et al. (2010)
MtCLE13	84	MtSUNN	Systemic	Rhizobia	Early (<4 days)	Y (100%)	Y	Mortier et al. (2010)

ND, not determined

Table 2.2 Amino acid sequence similarity (%) amongst the known nodulation CLE prepropeptides

Similarities are based on alignments obtained using CLUSTALW multiple alignment tool in Geneious Pro 6.0

	GmRIC1	GmRIC2	GmNIC1	PvRIC1	PvRIC2	PvNIC1	LjCLE-RS1	LjCLE-RS2	MtCLE12
GmRIC2	49.9	–	–	–	–	–	–	–	–
GmNIC1	26.4	24.1	–	–	–	–	–	–	–
PvRIC1	68.7	61.1	15.4	–	–	–	–	–	–
PvRIC2	47.4	82.1	21.8	40.4	–	–	–	–	–
PvNIC1	21.4	30.7	69.1	22.0	23.0	–	–	–	–
LjCLE-RS1	42.3	47.9	26.2	40.4	48.4	24.4	–	–	–
LjCLE-RS2	45.8	41.5	24.4	43.0	34.0	27.2	34.4	–	–
MtCLE12	32.6	33.0	22.6	30.9	32.6	25.0	40.2	28.4	–
MtCLE13	37.9	44.1	25.0	37.8	43.0	26.3	51.6	42.0	30.6

2.4 Key functional domains of CLE peptides

Mature CLE peptide signals are derived from prepropeptides consisting of 3–4 domains: a *N*-terminal signal peptide, a variable region and a CLE domain, with some also having a *C*-terminal extension (Fig. 2.1). Sequence similarities amongst the nodulation CLE prepropeptides shows the orthologous copies are most similar (Fig. 2.2; Table 2.2); however, it is likely that similarities and differences in the individual domains are most critical for driving specificity. Here we discuss the function, conservation and importance of each domain, particularly in respect to their role in the suppression of nodulation.

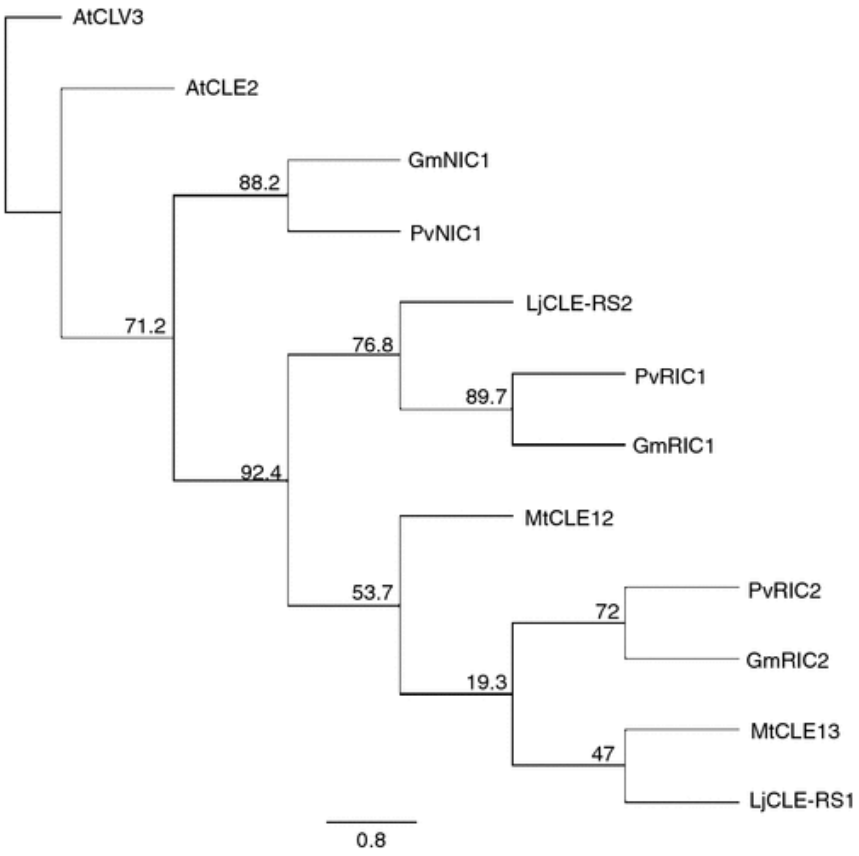


Fig. 2.2. Phylogenetic tree of nodulation CLE prepropeptides. The known nodulation-suppressing CLE peptides of *Medicago truncatula*, *Lotus japonicus*, *Glycine max* (soybean) and *Phaseolus vulgaris* (common bean) are shown, together with AtCLV2, the *Arabidopsis* CLE peptide most similar to the nodulation CLE peptides, and AtCLV3 as an outgroup. The tree was generated using PhyML 3.0 (Guindon and Gascuel 2003) in Geneious Pro 6.0 and constructed using the maximum likelihood approach. A branch was supported in 1000 bootstrap replications, with bootstrap confidence values expressed as a percentage of the 1000 bootstrap replications (Felsenstein 1985).

2.4.1 Signal peptide

The *N*-terminal hydrophobic signal peptide (also referred to as a transit peptide) is widely thought to be responsible for exporting the prepropeptide out of the cell (Ia Cour et al. 2004; Lim et al. 2011). It is ~30 amino acids in length and is critical to the specificity of the peptide (Fletcher et al. 1999; Reid et al. 2013). This domain has a role in exporting the AtCLV3 propeptide into the extracellular space (Rojo et al. 2002). A similar role for this domain is predicted for the nodulation CLE peptides.

Amongst the known nodulation CLE prepropeptides, the signal peptide domain has a 37% pairwise identity and contains a leucine-rich motif (Fig. 2.1), commonly observed in exported proteins (Ia Cour et al. 2004). Also present within this domain is a conserved motif of five amino acids (TLQAR; Table 2.1), which is predicted to be a site of cleavage (Okamoto et al. 2009). It has been noted that GmNIC1, PvNIC1 and MtCLE12 show lower conservation of amino acid residues within this motif. They are also the only known nodulation CLE peptides to lack the *C*-terminal extension domain (Table 2.1; Fig. 2.1). Outside of this motif, conserved sequence residues within the signal peptide can be seen amongst predicted orthologues of the nodulation CLE peptides (Fig. 2.1; Reid et al. 2011a; Ferguson et al. 2014).

2.4.2 Variable region

The functional importance of the variable domain, the least conserved of the four domains, remains unknown (Ni and Clark 2006; Meng et al. 2010; Reid et al. 2013). Indeed, AtCLV3 shows function without this domain (Fiers et al. 2006). The size of the domain is also highly variable (31–50 amino acids). However, recognition and cleavage immediately before the Arg1 residue of the CLE domain requires at least four to five residues of the variable domain to be present for correct processing of AtCLV3 (Kondo et al. 2008; Ni et al. 2011; Xu et al. 2013). An additional amino acid at residue 39 within the variable domain of AtCLV3 is also predicted to be a cleavage site (Xu et al. 2013).

There are no residues that are 100% conserved across the variable domain between the known nodulation CLE peptides (Fig. 2.1), although it shows a 19.4% pairwise identity and, as with other domains, residues are conserved between orthologues. This is particularly evident between the nodulation CLE peptides of the closely-related bean and soybean species (Fig. 2.1; Ferguson et al. 2014).

2.4.3 CLE domain

The CLE domain, from which the peptide is named, denotes the mature/active peptide sequence. It is located at the *C*-terminus and is the most conserved region (Cock and McCormick 2001; Oelkers et al. 2008). The consensus amino acid sequence of the nodulation-suppressing CLE peptides is RL(A/S)PGGPDPQHN(X) (Fig. 2.1). The domain is 12 or 13 amino acids in length and contains 50% identical sites, with 77.4% (12 amino acids) and 75.2% (13 amino acids) pairwise identity between the known nodulation CLE peptides. LjCLE-RS2 of *L. japonicus* is the only structurally-confirmed nodulation CLE peptide, and is 13 amino acids in length (Okamoto et al. 2013). However, the nitrate-induced GmNIC1 peptide of soybean and its orthologue in bean, PvNIC1, have a stop codon at position 13 and therefore can only be 12 amino acids in length (Fig. 2.1; Reid et al. 2011a; Ferguson et al. 2014). This may influence their functional properties, such as their apparent lack of long distance transport.

Notably, GmRIC1, PvRIC1, GmRIC2, and PvRIC2 are the only nodulation CLE peptides known to contain an Ala residue at position 3 of the CLE domain, presumably a result of polyploidisation and subsequent species divergence amongst the legumes (Figs 2.1, 2.2; Stefanović et al. 2009; Schmutz et al. 2010). There are four other residues within the CLE domain of the known nodulation CLE peptides that contain sequence divergence from the consensus sequence: Gly5 > Glu5 (GmRIC1 and PvRIC1) or Ala5 (MtCLE13); Asp8 > Asn8 (MtCLE12); Pro9 > His9 (MtCLE12) or Gln9 (GmNIC1); and Gln10 > His10 (GmRIC1 and PvRIC1) or Ile10 (MtCLE12) (Fig. 2.1). It is not yet known how the activity of the CLE peptide is affected by these sequence divergences. Only the one at position 8 in MtCLE12 is predicted to be critical for function (i.e. the suppression of nodulation) based on site-directed mutagenesis work using soybean (Reid et al. 2013); however, this nonsynonymous substitution from an uncharged asparagine to a negatively charged aspartic acid is conservative and may not affect activity.

Recent research has indicated that, despite sequence-redundancy of the CLE domain, there is likely some specificity between pathways and/or species that are dependent on sequence. Okamoto et al. (2013) were unable to elicit a plant response in *L. japonicus* from exogenous application of the mature AtCLV3 peptide, but saw a reduction in nodules when LjCLE-RS2 was applied with the correct post-translational modifications. Chimeric genes that swapped the CLE domains of GmNIC1 and GmRIC1 also impacted on the suppression of nodulation compared with their respective native genes (Reid et al. 2013).

In contrast, *GmRIC1* overexpression in common bean and *MtCLE13* overexpression in pea strongly suppressed nodulation inter-specifically (Osipova et al. 2012; Ferguson et al. 2014), indicating that these CLE peptide-encoding genes can function in the AON pathways of other legume species. However, overexpression results of any kind should always be interpreted with care.

2.4.4 C-terminal extension domain

The *C*-terminal domain of the known nodulation-suppressing CLE peptides is small, at ~6–9 residues in length, and is even completely absent from some (Fig. 2.1; Table 2.1). Indeed, of the known nodulation CLE peptides, GmNIC1, PvNIC1 and MtCLE12 all lack the *C*-terminal extension in its entirety (Mortier et al. 2010; Reid et al. 2011a; Ferguson et al. 2014). However, the remaining nodulation CLE peptides all contain the domain, as do AtCLV3 and GmCLV3 of the CLAVATA pathway (Fiers et al. 2006; Wong et al. 2013).

The *C*-terminal domain is thought to act as a protective mechanism from degradative protease enzymes in the xylem, which the peptides would encounter during systemic transport (Oelkers et al. 2008; Okamoto et al. 2009; Ni et al. 2011; Reid et al. 2011a). It is characteristic of the rhizobia-dependent, systemically acting, CLE peptides, and is not present in the nitrate-induced, locally-acting GmNIC1 of soybean and its orthologue in bean, PvNIC1 (Reid et al. 2011a; Ferguson et al. 2014). This would appear to further support a role for the domain in protection during long-distance xylem transport. Moreover, overexpressing a chimeric construct that added the *C*-terminal domain of GmRIC1 to GmNIC1 enhanced the suppression of nodulation compared with that of the native GmNIC1 (Reid et al. 2013). In contrast, the removal of the domain from GmRIC1 did not alter its ability to suppress when overexpressed (Reid et al. 2013), but this may be due to the overexpression technique masking or over-compensating for the true function of the modified construct. MtCLE12 also lacks the *C*-terminal domain and is both induced by rhizobia and predicted to be transported systemically (Mortier et al. 2010), so the exact need for the domain remains puzzling.

Two conserved proline residues are present within the *C*-terminal extension of all seven nodulation CLE peptides that contain the domain (Fig. 2.1; Okamoto et al. 2009; Mortier et al. 2010; Reid et al. 2011a; Ferguson et al. 2014). Site-directed mutagenesis and overexpression of *GmRIC1* modified to encode two alanine residues in place of these two

proline residues did not alter the suppressive activity of the peptide, consistent with the unclear role of this domain (Reid et al. 2013).

2.5 Post-translational modifications and critical residues of the CLE domain

The mature nodulation-suppressive CLE peptide of *L. japonicus*, LjCLE-RS2, is 13 amino acids in length and is hydroxylated at Pro4 and Pro7, with Hyp7 further modified to contain three arabinose sugars connected via β-1-2-linkages. These modifications are predicted to be made in the extracellular fluids (Okamoto et al. 2013). This is consistent with mature AtCLV3, AtCLE2 and AtCLE9 glycopeptides, which also contain a Hyp7 having three linked L-arabinose sugars (Kondo et al. 2006; Ohyama et al. 2009; Okamoto et al. 2013; Shinohara and Matsubayashi 2013). All of the nodulation CLE peptides contain motifs associated with arabinose modifications that are present in other plant proteins/peptides (Matsubayashi 2014). The hydroxyproline O-arabinosyltransferase (HPAT) gene that controls CLE arabinosylation in Arabidopsis is called *AtHPAT3* (Ogawa-Ohnishi et al. 2013). *MtRDN1* and *PsNOD3* are likely orthologues of *AtHPAT3* and are thought to be responsible for the arabinosylation of the nodulation-suppressing CLE peptides (Ogawa-Ohnishi et al. 2013). Mutations in these genes result in a supernodulation phenotype (Jacobsen and Feenstra 1984; Postma et al. 1988; Sagan and Duc 1996; Li et al. 2009; Schnabel et al. 2011), indicating that the peptides require the arabinose sugars for their activity.

Application of synthesised arabinosylated-LjCLE-RS2 to leaves of *L. japonicus* plants caused a reduction in nodulation in an *LjHAR1*-dependent manner (Okamoto et al. 2013). However, root or shoot application of synthetic nodulation CLE peptides devoid of modifications did not affect nodulation, although altered root growth was observed (Okamoto et al. 2009; Saur et al. 2011). Moreover, application of AtCLV3 with the arabinose modifications also had no effect on nodulation (Okamoto et al. 2013). Shinohara and Matsubayashi (2013) demonstrated that the binding of the AtCLV3 CLE peptide to the AtCLV1 LRR receptor-kinase declined as the arabinose chain length decreased, whereas AtCLE9 showed no change in receptor binding efficacy to its receptor, BAM1, a CLV1/BAM-family LRR RK, in the absence of the arabinose chain (Shinohara et al. 2012). Further, tracheary element differentiation inhibitory factor (TDIF) peptides synthesised with

or without hydroxyproline residues can mimic the function of the naturally occurring peptide, which contains Hyp4 and Hyp7 (Sawa et al. 2006).

In addition to post-translational modifications to critical residues, the structural configuration of the CLE peptide ligand is also likely to impact markedly on receptor interactions. Gly6 is proposed to allow for rotation, most likely because of its small size, complementing evidence for a boomerang curve in the peptide's configuration, with both ends of the peptide bending away from the arabinosylation at Hyp7 (Okamoto et al. 2013; Shinohara and Matsubayashi 2013; Song et al. 2013). Notably, Gly6 is 100% conserved amongst the known nodulation CLE peptides (Fig. 2.1). Site-directed mutagenesis of Gly6 to Ala6 significantly reduced the nodule suppressive activity of GmRIC1 (Reid et al. 2013). Song et al. (2013)altered Gly6 of AtCLV3 into 18 other amino acids; no substitution was able to rescue the phenotype of *Atclv3* mutant plants. Similar specificity is expected for the nodulation CLE peptides. In addition to Gly6 and Pro7, residues Arg1, Pro4, Asp8, His11 and Asn12 of GmRIC1 were required for full nodulation-suppression activity in soybean (Reid et al. 2013). Similarly, TDIF also lost activity when the CLE domain residues His1, Val3, Gly6, Asn8, Pro9 and Asn12 were changed into an alanine residue via site-directed mutagenesis (Ito et al. 2006; Sawa et al. 2006).

It has been noted that locally-acting CLE peptides, including GmNIC1 and PvNIC1 (Fig. 2.1), in addition to AtCLV3, GmCLV3 and LjCLV3, all contain His12 (Reid et al. 2011a; Okamoto et al. 2011; Wong et al. 2013; Ferguson et al. 2014). This may indicate a role for this residue in local, but not systemic, transport of the peptide. Constructs having swapped the CLE domain of the systemically-acting GmRIC1 and the locally-acting GmNIC1 showed an altered inhibition of nodulation when overexpressed compared with the native peptides (Reid et al. 2013). Whether residue 12 plays a specific role in the transport or recognition of the peptide is of interest to determine.

As noted above, Arg1 of the AtCLV3 CLE domain has been shown to be critical for binding and processing of the mature CLE peptide, with at least 4–5 residues upstream of Arg1 required for proper recognition of the signal (Kondo et al. 2008; Ni et al. 2011; Xu et al. 2013). It is hypothesised that a subtilisin with endoproteolytic activity cleaves the CLE peptide, with a carboxypeptidase processing the C-terminal extension where present (Ni et al. 2011; Djordjevic et al. 2011). However, to date, there is little known about the mechanisms and sites of proteolytic cleavage in the nodulation CLE peptides.

2.6 Mode of induction of the nodulation-suppressing CLE peptides

All of the known nodulation-suppressing CLE peptides are upregulated in expression by the presence of rhizobia and/or the available soil nitrogen content (Table 2.1). Phylogenetic analysis shows that they cluster according to their mode of induction (Fig. 2.2). Evidence for other environmental factors such as phosphate and soil acidity, inducing or influencing the expression of CLE peptide-encoding genes also exists.

2.6.1 Rhizobia-induced CLE peptides

The presence of compatible rhizobia, and possibly more specifically the rhizobia-produced Nod factor signal, elicits the expression of systemically-acting CLE peptide-encoding genes that function in AON. These CLE peptides include: *LjCLE-RS1*, *LjCLE-RS2*, *MtCLE12*, *MtCLE13*, *GmRIC1*, *GmRIC2*, *PvRIC1*, and *PvRIC2*(Table 1; Okamoto et al. 2009, 2013; Mortier et al. 2010, 2012; Lim et al. 2011; Reid et al. 2011a, 2013; Saur et al. 2011; Hayashi et al. 2012; Ferguson et al. 2014). Overexpression of these peptides in wild-type legume plants results in a complete abolishment of nodulation, but does not alter the nodulation pattern in NARK mutants, demonstrating that they act in a NARK-dependent manner (Okamoto et al. 2009; Mortier et al. 2010; Reid et al. 2011a; Lim et al. 2011).

Laser microdissection of root sections indicate that *LjCLE-RS1* and *-RS2* are expressed in the stele and outside of the endodermis (cortex and epidermis) (Okamoto et al. 2009). Promoter:*GUS* reporter fusion studies have shown that *MtCLE13* is expressed in the inner cortex during early nodulation and later in dividing cells of the cortex and pericycle. In contrast, MtCLE12 is not expressed early but instead is expressed throughout young nodules and in meristematic tissues of the elongating indeterminate nodule (Mortier et al. 2010). Finally, Lim et al. (2011) have shown that *GmRIC2* is expressed in the pericycle and inner cortex during early nodule development, and later in the outer cortex of more developed nodules.

Time-course experiments have revealed different but overlapping expression patterns for these genes within a species (Table 2.1). Soybean *GmRIC1* is induced early (within 12 h) after inoculation with infection-capable (Nod factor producing) *Bradyrhizobium japonicum*, whereas *GmRIC2* expression is induced later (48–72 h) and remains elevated in

expression for longer (Reid et al. 2011a; Hayashi et al. 2012). The rhizobia-induced peptide encoding genes of common bean, *PvRIC1* and *PvRIC2*, exhibit a similar pattern of expression (Ferguson et al. 2014). Likewise, *M. truncatula MtCLE13* is expressed earlier than *MtCLE12*, although both are also expressed in later stages of nodulation (Mortier et al. 2010, 2012). *LjCLE-RS1* and *LjCLE–RS2* are both upregulated within 3 h of inoculation (Okamoto et al. 2009). Similar to *GmRIC2*, *MtCLE13* and *LjCLE-RS1* transcript levels appear to remain elevated for longer compared with *MtCLE12*and *LjCLE–RS2*, respectively (Okamoto et al. 2009; Mortier et al. 2010; Reid et al. 2011a).

When compared with wild-type plants, a significant increase in expression of both *LjCLE-RS1* and *LjCLE–RS2* was also observed in the hypernodulating mutant of *L. japonicus*, *too much love*, possibly indicating that their synthesis is directly linked to the number of nodules being formed (Magori and Kawaguchi 2010). Interestingly, the plant hormone cytokinin, which has a role in early nodule development (reviewed in Ferguson and Mathesius 2014), has also been shown to induce the expression of some nodulation-suppressing CLE peptide genes (Lim et al. 2011; Mortier et al. 2010, 2012), consistent with the idea that the initiation of the AON pathway is linked to early cell divisions. Additional studies are required to further understand the expression patterns of these nodulation-suppressing CLE peptides, both within and between species.

2.6.2 Nitrate-induced CLE peptides

GmNIC1 and *LjCLE-RS2* are the only nodulation-suppressing CLE peptide-encoding genes that are confirmed to respond to nitrate (Table 2.1; Okamoto et al. 2009; Reid et al. 2011a). *GmNIC1* is specifically induced by nitrate and not co-induced by the rhizobial microsymbiont, whereas *LjCLE-RS2* is reported to be induced by both. *PvNIC1* is also likely to be induced by nitrate as the candidate orthologue of *GmNIC1*(Ferguson et al. 2014). To date, no CLE peptide-encoding gene of *M. truncatula* has been reported to respond to nitrate, although evidence suggests the existence of a locally acting, nitrate-responsive mechanism that acts in a MtSUNN-dependent manner to regulate nodulation (Jeudy et al. 2010). We note that *AtCLE2*, the Arabidopsis gene most similar to the nodulation-suppressing CLE peptides, has a role in root development and is also induced by nitrate (Scheible et al. 2004; Araya et al. 2014).

Overexpression of the locally-acting *GmNIC1* in wild-type soybean reduces nodule numbers by ~50% compared with empty vector controls (Reid et al. 2011a). Although

significant, this suppressive ability is far from that of *GmRIC1* and *GmRIC2*, as discussed above. Confirmation is required to determine whether this is unique to soybean or is shared with the closely related orthologues identified in common bean (Ferguson et al. 2014).

2.6.3 Other inducing factors

Numerous factors can influence the extent of nodulation and it is possible that some do so by inducing, or otherwise influencing, the production, transport, perception or response to a CLE peptide(s). Recently, split-root and grafting studies using soybean grown in low pH conditions revealed a novel systemic mechanism that acts via GmNARK in the shoot to inhibit nodulation of the root (Lin et al. 2012; Ferguson et al. 2013). This suggests that soil acidity may act via a CLE peptide to suppress nodulation. Two CLE peptide-encoding genes of *L. japonicus*, called *LjCLE19* and *LjCLE20*, have been shown to be upregulated in the presence of phosphate (Funayama-Noguchi et al. 2011); however, a specific role for these peptides in plant development has not been reported. It has been noted that although CLE peptides are nearly-exclusive to plants, they also exist in plant-parasitic nematodes (e.g. Bakhetia et al. 2007), which appear to use the peptides to initiate the formation of feeding structures in host roots (reviewed by Mitchum et al. 2012). Also noted is that nematodes are easily genetically transformed through simple feeding, suggesting that perhaps nematode CLE genes were plant-derived. Whether nematodes, or any other pathogen, can also induce a plant-encoded CLE peptide(s) is of great interest to determine.

2.6.4 Rhizobia-induced CLE peptides that do not suppress nodulation

In addition to the nodulation-suppressing CLE peptides reported above, two further CLE peptide-encoding genes have been identified that are expressed in response to rhizobia inoculation, namely LjCLE3 (Okamoto et al. 2009) and MtCLE4 (Mortier et al. 2010). Overexpression of these genes does not alter the nodulation phenotype when compared with empty vector controls. It is now of interest to determine the role of these peptides in nodulation to determine why they are responsive to rhizobia inoculation and how they function in this symbiosis.

2.7 Future perspectives

In addition to the nodulation-suppressing CLE peptides reported above, two further CLE peptide-encoding genes have been identified that are expressed in response to rhizobia inoculation, namely *LjCLE3* (Okamoto et al. 2009) and *MtCLE4* (Mortier et al. 2010). Overexpression of these genes does not alter the nodulation phenotype when compared with empty vector controls. It is now of interest to determine the role of these peptides in nodulation to determine why they are responsive to rhizobia inoculation and how they function in this symbiosis.

2.8 Acknowledgements

We would like to thank the Australian Research Council for provision of a Centre of Excellence grant (CEO348212) and Discovery Project grants (DP130103084 and DP130102266), as well as the Hermon Slade Foundation and the University of Queensland for strategic funds. We thank Alina Tollenaere for technical assistance in developing the phylogenetic tree in Fig. 2.2.

2.9 References

Araya T, Miyamoto M, Wibowo J, Suzuki A, Kojima S, Tsuchiya YN, Sawa S, Fukuda H, Von Wirén N, Takahashi H. 2014. CLE-CLAVATA1 peptide-receptor signaling module regulates the expansion of plant root systems in a nitrogen-dependent manner. *Proceedings of the National Academy of Sciences* **111**(5): 2029-2034.

Bakhetia M, Urwin P, Atkinson HJ. 2007. qPCR analysis and RNAi define pharyngeal gland cell-expressed genes of *Heterodera glycines* required for initial interactions with the host. *Molecular Plant-Microbe Interactions* **20**(3): 306-312.

Carroll BJ, McNeil DL, Gresshoff PM. 1985a. A supernodulation and nitrate-tolerant symbiotic (nts) soybean mutant. *Plant Physiology* **78**(1): 34-40.

Carroll BJ, McNeil DL, Gresshoff PM. 1985. 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* **82**(12): 4162-4166.

Clark SE, Williams RW, Meyerowitz EM. 1997. The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in Arabidopsis. *Cell* **89**(4): 575-585.

Cock JM, McCormick S. 2001. A large family of genes that share homology with *CLAVATA3*. *Plant Physiology* **126**(3): 939-942.

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**(2): 588-590.

Desbrosses GJ, Stougaard J. 2011. Root nodulation: a paradigm for how plant-microbe symbiosis influences host developmental pathways. *Cell Host & Microbe* **10**(4): 348-358.

Djordjevic MA, Oakes M, Wong CE, Singh M, Bhalla P, Kusumawati L, Imin N. 2011. Border sequences of *Medicago truncatula* CLE36 are specifically cleaved by endoproteases common to the extracellular fluids of *Medicago* and soybean. *Journal of Experimental Botany* **62**(13): 4649-4659.

Erisman JW, Sutton MA, Galloway J, Klimont Z, Winiwarter W. 2008. How a century of ammonia synthesis changed the world. *Nature Geoscience* **1**(10): 636-639.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**(4): 783-791.

Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE, Gresshoff PM. 2010. Molecular analysis of legume nodule development and autoregulation. *Journal of Integrative Plant Biology* **52**(1): 61-76.

Ferguson BJ, Li D, Hastwell AH, Reid DE, Li Y, Jackson SA, Gresshoff PM. 2014. The soybean (*Glycine max*) nodulation-suppressive CLE peptide, GmRIC1, functions interspecifically in common white bean (*Phaseolus vulgaris*), but not in a supernodulating line mutated in the receptor PvNARK. *Plant Biotechnology Journal* **12**(8): 1085-1097.

Ferguson BJ, Mathesius U. 2014. Phytohormone regulation of legume-rhizobia interactions. *Journal of Chemical Ecology* **40**(7): 770-790.

Ferguson BJ, Lin M-H, Gresshoff PM. 2013. Regulation of legume nodulation by acidic growth conditions. *Plant Signaling & Behavior* **8**(3): e23426.

Fletcher JC, Brand U, Running MP, Simon R, Meyerowitz EM. 1999. Signaling of cell fate decisions by CLAVATA3 in Arabidopsis shoot meristems. *Science* **283**(5409): 1911-1914.

Funayama-Noguchi S, Noguchi K, Yoshida C, Kawaguchi M. 2011. Two CLE genes are induced by phosphate in roots of *Lotus japonicus*. *Journal of Plant Research* **124**(1): 155-163.

Gresshoff PM, Delves AC 1986. Plant genetic approaches to symbiotic nodulation and nitrogen fixation in legumes. In: Blaustein A, King P eds. *A Genetic Approach to Plant Biochemistry*. New York: Springer-Verlag, 159-206.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**(5): 696-704.

Hayashi S, Reid DE, Lorenc MT, Stiller J, Edwards D, Gresshoff PM, Ferguson BJ. 2012. Transient Nod factor-dependent gene expression in the nodulation-competent zone of soybean (*Glycine max* [L.] Merr.) roots. *Plant Biotechnology Journal* **10**(8): 995-1010.

Herridge DF, Peoples MB, Boddey RM. 2008. Global inputs of biological nitrogen fixation in agricultural systems. *Plant and Soil* **311**(1-2): 1-18.

Ito Y, Nakanomyo I, Motose H, Iwamoto K, Sawa S, Dohmae N, Fukuda H. 2006. Dodeca-CLE peptides as suppressors of plant stem cell differentiation. *Science* **313**(5788): 842-845.

Jacobsen E, Feenstra W. 1984. A new pea mutant with efficient nodulation in the presence of nitrate. *Plant Science Letters* **33**(3): 337-344.

Jensen ES, Peoples MB, Boddey RM, Gresshoff PM, Hauggaard-Nielsen H, Alves BJ, Morrison MJ. 2012. Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. A review. *Agronomy for Sustainable Development* **32**(2): 329-364.

Jeudy C, Ruffel S, Freixes S, Tillard P, Santoni AL, Morel S, Journet EP, Duc G, Gojon A, Lepetit M. 2010. Adaptation of *Medicago truncatula* to nitrogen limitation is modulated via local and systemic nodule developmental responses. *New Phytologist* **185**(3): 817-828.

Kondo T, Sawa S, Kinoshita A, Mizuno S, Kakimoto T, Fukuda H, Sakagami Y. 2006. A plant peptide encoded by CLV3 identified by in situ MALDI-TOF MS analysis. *Science* **313**(5788): 845-848.

Kondo T, Nakamura T, Yokomine K, Sakagami Y. 2008. Dual assay for MCLV3 activity reveals structure–activity relationship of CLE peptides. *Biochemical and Biophysical Research Communications* **377**(1): 312-316.

Kosslak RM, Bohlool BB. 1984. Suppression of nodule development of one side of a split-root system of soybeans caused by prior inoculation of the other side. *Plant Physiology* **75**(1): 125-130.

Krusell L, Madsen LH, Sato S, Aubert G, Genua A, Szczyglowski K, Duc G, Kaneko T, Tabata S, de Bruijn F. 2002. Shoot control of root development and nodulation is mediated by a receptor-like kinase. *Nature* **420**(6914): 422-426.

Krusell L, Sato N, Fukuhara I, Koch BE, Grossmann C, Okamoto S, Oka-Kira E, Otsubo Y, Aubert G, Nakagawa T. 2011. The *Clavata2* genes of pea and *Lotus japonicus* affect autoregulation of nodulation. *The Plant Journal* **65**(6): 861-871.

La Cour T, Kiemer L, Mølgaard A, Gupta R, Skriver K, Brunak S. 2004. Analysis and prediction of leucine-rich nuclear export signals. *Protein Engineering Design and Selection* **17**(6): 527-536.

Larkin MA, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R. 2007. Clustal W and Clustal X version 2.0. B*ioinformatics* **23**(21): 2947-2948.

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**(9): 955-967.

Lim CW, Lee YW, Hwang CH. 2011. Soybean nodule-enhanced CLE peptides in roots act as signals in GmNARK-mediated nodulation suppression. *Plant and Cell Physiology* **52**(9): 1613-1627.

Lin YH, 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**(4): 1074-1086.

Lin Y-H, Lin M-H, Gresshoff PM, Ferguson BJ. 2011. An efficient petiole-feeding bioassay for introducing aqueous solutions into dicotyledonous plants. *Nature Protocols* **6**(1): 36-45.

Lin M-H, Gresshoff PM, Ferguson BJ. 2012. Systemic regulation of soybean nodulation by acidic growth conditions. *Plant Physiology* **160**(4): 2028-2039.

Magori S, Kawaguch M. 2010. Analysis of two potential long-distance signaling molecules, LjCLE-RS1/2 and jasmonic acid, in a hypernodulating mutant too much love. *Plant Signaling & Behavior* **5**(4): 403-405.

Matsubayashi Y. 2014. Posttranslationally modified small-peptide signals in plants. *Annual Review of Plant Biology* **65**: 385-413.

Meng L, Ruth KC, Fletcher JC, Feldman L. 2010. The roles of different CLE domains in Arabidopsis CLE polypeptide activity and functional specificity. *Molecular Plant* **3**(4): 760-772.

Mitchum MG, Wang X, Wang J, Davis EL. 2012. Role of nematode peptides and other small molecules in plant parasitism. *Annual Review of Phytopathology* **50**: 175-195.

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**(37): 25381-25391.

Miyazawa H, Oka-Kira E, Sato N, Takahashi H, Wu G-J, Sato S, Hayashi M, Betsuyaku S, Nakazono M, Tabata S. 2010. The receptor-like kinase KLAVIER mediates systemic regulation of nodulation and non-symbiotic shoot development in *Lotus japonicus*. *Development* **137**(24): 4317-4325.

Mortier V, Den Herder G, Whitford R, Van de Velde W, Rombauts S, D'haeseleer K, Holsters M, Goormachtig S. 2010. CLE peptides control *Medicago truncatula* nodulation locally and systemically. *Plant Physiology* **153**(1): 222-237.

Mortier V, De Wever E, Vuylsteke M, Holsters M, Goormachtig S. 2012. Nodule numbers are governed by interaction between CLE peptides and cytokinin signaling. *The Plant Journal* **70**(3): 367-376.

Ni J, Clark SE. 2006. Evidence for functional conservation, sufficiency, and proteolytic processing of the CLAVATA3 CLE domain. *Plant Physiology* **140**(2): 726-733.

Ni J, Guo Y, Jin H, Hartsell J, Clark SE. 2011. Characterization of a CLE processing activity. *Plant Molecular Biology* **75**(1-2): 67-75.

Nishimura R, Hayashi M, Wu G-J, Kouchi H, Imaizumi-Anraku H, Murakami Y, Kawasaki S, Akao S, Ohmori M, Nagasawa M. 2002. HAR1 mediates systemic regulation of symbiotic organ development. *Nature* **420**(6914): 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**(7): 769-780.

Oelkers K, Goffard N, Weiller GF, Gresshoff PM, Mathesius U, Frickey T. 2008. Bioinformatic analysis of the CLE signaling peptide family. *BMC Plant Biology* **8**(1): 1.

Ogawa M, Shinohara H, Sakagami Y, Matsubayashi Y. 2008. Arabidopsis CLV3 peptide directly binds CLV1 ectodomain. *Science* **319**(5861): 294-294.

Ogawa-Ohnishi M, Matsushita W, Matsubayashi Y. 2013. Identification of three hydroxyproline O-arabinosyltransferases in *Arabidopsis thaliana*. *Nature Chemical Biology* **9**(11): 726-730.

Ohyama K, Shinohara H, Ogawa-Ohnishi M, Matsubayashi Y. 2009. A glycopeptide regulating stem cell fate in *Arabidopsis thaliana*. *Nature Chemical Biology* **5**(8): 578-580.

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**(1): 67-77.

Okamoto S, Nakagawa T, Kawaguchi M. 2011. Expression and functional analysis of a CLV3-like gene in the model legume Lotus japonicus. *Plant and Cell Physiology* **52**(7): 1211-1221.

Okamoto S, Shinohara H, Mori T, Matsubayashi Y, Kawaguchi M. 2013. Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. *Nature Communications* **4**: 2191.

Oldroyd GE. 2013. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews Microbiology* **11**(4): 252-263.

Osipova MA, Mortier V, Demchenko KN, Tsyganov VE, Tikhonovich IA, Lutova LA, Dolgikh EA, Goormachtig S. 2012. *WUSCHEL-RELATED HOMEOBOX5* gene expression and interaction of CLE peptides with components of the systemic control add two pieces to the puzzle of autoregulation of nodulation. *Plant Physiology* **158**(3): 1329-1341.

Postma J, Jacobsen E, Feenstra W. 1988. Three pea mutants with an altered nodulation studied by genetic analysis and grafting. *Journal of Plant Physiology* **132**(4): 424-430.

Reid DE, Ferguson BJ, Gresshoff PM. 2011a. Inoculation-and nitrate-induced CLE peptides of soybean control NARK-dependent nodule formation. *Molecular Plant-Microbe Interactions* **24**(5): 606-618.

Reid DE, Ferguson BJ, Hayashi S, Lin Y-H, Gresshoff PM. 2011b. Molecular mechanisms controlling legume autoregulation of nodulation. *Annals of Botany* **108**(5): 789-795.

Reid DE, Hayashi S, Lorenc M, Stiller J, Edwards D, Gresshoff PM, Ferguson BJ. 2012. Identification of systemic responses in soybean nodulation by xylem sap feeding and complete transcriptome sequencing reveal a novel component of the autoregulation pathway. *Plant Biotechnology Journal* **10**(6): 680-689.

Reid DE, Li D, Ferguson BJ, Gresshoff PM. 2013. Structure–function analysis of the Gm RIC1 signal peptide and CLE domain required for nodulation control in soybean. *Journal of Experimental Botany* **64**(6): 1575-1585.

Rojo E, Sharma VK, Kovaleva V, Raikhel NV, Fletcher JC. 2002. CLV3 is localized to the extracellular space, where it activates the Arabidopsis CLAVATA stem cell signaling pathway. *The Plant Cell* **14**(5): 969-977.

Sagan M, Duc G. 1996. *Sym28* and *Sym29*, two new genes involved in regulation of nodulation in pea (*Pisum sativum* L.). *Symbiosis* **20**(3): 229-245.

Sasaki T, Suzaki T, Soyano T, Kojima M, Sakakibara H, Kawaguchi M. 2014. Shoot-derived cytokinins systemically regulate root nodulation. *Nature Communications* **5**: 4983.

Saur IM, Oakes M, Djordjevic MA, Imin N. 2011. Crosstalk between the nodulation signaling pathway and the autoregulation of nodulation in *Medicago truncatula*. *New Phytologist* **190**(4): 865-874.

Sawa S, Kinoshita A, Nakanomyo I, Fukuda H. 2006. CLV3/ESR-related (CLE) peptides as intercellular signaling molecules in plants. *The Chemical Record* **6**(6): 303-310.

Scheible W-R, Morcuende R, Czechowski T, Fritz C, Osuna D, Palacios-Rojas N, Schindelasch D, Thimm O, Udvardi MK, Stitt M. 2004. Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of Arabidopsis in response to nitrogen. *Plant Physiology* **136**(1): 2483-2499.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J. 2010. Genome sequence of the palaeopolyploid soybean. N*ature* **463**(7278): 178-183.

Schnabel E, Journet E, 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**(6): 809-822

Schnabel E, Kassaw T, Smith L, Marsh J, Oldroyd GE, Long SR, Frugoli J. 2011. The *ROOT DETERMINED NODULATION 1* gene regulates nodule number in roots of *M. truncatula* and defines a highly conserved, uncharacterized plant gene family. *Plant Physiology*: **157**(1): 328-340.

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**(5603): 109-112.

Shinohara H, Matsubayashi Y. 2013. Chemical synthesis of Arabidopsis CLV3 glycopeptide reveals the impact of hydroxyproline arabinosylation on peptide conformation and activity. *Plant and Cell Physiology* **54**(3): 369-374.

Shinohara H, Moriyama Y, Ohyama K, Matsubayashi Y. 2012. Biochemical mapping of a ligand-binding domain within Arabidopsis BAM1 reveals diversified ligand recognition mechanisms of plant LRR-RKs. *The Plant Journal* **70**(5): 845-854.

Song X-F, Guo P, Ren S-C, Xu T-T, Liu C-M. 2013. Antagonistic peptide technology for functional dissection of CLV3/ESR genes in Arabidopsis. *Plant Physiology* **161**(3): 1076-1085.

Soyano T, Hirakawa H, Sato S, Hayashi M, Kawaguchi M. 2014. NODULE INCEPTION creates a long-distance negative feedback loop involved in homeostatic regulation of nodule organ production. *Proceedings of the National Academy of Sciences* **111**(40): 14607-14612.

Stefanović S, Pfeil BE, Palmer JD, Doyle JJ. 2009. Relationships among phaseoloid legumes based on sequences from eight chloroplast regions. *Systematic Botany* **34**(1): 115-128.

Strabala TJ, O'Donnell PJ, Smit A-M, Ampomah-Dwamena C, Martin EJ, Netzler N, Nieuwenhuizen NJ, Quinn BD, Foote HC, Hudson KR. 2006. Gain-of-function phenotypes of many CLAVATA3/ESR genes, including four new family members, correlate with tandem variations in the conserved CLAVATA3/ESR domain. *Plant Physiology* **140**(4): 1331-1344.

Sutton MA, Oenema O, Erisman JW, Leip A, van Grinsven H, Winiwarter W. 2011. Too much of a good thing. *Nature* **472**(7342): 159-161.

Takahara M, Magori S, Soyano T, Okamoto S, Yoshida C, Yano K, Sato S, Tabata S, Yamaguchi K, Shigenobu S. 2013. TOO MUCH LOVE, a novel kelch repeat-containing F-box protein, functions in the long-distance regulation of the legume–*Rhizobium* symbiosis. *Plant and Cell Physiology* **54**(4): 433-447.

Timmers AC, Auriac MC, Truchet G. 1999. Refined analysis of early symbiotic steps of the *Rhizobium-Medicago* interaction in relationship with microtubular cytoskeleton rearrangements. *Development* **126**(16): 3617-3628.

Udvardi MK, Price GD, Gresshoff PM, Day DA. 1988. A dicarboxylate transporter on the peribacteroid membrane of soybean nodules. *FEBS Letters* **231**(1): 36-40.

Wong CE, Singh MB, Bhalla PL. 2013. Spatial expression of *CLAVATA3* in the shoot apical meristem suggests it is not a stem cell marker in soybean. *Journal of Experimental Botany* **64**(18): 5641-5649.

Xu T-T, Song X-F, Ren S-C, Liu C-M. 2013. The sequence flanking the N-terminus of the CLV3 peptide is critical for its cleavage and activity in stem cell regulation in Arabidopsis. *BMC Plant Biology* **13**(1): 225.

Chapter 3

Genome-wide annotation and characterization of CLAVATA/ESR (CLE) peptide hormones of soybean (*Glycine max*) and common bean (*Phaseolus vulgaris*), and their orthologues of *Arabidopsis thaliana*.

Preface

This chapter identifies and characterises the family of CLE peptide-encoding genes in soybean and common bean and has been published in the Journal of Experimental Botany (2015, DOI: 10.1093/jxb/erv351). The bioinformatic resources developed in this article can be used by subsequent studies aimed at characterising individual CLE peptide-encoding genes, as demonstrated by Corcilius et al. (2017) in Chapter 5 and Ferguson et al. (2014) in Chapter 7. Additional supplementary tables and figures may be found online at http://jxb.oxfordjournals.org/content/66/17/5271.abstract

3.1 Abstract

CLE peptides are key regulators of cell proliferation and differentiation in plant shoots, roots, vasculature, and legume nodules. They are C-terminally encoded peptides that are post-translationally cleaved and modified from their corresponding pre-propeptides to produce a final ligand that is 12–13 amino acids in length. In this study, an array of bionformatic and comparative genomic approaches was used to identify and characterize the complete family of CLE peptide-encoding genes in two of the world's most important crop species, soybean and common bean. In total, there are 84 CLE peptide-encoding genes in soybean (considerably more than the 32 present in Arabidopsis), including three pseudogenes and two multi-CLE domain genes having six putative CLE domains each. In addition, 44 CLE peptide-encoding genes were identified in common bean. *In silico* characterization was used to establish all soybean homeologous pairs, and to identify corresponding gene orthologues present in common bean and Arabidopsis. The soybean CLE pre-propeptide family was further analysed and separated into seven distinct groups based on structure, with groupings strongly associated with the CLE domain sequence and function. These groups provide evolutionary insight into the CLE peptide families of soybean, common bean, and Arabidopsis, and represent a novel tool that can aid in the functional characterization of the peptides. Transcriptional evidence was also used to provide further insight into the location and function of all CLE peptide-encoding members currently available in gene atlases for the three species. Taken together, this in-depth analysis helped to identify and categorize the complete CLE peptide families of soybean and common bean, established gene orthologues within the two legume species, and Arabidopsis, and provided a platform to help compare, contrast, and identify the function of critical CLE peptide hormones in plant development.

3.2 Introduction

CLAVATA/embryo surrounding region (ESR) peptide hormones (CLE peptides) are a group of post-translationally modified signal molecules involved in the regulation and differentiation of meristematic plant tissues. They have been shown to control cell divisions in the shoot apical meristem (SAM), root apical meristem (RAM), vasculature, and legume nodules (Matsubayashi 2014; Ferguson and Mathesius 2014; Grienenberger and Fletcher 2015; Hastwell et al. 2015). They arise from a structurally conserved gene family and are named after the first identified CLE peptide (AtCLV3 in *Arabidopsis thaliana*; Fletcher et al. 1999), and the structurally and functionally similar, but unrelated, ESR peptides (first identified in *Zea mays*; Opsahl-Ferstad et al. 1997; Cock and McCormick 2001).

Mature CLE peptides are typically 12–13 amino acids in length and are located at or near the C-terminus of their pre-propeptide. CLE pre-propeptides are cysteine-poor and have a tripartite domain structure, consisting of an N-terminal signal peptide, a central variable domain, and a highly conserved and functional CLE peptide domain (Matsubayashi 2014; Hastwell et al. 2015). Some also have a fourth domain, called a C-terminal extension, which is not highly conserved, except between orthologous genes. Multi-CLE domain-containing pre-propeptides have also been identified in several plant species (Kinoshita et al. 2007; Oelkers et al. 2008), but little is known about their processing in plants. There is also a group of CLE-Like (CLEL) peptides, whose functional domain shares a similar structure but exhibits unrelated activity (Meng et al. 2012). Interestingly, one gene identified in Arabidopsis (*AtCLE18*) contains both a CLE and a CLEL domain (Meng et al. 2012).

The mature CLE peptide ligand is post-translationally cleaved and modified from its pre-propeptide. Hydroxylatation of proline residues is common, with one central hydroxyproline having a tri-arabinose moiety attached (Matsubayashi 2014); however, it is important to note that all arabinose post-translational modifications identified in plants to date are limited to three peptides in *A. thaliana* (AtCLV3, AtCLE2, and AtCLE9) and one in *Lotus japonicus* (LjCLE-RS2) (Ohyama et al. 2009; Okamoto et al. 2013; Shinohara and Matsubayashi 2013; Matsubayashi 2014). Mature CLE peptides are ligands for leucine-rich repeat receptor kinases (LRR-RKs), with the first identified ligand receptor pair being CLV3 and CLV1 of Arabidopsis (Fletcher et al. 1999), which has since expanded to include a number of additional binding partners and associated factors (Shinohara and

Matsubayashi 2015). A comprehensive list of putative CLE ligand–LRR-RK pairs was recently presented (Endo et al. 2014).

The role of many CLE peptides remains unknown, with the majority that have been functionally characterized found in Arabidopsis. The most widely studied is AtCLV3, which acts in the SAM to regulate stem cell numbers (Fletcher et al. 1999; Gaillochet et al. 2015). Additional Arabidopsis CLE peptides acting in the root have also been characterized, including AtCLE40 (Hobe et al. 2003; Sharma et al. 2003; Stahl et al. 2009), which regulates cell proliferation in the RAM as part of a mechanism mirroring that acting in the SAM (van der Graff et al. 2009). Other root-acting CLE peptides of Arabidopsis include AtCLE1, 2, 3, 4, and 7, which are involved in nitrate-responsive mechanisms, with some also involved in lateral root development (Scheible et al. 2004; Araya et al. 2014). Additional CLE peptide-encoding genes involved in cell proliferation and differentiation include *AtCLE8*, which acts in embryogenesis (Fiume and Fletcher 2012), and *AtCLE45*, which has been implicated in both root protophloem and pollen development (Depuydt et al. 2013; Endo et al. 2013; Rodriguez-Villalon et al. 2014). Three CLE peptides, known as tracheary element differentiation factors (TDIFs), control vascular meristematic tissue proliferation and differentiation (encoded by *AtCLE41*, *AtCLE42*, and *AtCLE44*; Sawa et al. 2006; Ito et al. 2006; Hirakawa et al. 2010). This group has the highest conservation amongst gymnosperms and angiosperms (Strabala et al. 2014), and consists of the only CLE peptides to begin with a histidine, rather than the archetypical arginine residue that is characteristic of all other CLE peptides (with the sole exception of AtCLE46, whose CLE domain begins with a histidine, and whose function remains unknown; Hirakawa et al. 2011).

In addition to those identified in Arabidopsis, a number of CLE peptides have been identified in various legume species. This includes CLE peptides acting to control the highly important nodulation process, which is a symbiotic relationship legumes enter into with nitrogen-fixing rhizobia bacteria (Okamoto et al. 2009, 2013; Mortier et al. 2010, 2012; Reid et al. 2011*a*, 2013; Ferguson et al. 2014; reviewed in Hastwell et al. 2015). By regulating nodulation, these CLE peptides essentially enable the host plant to balance nitrogen uptake from the bacteria with resource allocation to form and maintain nodules (Ferguson et al. 2010). Prominent pathways involved in this regulation are the systemic autoregulation of nodulation (AON) and the local nitrogen regulation pathways, both of which commence with the induction of CLE peptide signals (reviewed in Ferguson et al. 2010; Reid et al. 2011*b*). Similarly, a number of legume CLE peptides have also been

shown to respond to phosphate application (Funayama-Noguchi et al. 2011) and more recently mycorrhiza infection (Handa et al. 2015).

Aside from plants, cyst nematodes are the only other known organism to have CLE peptide-encoding genes (Mitchum et al. 2013). These genes have multiple CLE domains that are processed into a single mature peptide ligand (Chen et al. 2015). The peptides are thought to assist in nematode infection, possibly by manipulating the host to gain entry into the plant (Olsen and Skriver 2003; Wang et al. 2005; reviewed in Mitchum et al. 2013). They are post-translationally modified and processed by the host plant's machinery, and are perceived by plant receptors (Replogle et al. 2011; Chen et al. 2015), suggesting that they may have evolved through horizontal gene transfer.

Here, advantage was taken of recent advances in genomics and bioinformatics to identify, categorize, and functionally characterize the highly important CLE peptide families of soybean and common bean, two agriculturally important crop species. Soybean and common bean share a common ancestor whose genome duplicated ~59 million years ago (MYA), from which soybean subsequently diverged (19 MYA) and duplicated again 13 MYA (Lavin et al. 2005; Schmutz et al. 2010, 2014). As a result, 75% of soybean genes have more than one copy across the genome (a homeologous or duplicate copy; Schmutz et al. 2010, 2014; Roulin et al. 2013), whereas common bean does not. Indeed, for these reasons, soybean and common bean are commonly used for comparative and evolutionary studies in genomics and genetics (e.g. McClean et al. 2008; Lin et al. 2010; Ferguson et al. 2014; Schmutz et al. 2014).

The present investigations identified a total of 84 CLE peptide-encoding genes in soybean and 44 in common bean. In-depth sequence analyses enabled the identification of all homeologous copies within soybean, in addition to all orthologous copies existing between soybean, common bean, and Arabidopsis. Transcriptional analysis of all CLE peptide-encoding genes available in gene atlases of soybean, common bean, and Arabidopsis were evaluated to provide further insight into the localization and function of the genes. Moreover, using the complete family in soybean, seven distinct CLE peptide groups were defined based on both sequence similarity and phylogenetic analysis, with consensus sequences subsequently derived for each. Collectively, the findings provide new insight into the sequence, structure, and evolution of critical CLE peptide hormones of plants.

3.3 Materials and methods

3.3.1 Gene identification

To identify CLE peptide-encoding genes, multiple TBLASTN and BLASTN searches using known soybean sequences were conducted in Phytozome against the *Glycine max Wm82.a2.v1* and *Phaseolus vulgaris v1.0* genomes (http://www.phytozome.net/; Schmutz et al. 2010, 2014; Goodstein et al. 2012). Searches were conducted using less stringent parameters [expected threshold (E)=10] to enhance the identification of genes of interest. Results were then manually validated to confirm the presence of a CLE domain in an open reading frame. Subsequent searches based on the preliminary findings were performed using BLASTN to identify additional genes, including common bean orthologues and soybean duplicates, particularly where no duplicate/orthologue was identified in the initial queries. These subsequent searches were conducted using a slightly more stringent parameter of E=1. The open reading frames of homologous chromosome regions were also examined for potential unannotated or truncated duplicates. Additional BLASTP searches of mycorrhizal (http://genome.jgi.doe.gov/) and rhizobia genomes (Rhizobase; http://genome.microbedb.jp/rhizobase; Fujisawa et al. 2014), using both whole CLE pre-propeptide sequences and also CLE domain consensus sequences from soybean, were also performed using very low stringency (E=100) to identify CLE peptide encoding genes in these species.

3.3.2 Genomic environments

Synteny between genomic environments was individually obtained for each gene of interest. This was achieved using Phytozome JBrowse of the *Glycine max Wm82.a2.v1*, *Phaseolus vulgaris v1.0*, *Arabidopsis thaliana TAIR10*, *Oryza sativa v7.0* and *Medicago truncatula Mt4.0v1* genomes (http://www.phytozome.net/; Ouyang et al. 2007; Schmutz et al. 2010, 2014; Young et al. 2011; Goodstein et al. 2012; Lamesch et al. 2012). For each genomic environment investigated, the five genes located directly up- and downstream of the gene of interest were assessed for their orientation, gene family, and predicted homologues.

3.3.3 Sequence characterization

Clustal Omega, hosted on EMBL-EBI (http://www.ebi.ac.uk/Tools/msa/clustalo/), was used to generate multiple sequence alignments (Goujon et al. 2010; Sievers et al. 2011; McWilliam et al. 2013). Manual adjustments were subsequently made to some of the sequences predicted in Phytozome, particularly in regards to their start codon. This was based on sequence similarity to duplicate genes, similarly clustering genes, and/or likely orthologous genes, in addition to signal peptide domain prediction results.

Logo diagrams used to define consensus sequences were obtained using multiple sequence alignments for each CLE peptide group (I–VII) in Geneious Pro v6.1.8 (Kearse et al. 2012). Signal peptides were identified using the SignalP prediction program v4.1 (http://www.cbs.dtu.dk/services/SignalP/; Petersen et al. 2011). Hydrophobicity values were determined from amino acid scale values on ProtScale (http://web.expasy.org/protscale/; Gasteiger et al. 2005) using the Kyte and Doolittle (1982) hydrophobicity scale.

3.3.4 Phylogenetic analyses

Phylogenetic trees were constructed from multiple sequence alignments using the PHYML plugin in Geneious Pro v6.1.8 (Guindon and Gascuel 2003). They were derived using the maximum likelihood approach with 1000 bootstraps to support a branch, with the exception of the tree designed using all soybean, common bean, and Arabidopsis sequences, where 100 bootstraps were used. Multiple trees were constructed to identify homeologous soybean genes. Those appearing to lack a homeologous copy were identified and used to re-search the genome for a potential duplicate. All trees presented here include each distinct gene identified in the numerous searches made. A similar approach was used to identify all soybean gene orthologues in common bean and Arabidopsis.

3.3.5 Meta-analyses of transcriptome data

Transcriptional data for the meta-analysis was collected from publicly available data sets from the Soybean RNA-Seq Atlas (http://www.soybase.org/soyseq/; Severin et al. 2010); the Soybean eFP Browser (http://bar.utoronto.ca/efpsoybean/cgi-bin/efpWeb.cgi; Libault et

al. 2010*a*, b); A Common Bean Gene Expression Atlas (http://plantgrn.noble.org/PvGEA/index.jsp; Jamie et al. 2014); and the Arabidopsis eFP Browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi; Schmid et al. 2005). The entire list of gene identifiers for each species was searched in their respective databases, and only those with transcriptional data are presented. Normalized RPKM (reads per kilobase per million) values were taken where possible.

3.4 Results

3.4.1 Identification of CLE peptide-encoding genes in soybean and common bean, in addition to mycorrhiza and rhizobia species

To identify CLE peptide-encoding genes in soybean and common bean, a genome-wide analysis was performed involving multiple BLAST queries, followed by manual validation and the removal of false positives (i.e. no CLE domain). This resulted in the identification of 84 distinct soybean genes and 44 distinct common bean genes (Figs 3.1, 3.2; Tables 3.1, 3.2). BLAST queries were based on all known soybean CLE genes, and some Arabidopsis genes, and involved searching with both pre-propeptide and CLE domain sequences to enhance the likelihood of detecting all CLE peptide-encoding genes in the two genomes.

The identified genes are scattered across the genomes, with at least one located on every chromosome, except for chromosome 10 of common bean. Chromosome 13 of soybean contains the most CLE peptide-encoding genes, with a total of 12. Most of the identified genes lack predicted introns, with the exception of 12 soybean genes and nine common bean genes (Tables 3.1, 3.2).

Many of the genes identified here had not been discovered previously and therefore had not yet been assigned a name. In contrast, those which were previously reported had as many as five different aliases. To unify the nomenclature, designations were assigned based on the names of all previously characterized soybean CLE peptides (e.g. Cock and McCormick, 2001; Reid et al. 2011*a*; Wong et al. 2013), and the Arabidopsis phylogenetic approach was used for all non-characterized genes (Cock and McCormick 2001). The duplicated nature of the soybean genome was also accounted for by identifying *a* and *b* copies of homeologous gene pairs (described below). In common bean, the gene names

were assigned based on their orthologue in soybean (Table 3.1; Supplementary Fig. S1 available at *JXB* online). A comprehensive list of all soybean and common bean names, including all previous identifiers, is provided in Supplementary Table S1.

Aside from plants, cyst nematodes are the only known organisms to possess CLE peptide-encoding genes (Mitchum et al. 2013). These peptides appear to assist in parasitism of the host. To determine whether mutualistic symbiotic organisms also encode for CLE peptides that assist in infection, a protein search of mycorrhiza (http://genome.jgi.doe.gov/) and rhizobia (Rhizobase; http://genome.microbedb.jp/rhizobase; Fujisawa et al. 2014) species was conducted using CLE domain consensus sequences and also pre-propeptide sequences. This thorough search yielded the identification of no CLE peptide-encoding genes in these organisms.

3.4.2 Identification of homeologues and orthologues in soybean and common bean

To characterize their amino acid sequences, all identified CLE peptide-encoding genes were translated and successive multiple sequence alignments were conducted using entire CLE pre-propeptide sequences. Despite having large variable domains, the pre-propeptides grouped strongly according to their CLE domain sequence in both soybean (Fig. 3.1) and common bean (Fig. 3.2). This helped in identifying likely homeologous (duplicate) copies of genes in the palaeopolyploid genome of soybean, with 39 pairs identified compared with only six genes having no duplicate (Fig. 3.1; Table 3.1). The six genes lacking a duplicate were re-blasted against the soybean genome to confirm their lack of a duplicate, and their homeologous chromosome region was checked for unannotated genes. The presence of a common bean orthologue confirmed they were not triplicated within the soybean genome.

To identify likely orthologues between soybean and common bean, an additional multiple sequence alignment was produced using the CLE peptide-encoding gene families of both species (data not shown). This alignment was also useful in confirming the 39 homeologous gene pairs of soybean. As expected, all previously reported gene orthologues of soybean and common bean clustered together (e.g. RIC, NIC; Ferguson et al. 2014). Additional orthologue candidates also clustered; however, soybean has four homeologous gene pairs and one individual gene lacking an apparent duplicate that appear to have no orthologue in common bean (*GmCLE2a* and *b*; *GmCLE31a* and *b*; *GmCLE32*; *GmCLE36a* and *b*; and *GmCLE37a* and *b*; Table 3.1).

Table 3.1 Features of the soybean (*Glycine max*) CLE genes

Listed are the genetic location, pre-propeptide length, predicted intron presence, gene orientation, soybean and common bean homologue, pre-propeptide similarity (%). and SignalP signal peptide (SP) cleavage site.

Name	Chromosome location	Orientation	Pre-propeptide length[a]	Predicted intron	SP cleavage site[b]	Homeologue similarity (%)	Common bean orthologue	Soybean and common bean pairwise identity (%)
GmCLE1a	Chr11:10740675..10741635	Reverse	84	Y	23	82.1	PvCLE1	74.6
GmCLE1b	Chr12:4724973..4727049	Reverse	83	Y	23			
GmCLE2a	Chr20:46634836..46635799	Reverse	76	N	30	92.1	–	–
GmCLE2b	Chr10:38974407..38975417	Forward	74	N	28			
GmCLE3a	Chr03:43793053..43794104	Forward	81	N	27	89.5	PvCLE3	80.2
GmCLE3b	Chr19:48528559..48529545	Forward	75	N	27			
GmCLE4a	Chr01:53094482..53095085	Forward	67	N	21	92.5	PvCLE4	82.6
GmCLE4b	Chr11:3319115..3320325	Reverse	67	N	21			
GmCLE5	Chr08:46805591..46806636	Reverse	99	N	25	-	PvCLE5	69.9
GmCLE6a	Chr20:35756760..35757955	Reverse	97	N	26	91.8	PvCLE6	76.3
GmCLE6b	Chr10:49704427..49706416	Forward	96	N	26			
GmCLE7a	Chr01:5559528..5560353	Forward	108	N	23	89.8	PvCLE7	85.8
GmCLE7b	Chr02:10245905..10246706	Reverse	108	N	23			
GmCLE8a	Chr06:17294801..17295629	Reverse	96	N	21	83.9	PvCLE8	85.4
GmCLE8b	Chr04:42380768..42381923	Forward	95	N	28			
GmCLE9a	Chr05:2299498..2299782	Forward	79	Y	19	93.8	PvCLE9	80.3
GmCLE9b	Chr17:7902958..7904070	Reverse	79	Y	19			
GmCLE10a	Chr01:4182744..4185349	Reverse	108	Y	42	83.3	PvCLE10	79.7
GmCLE10b	Chr02:2311001..2311717	Forward	102	Y	40			
GmCLE11a	Chr14:7781256..7782013	Reverse	82	N	27	89.3	PvCLE11	65.4
GmCLE11b	Chr17:39269471..39270222	Forward	84	N	27			
GmCLE12a	Chr13:16671710..16673786	Forward	97	Y	34	94.8	PvCLE12	93.1
GmCLE12b	Chr19:1819967..1821863	Reverse	97	Y	34			

Name	Chromosome location	Orientation	Pre-propeptide length[a]	Predicted intron	SP cleavage site[b]	Homeologue similarity (%)	Common bean orthologue	Soybean and common bean pairwise identity (%)
GmCLE13	Chr13:36676213..36676962	Forward	86	Y	24	–	PvCLE13	73.8
GmCLE14	Chr10:46589943..46590137	Forward	83	N	25	–	PvCLE14	72.7
GmCLE15a	Chr10:46586624..46587350	Forward	86	N	25	51.1	PvCLE15a, PvCLE15b, PvCLE15c, PvCLE15d	48.3, 47.9, 45.4, 45.6
GmCLE15b	Chr06:27528956..27529216	Forward	86	N	26			
GmCLE16a	Chr09:34804635..34806006	Forward	86	N	27	90.7	PvCLE16	85.7
GmCLE16b	Chr16:35643819..35644747	Forward	86	N	27			
GmCLE17a	Chr05:38846465..38847260	Reverse	87	N	28	86.2	PvCLE17	85.1
GmCLE17b	Chr08:969117..970012	Reverse	87	N	24			
GmCLE18a	Chr13:21801637..21802409	Forward	85	N	19	85.9	PvCLE18	80.8
GmCLE18b	Chr17:4258185..4258436	Reverse	83	N	19			
GmCLE19a	Chr07:39333907..39334972	Forward	119	N	32	83.9	PvCLE19	67.0
GmCLE19b	Chr20:1750676..1751787	Forward	114	N	32			
GmCLE20a	Chr03:33954213..33955592	Forward	100	N	36	91.0	PvCLE20	78.9
GmCLE20b	Chr19:38764138..38765477	Forward	94	N	31			
GmCLE21a	Chr02:46067116..46071548	Forward	81	N	26	88.9	PvCLE21	75.4
GmCLE21b	Chr14:2730030..2731670	Reverse	80	N	26			
GmCLE22a	Chr07:41652868..41653137	Reverse	89	N	27	91.0	PvCLE22	74.0
GmCLE22b	Chr20:7721313..7721576	Reverse	87	N	27			
GmCLE23a	Chr02:45459965..45460989	Reverse	73	N	23	85.9	PvCLE23	79.0
GmCLE23b	Chr14:3533265..3534446	Forward	71	N	21			
GmCLE24a	Chr10:43660111..43661108	Forward	110	N	23	88.4	PvCLE24	82.9
GmCLE24b	Chr20:42379994..42380805	Reverse	111	N	23			
GmCLE25a	Chr05:1295698..1296578	Forward	118	N	29	80.0	PvCLE25	68.8
GmCLE25b	Chr17:9746590..9748712	Forward	114	N	29			
GmCLE26	Chr20:2984627..2986271	Forward	99	N	27	–	PvCLE26	52.6
GmCLE27a	Chr02:11156483..11156827	Reverse	114	N	30	83.3	PvCLE27	78.6

Name	Chromosome location	Orientation	Pre-propeptide length[a]	Predicted intron	SP cleavage site[b]	Homeologue similarity (%)	Common bean orthologue	Soybean and common bean pairwise identity (%)
GmCLE27b	Chr01:7300791..7302992	Reverse	107	N	30			
GmCLE28a	Chr13:37349043..37349282	Reverse	83	N	27	80.3	PvCLE28	69.4
GmCLE28b	Chr12:38835186..38835383	Reverse	65	N	26			
GmCLE29a	Chr12:27615321..27615566	Forward	82	N	26	92.8	PvCLE29	84.3
GmCLE29b	Chr06:36330866..36331117	Reverse	83	N	26			
GmCLE30a	Chr06:36324860..36325095	Reverse	78	N	22	61.5	PvCLE30	60.5
GmCLE30b	Chr06:36255159..36255402	Reverse	81	N	26			
GmCLE31a	Chr07:37351348..37351668	Forward	106	N	22	92.5	–	–
GmCLE31b	Chr13:28570341..28570661	Reverse	106	N	22			
GmCLE32	Chr13:28559073..28559703	Reverse	68	N	23	–	–	–
GmCLE33a	Chr06:36402219..36402452	Reverse	78	N	23	84.4	PvCLE33	66.3
GmCLE33b	Chr12:27380684..27380911	Forward	76	N	24			
GmCLE34a	Chr12:38840660..38840902	Reverse	81	N	22	88.9	PvCLE34	78.6
GmCLE34b	Chr13:37353930..37354172	Reverse	81	N	22			
GmCLE35	Chr13:28564185..28564418	Reverse	78	N	23	–	PvCLE35	70.5
GmCLE36a	Chr13:34350525..34350935	Reverse	76	N	24	83.1	–	–
GmCLE36b	Chr15:6162182..6162415	Forward	77	N	25			
GmCLE37a	Chr16:4533525..4534140	Forward	185	Y	18	40.8	–	–
GmCLE37b	Chr19:35239153..35240209	Reverse	190	Y	24			
GmCLE40a	Chr12:3979297..3980162	Forward	82	Y	23	40.0	PvCLE40	47.9
GmCLE40b	Chr11:9961342..9961800	Forward	35	N	-			
GmCLV3a	Chr12:34902722..34903650	Forward	105	Y	28	93.3	PvCLV3	91.1
GmCLV3b	Chr13:40867356..40867942	Reverse	105	Y	29			
GmNIC1a	Chr12:36837550..36838464	Forward	80	N	22	86.3	PvNIC1	75.9
GmNIC1b	Chr13:39224711..39225630	Reverse	79	N	22			
GmRIC1a	Chr13:39215403..39216108	Reverse	95	N	28	77.3	PvRIC1	68.8
GmRIC1b	Chr12:36848528..36849475	Forward	96	N	27			
GmRIC2a	Chr06:47247215..47248215	Reverse	93	N	26	87.2	PvRIC2	74.5

Name	Chromosome location	Orientation	Pre-propeptide length[a]	Predicted intron	SP cleavage site[b]	Homeologue similarity (%)	Common bean orthologue	Soybean and common bean pairwise identity (%)
GmRIC2b	Chr12:13187190..13187511	Forward	94	N	26			
GmTDIF1a	Chr07:41652868..41653137	Reverse	104	N	42	92.4	PvTDIF1	82.5
GmTDIF1b	Chr18:40563162..40564249	Reverse	104	N	41			
GmTDIF2a	Chr05:32724420..32724761	Reverse	113	N	28	92.2	PvTIDF2	87.9
GmTDIF2b	Chr08:6781787..6783296	Reverse	113	N	28			
GmTDIF3a	Chr09:4193781..4194815	Forward	125	N	31	76.7	PvTDIF3	68.6
GmTDIF3b	Chr15:13038523..13039541	Forward	127	N	29			

[a] Number of amino acid residues.

[b] After amino acid number listed.

Table 3.2 Features of the common bean (*Phaseolus vulgaris*) CLE genes

Listed are the genetic location, pre-propeptide length, and predicted intron presence.

Name	Phytozome v10 ID	Pre-propeptide length[a]	Predicted intron	Chromosome location	Orientation	Oelkers et al. (2008)	uniprot.org
PvCLE1	Phvul.011G065200	96	Y	Chr11:5675757..5676469	Reverse	–	XP_007132079
PvCLE3	Phvul.006G092600	99	Y	Chr06:21113605..21114127	Forward	PvCLE169	XP_007147057
PvCLE4	Phvul.002G008500	67	N	Chr02:960456..961284	Reverse	–	XP_007156683
PvCLE5	Phvul.003G035700	121	N	Chr03:3588969..3589711	Forward	–	XP_007153443
PvCLE6	Phvul.007G027300	94	Y	Chr07:2049797..2054614	Reverse	PvCLE176	XP_007142910
PvCLE7	Phvul.002G085300	108	N	Chr02:13297480..13297806	Forward	–	XP_007157625
PvCLE8	Phvul.009G187200	95	N	Chr09:27684592..27685489	Forward	–	XP_007138182
PvCLE9	Phvul.003G190100	95	N	Chr03:40210422..40210709	Forward	–	XP_007155310
PvCLE10	Phvul.002G079000	101	Y	Chr02:11819569..11820862	Reverse	–	XP_007157554
PvCLE11	Phvul.001G025500	77	N	Chr01:2309373..2309606	Reverse	–	XP_007160889
PvCLE12	Phvul.004G023800	108	Y	Chr04:2459046..2460734	Reverse	–	XP_007151170
PvCLE13	Phvul.005G069900	102	Y	Chr05:11484552..11485119	Reverse	–	XP_007149431
PvCLE14	Phvul.007G068800	88	N	Chr07:6196473..6196739	Reverse	–	XP_007143392
PvCLE15a	Phvul.007G068400	85	N	Chr07:6165176..6165433	Reverse	–	XP_007143388
PvCLE15b	Phvul.007G068500	83	N	Chr07:6181155..6181406	Forward	–	XP_007143389
PvCLE15c	Phvul.007G068600	87	N	Chr07:6184216..6184479	Reverse	–	XP_007143390
PvCLE15d	Phvul.007G068700	84	N	Chr07:6189914..6190168	Forward	–	XP_007143391
PvCLE16	Phvul.004G117600	86	N	Chr04:38385127..38385862	Forward	–	XP_007152295
PvCLE17	Phvul.002G287300	97	N	Chr02:45090923..45091742	Reverse	–	XP_007160038
PvCLE18	Phvul.003G137800	85	N	Chr03:33013056..33013313	Reverse	–	XP_007154669
PvCLE19	Phvul.002G095900	104	Y	Chr02:17549689..17550064	Forward	–	XP_007157755
PvCLE20	Phvul.001G120900	92	N	Chr01:34104465..34105721	Forward	–	XP_007162068
PvCLE21	Phvul.008G203000	88	N	Chr08:51319273..51319539	Forward	–	XP_007141519
PvCLE22	Phvul.006G016000	90	N	Chr06:7671543..7672241	Reverse	-	XP_007146145
PvCLE23	Phvul.008G211300	74	N	Chr08:52313956..52316136	Forward	–	XP_007141620

Name	Phytozome v10 ID	Pre-propeptide length[a]	Predicted intron	Chromosome location	Orientation	Oelkers et al. (2008)	uniprot.org
PvCLE24	Phvul.007G101800	109	N	Chr07:11339237..11339566	Reverse	–	XP_007143789
PvCLE25	Phvul.003G177600	110	N	Chr03:38979082..38979719	Forward	–	XP_007155150
PvCLE26	Phvul.002G168200	85	Y	Chr02:31082684..31084138	Reverse	–	XP_007158622
PvCLE27	Phvul.002G081400	106	N	Chr02:12270950..12272253	Reverse	–	XP_007157583
PvCLE28	Phvul.005G067900	83	N	Chr05:10636536..10636787	Reverse	–	XP_007149409
PvCLE29	Phvul.011G160600	81	N	Chr11:42316953..42317385	Forward	–	XP_007133207
PvCLE30	Phvul.011G160700	82	N	Chr11:42325813..42326352	Forward	–	XP_007133208
PvCLE31	Chr01:14906066..14906353	95	N	Chr01:14906066..14906353	Forward	–	–
PvCLE33	Chr11:42291102..42291350	82	N	Chr11:42291102..42291350	Reverse	-	-
PvCLE34	Chr05:10644869..10645097	75	N	Chr05:10644869..10645097	Reverse	–	–
PvCLE35	Phvul.003G057900	75	N	Chr03:7610340..7610764	Forward	–	XP_007153705
PvCLE40	Phvul.011G056800	114	Y	Chr11:4877577..4878010	Forward	–	XP_007131981
PvCLV3	Phvul.005G120600	104	Y	Chr05:34343926..34344486	Reverse	–	XP_007150035
PvNIC1	Phvul.005G097000	80	N	Chr05:28793851..28794118	Reverse	–	XP_007149764
PvRIC1	Phvul.005G096900	115	Y	Chr05:28775368..28775758	Reverse	–	–
PvRIC2	Phvul.011G135900	93	N	Chr11:30985821..30986626	Reverse	–	XP_007132915
PvTDIF1	Phvul.008G124100	118	N	Chr08:17187233..17187933	Forward	–	XP_007140575
PvTDIF2	Phvul.002G187400	108	N	Chr02:34265616..34266385	Forward	–	XP_007158853
PvTDIF3	Phvul.009G244400	115	N	Chr09:35772334..35773004	Reverse	–	XP_007138869

[a] Number of amino acid residues.

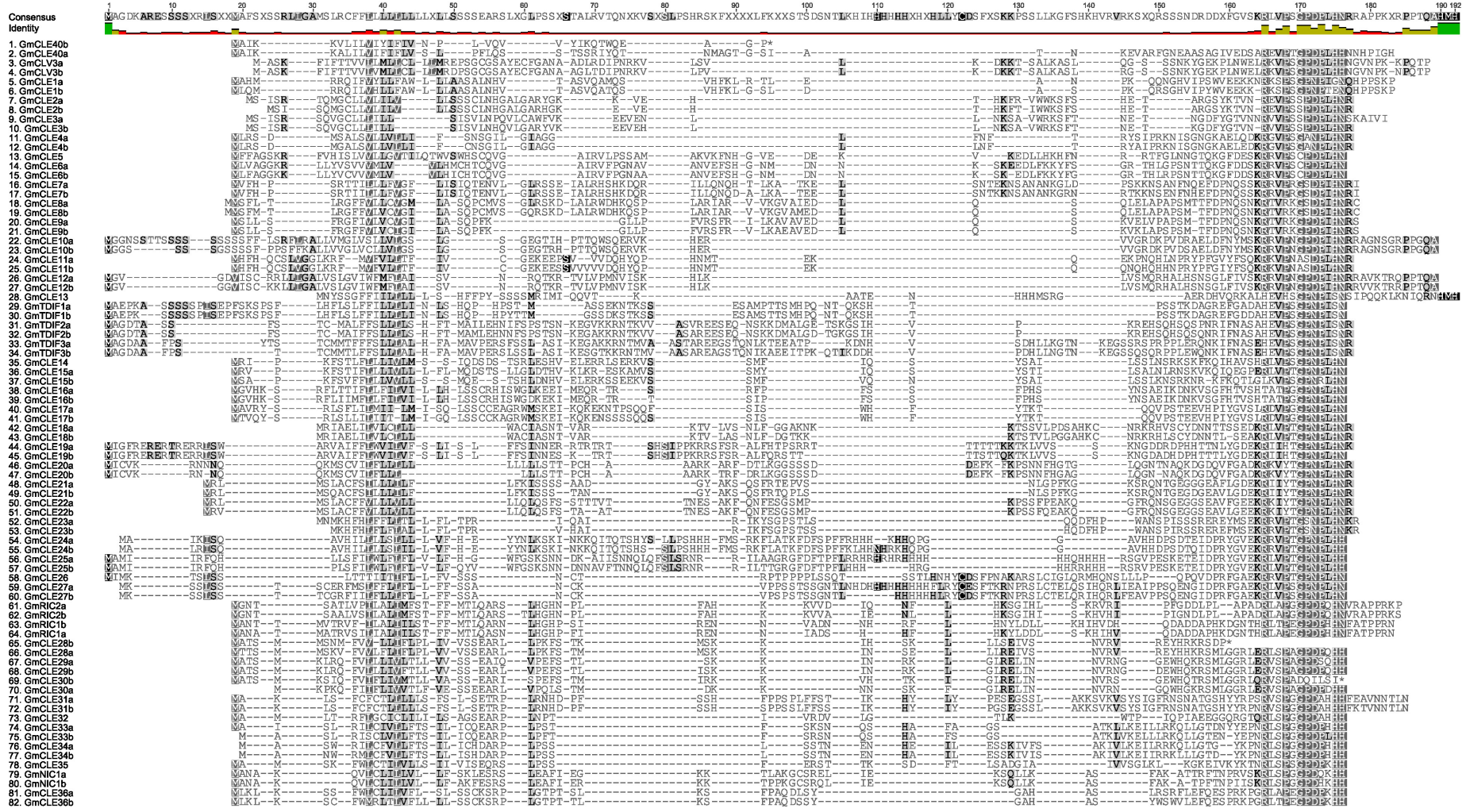


Figure 3.1 Multiple sequence alignment of soybean (*Glycine max*) CLE pre-propeptides. Homeologous copies consistently align together, as do other closely related sequences. Shading of amino acid residues represents conservation, with the darker the shading the more highly conserved the residues. The CLE domain and the leucine-rich region of the signal peptide domain exhibit the greatest degree of conservation across the entire pre-propeptide family.

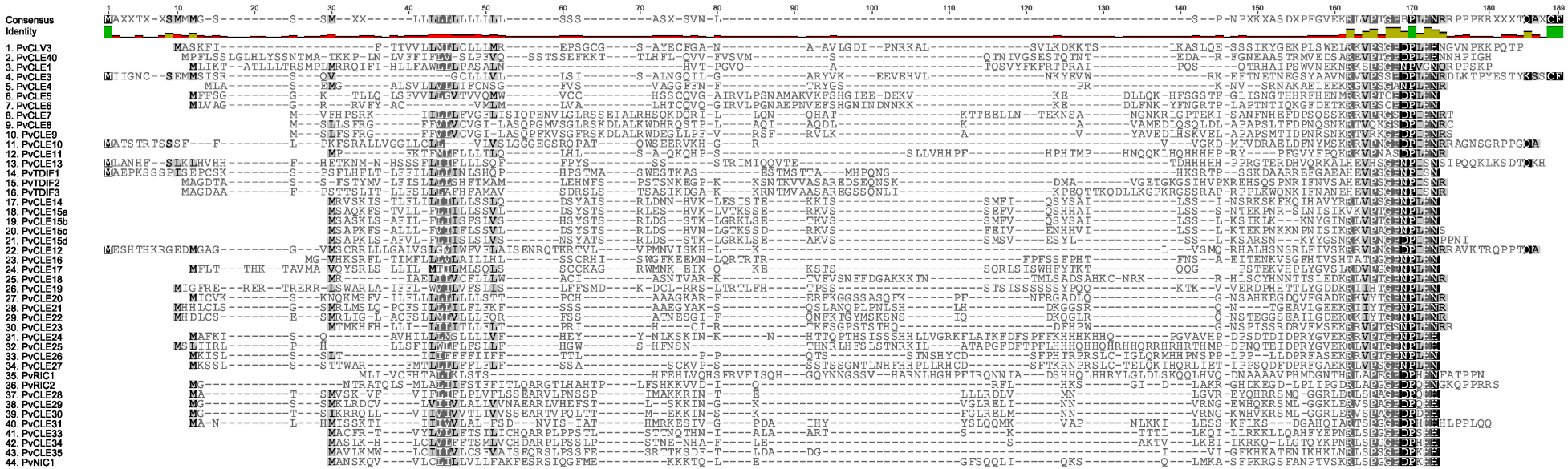


Figure 3.2 Multiple sequence alignment of common bean (*Phaseolus vulgaris*) CLE pre-propeptides. Related sequences tend to align closer together. Shading of amino acid residues represents conservation, with the darker the shading the more highly conserved the residues. As with the soybean prepropeptides shown in Figure 3.1, the CLE domain and the leucine-rich region of the signal peptide domain exhibit the greatest degree of conservation across the entire pre-propeptide family.

When identifying gene orthologues, it was noticed that three of the 44 genes identified in common bean did not have an apparent orthologue in soybean (Table 3.1; Supplementary Fig. S1 at *JXB* online). These genes are all part of a group of four tandemly duplicated genes located on chromosome 7, called *PvCLE15a*, *b*, *c*, and *d*, and thus can all be considered orthologous to the same genes in soybean, *GmCLE15a* and *b*. This indicates that the tandem duplication occurred in common bean after it diverged ~19 MYA from soybean. Directly upstream of these tandemly duplicated genes and adjacent to *PvCLE15d* is another CLE peptide-encoding gene, *PvCLE14* (Fig. 3.3A). This tandem duplication also occurs in soybean (*GmCLE14* and *GmCLE15a*) and thus must have occurred prior to the two species diverging.

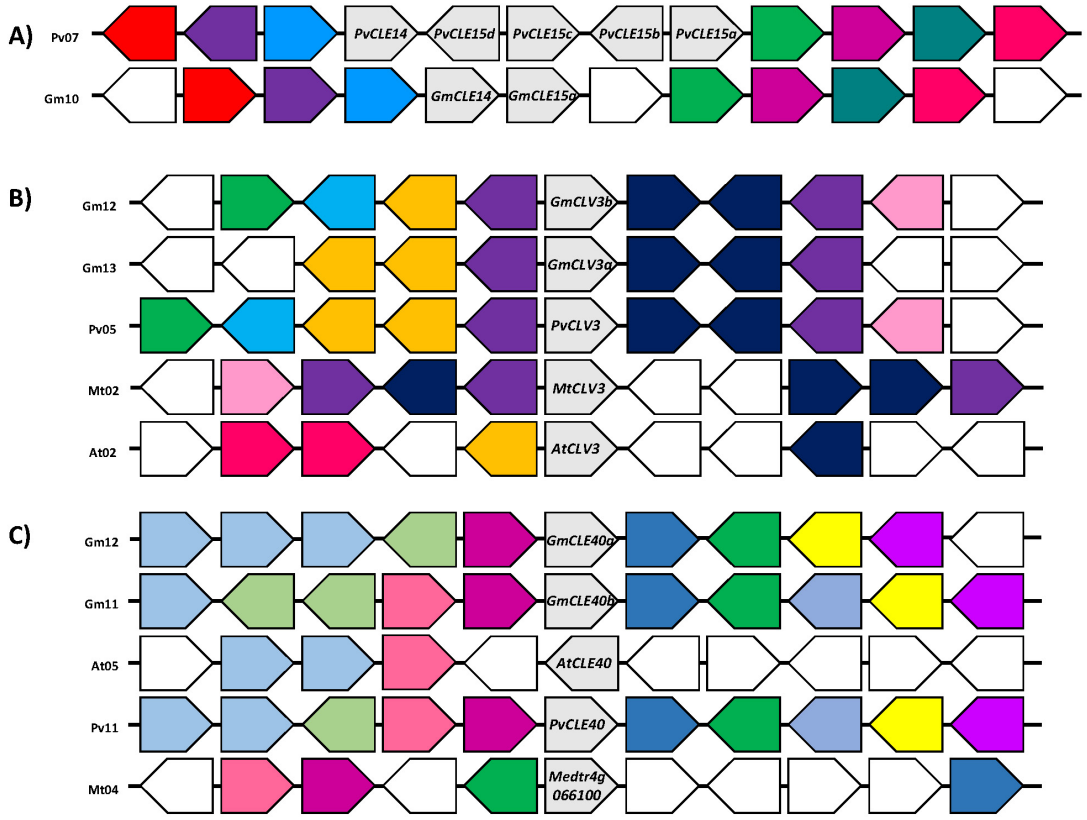


Figure 3.3 Genomic environment of PvCLE15 tandemly duplicate genes of common bean, and the CLV3 and CLE40 genes of different species. The genes of interest are positioned centrally and shaded in grey. Species and chromosome number are indicated to the left of each genomic segment. Surrounding genes similar in putative function are indicated by the same colour and genes with unrelated putative functions are uncolored. The direction of the arrow represents the orientation of the gene compared with that of the CLE gene. **(A)** Common bean chromosome 7 containing a tandem gene duplication not found on the orthologous region of soybean on chromosome 10. Orthologues of **(B)** CLV3 and **(C)** CLE40 in soybean, common bean, Arabidopsis, and M. truncatula. A high level of genetic synteny is shown here for each of these CLE genes.

Two additional sets of genes occur in tandem in common bean: *PvCLE29* and *PvCLE30*, and *PvNIC1* and *PvRIC1*. In soybean, the *NIC1* and *RIC1* genes also occur in tandem, suggesting that this duplication occurred prior to the divergence of soybean and common bean. However, due to the whole-genome duplication, soybean has homeologous regions that include these genes, resulting in two tandem repeats: *GmNIC1a* and *GmRIC1b* on chromosome 12 and *GmNIC1b* and *GmRIC1a* on chromosome 13.

Manual adjustments were made to some coding sequences predicted in Phytozome regarding the placement of their start codon. These adjustments were based on sequence similarity to their duplicate gene, to clustering sequences in common bean (i.e. probable orthologues), and/or to signal peptide domain prediction results (described below). In total, eight soybean sequences were trimmed slightly to place their start codon downstream of where it was predicted in Phytozome (GmCLE10b, GmCLE16b, GmCLE21b, GmCLV3b, GmTDIF1a, GmTDIF1b, GmRIC1a, and GmRIC2b). An additional five sequences were extended to include a start codon slightly upstream of that predicted in Phytozome (GmCLE3a, GmCLE16a, GmCLE20a, GmCLE27a, and GmCLE28a).

3.4.3 Characterization of CLE pre-propeptides in soybean and common bean

CLE pre-propeptides typically consist of a signal peptide, a variable domain, and a CLE domain, with some also having a C-terminal extension (Hastwell et al. 2015). All of the CLE pre-propeptides identified here have this structure. Moreover, they are rich in lysine (11.4%) and serine (11.3%), and are notably poor in cysteine (1.3%), tyrosine (1.3%), and tryptophan (0.7%; often poorly represented in plants) (Supplementary Table S2 at *JXB* online), which is typical amongst CLE peptides (Hastwell et al. 2015). The length of the CLE pre-propeptides varies, with the smallest being 67 residues in both soybean and common bean (excluding likely pseudogenes reported below), and the longest being 127 and 121 residues, respectively. Some contain histidine repeats in their variable domain, but this does not correlate with sequence length.

The signal peptide located at the N-terminus of the pre-propeptide is typically hydrophobic and is responsible for exporting the propeptide from the cell (Rojo et al. 2002). Hydrophobicity analysis confirmed that the signal peptide is the most hydrophobic region of the CLE pre-propeptides investigated here, whereas the remaining propeptide is more hydrophilic, as determined by Kyte and Doolittle (1982) scores (Supplementary Fig. S2 at *JXB* online). Indeed, 61.4% of the amino acid residues occurring in the signal peptide

domain are hydrophobic (Supplementary Fig. S2). SignalP prediction software was used to determine the putative cleavage site of the signal peptide (Table 3.1). Using these predicted signal peptide sequences, a multiple sequence alignment and phylogenetic tree was constructed that showed less conserved and confident groupings (data not shown) compared with entire pre-propeptides. One pre-propeptide, GmCLE40b, is not predicted to have a signal peptide, as it is truncated and only 34 amino acids in length (Table 3.1; Fig. 3.1).

Directly following the signal peptide domain in the pre-propeptide is the variable domain. This region only shows conservation between homeologous and/or orthologous genes (Figs 3.1, 3.2). However, the final residue of the variable domain positioned directly before the CLE domain is commonly a lysine (48.4%), with asparagine (13.9%), glutamic acid (9.0%), alanine (7.4%), and histidine (5.7%) as the next four highest represented amino acids at this position.

The CLE domain represents the region of the pre-propeptide that is cleaved and modified to become the functional CLE peptide product. Of the 126 CLE peptide-encoding genes of soybean and common bean, there are 54 unique CLE domain sequences that are 12 amino acids in length (with 44 of 82 in soybean and 40 of 44 in common bean). This number increases to 60 sequences if 13 amino acids are taken into account. All mature CLE peptides that have been biochemically confirmed to date have been 13 amino acids in length (Ohyama et al. 2009; Shinohara et al. 2012; Okamoto et al. 2013; Chen et al. 2015); however, only 54.8% of the pre-propeptide CLE sequences of soybean and common bean have a residue in position 13, with the others having a stop codon preventing them from being any more than 12 amino acids in length.

Sequence similarity within the CLE pre-propeptides of soybean and common bean is highest in the CLE domain (Figs 3.1, 3.2). There is no 100% conserved residue, although position 12 has a highly conservative histidine/asparagine substitution. The least conserved residues are at position 2 (15.8% pairwise identity) and position 5 (19.7% pairwise identity). Of the critical residues previously identified in the CLE domain (e.g. Ni et al. 2011; Reid et al. 2013), position 1 is predominantly arginine, or, in some cases, histidine (i.e. TDIF peptides). An additional group has threonine at position 1 (GmCLE16a, GmCLE16b, and PvCLE16). Three others that group together have valine, lysine, and leucine residues at this position (PvCLE15a, PvCLE15d, and GmCLE15b, respectively; Figs 3.1, 3.2), which includes two of the four common bean genes that are tandemly

duplicated (described above). Position 7, which is often post-translationally modified, is predominately a proline. However, there are 10 soybean homeologues and five associated common bean orthologues where a serine (CLE7; CLE8; CLE11 and CLE23 orthologous) or alanine (CLE4 orthologues) is in that position. Interestingly, soybean has six pairs (i.e. 12 genes) of homeologous CLE peptide-encoding genes that have a mismatch within their CLE domain as a result of naturally occurring mutations (Fig. 3.1). The impact of amino acid changes on the function and activity of various Arabidopsis and legume CLE pre-propeptides was recently reviewed (Hastwell et al. 2015).

Some CLE pre-propeptides contain a fourth domain directly following the CLE domain, called the C-terminal extension. The precise function of this domain remains unclear. Only 32.5% of the CLE pre-propeptides in soybean and common bean have this domain, similar to the CLE pre-propeptide family of *A. thaliana* (31.3%; Cock and McCormick, 2001). The only prevalent feature of the C-terminal extension appears to be the common presence of proline (19.5%). Indeed, the sequence is highly variable in length and amino acid residues, except between homeologous and/or orthologous genes (Fig. 3.1). Interestingly, the domain is present in 83.3% of the CLE genes that contain a predicted intron. It is also present in CLV3 orthologues and in almost all rhizobia-induced nodulation-suppressing CLE peptides (with the exception of MtCLE12; Hastwell et al. 2015).

3.4.4 Pseudogenes and multi-CLE peptide-encoding genes of soybean and common bean

Due to insertion, duplication, and deletion events, some of the CLE peptide-encoding genes identified here do not fit the common tripartite domain structure. For example, in soybean, *GmCLE28b*, *GmCLE30b*, and *GmCLE40b* are all probably pseudogenes. *GmCLE28b* and *GmCLE40b* have nonsense mutations that result in a truncation prior to the CLE domain. However, the sequences downstream of these mutations align closely to *GmCLE28a* and *GmCLE40a*, respectively. *GmCLE30b* has low conservation in the CLE domain after residue five, when compared with its duplicate, *GmCLE30a*. This appears to be due to a deletion event causing a frameshift directly in the CLE domain. It is likely that none of these three pseudogenes genes produces a functional CLE peptide. They have been denoted as the *b* copy, consistent with the *RIC*, *NIC*, and *CLV3*genes, where the *b* copy may not be transcribed/functional (Reid et al. 2011*a*; Wong et al. 2013).

Genes encoding pre-propeptides that contain multi-CLE domains were also identified. This includes *GmCLE37a* and *GmCLE37b*, which have six possible CLE domains each (Fig. 3.4A). These were excluded from the alignment in Fig. 1as they do not have the archetypical domain structure. There are only two identical CLE domains within the soybean multi-CLE domain pre-propeptides and they both occur in GmCLE37b (Fig. 3.4A). A multi-CLE domain-containing pre-propeptide previously reported in *Medicago truncatula* by Oelkers et al.(2008) was identified here as *MtCLV3* (*MtCLV3* was previously discovered by Chen et al. 2009, but was not reported to encode a multi-CLE domain). Although *MtCLV3* encodes three CLE domains, only one is actually translated due to the presence of a previously undetected intron identified here. An additional pre-propeptide of *M. truncatula*, called MtCLE14, contains a multi-CLE domain with seven CLE peptide domains (Fig 3.4A; Mortier et al. 2011). MtCLE14 contains four identical 12 amino acid CLE domains in tandem, each followed by an asparagine residue (possible representing a 13th residue in the CLE peptide), and each preceded by the same two hydrophobic residues (Fig. 3.4A).

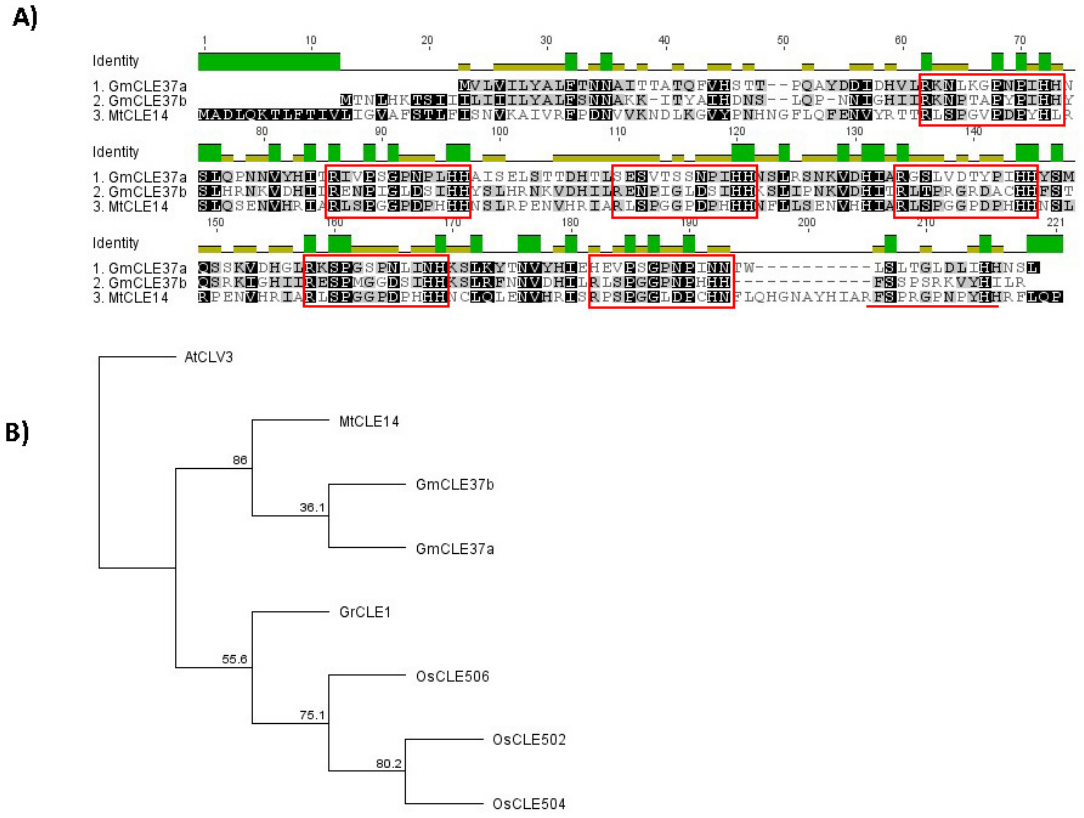


Figure 3.4 Multi-CLE domain pre-propeptides. **(A)** Multiple sequence alignment of the soybean and *M. truncatula* multi-CLE domain pre-propeptides, with putative 13 amino acid residue CLE domains highlighted by a red box. An additional CLE domain of MtCLE14 that is not detected in the two soybean pre-propeptides is underlined in red. Four MtCLE14 CLE domains are identical in sequence (CLE domains 2–5) while there are no 100% conserved 13 amino acid residue CLE domains in soybean.

However, there are two fully conserved 12 residue CLE domains in GmCLE37b (CLE domains 1 and 2). **(B)** Phylogenetic tree of known multi-CLE domain-containing pre-propeptides of rice (*Oryza sativa*), potato cyst nematode (*Globodera rostochiensis*), *MtCLE14* of *M. truncatula*, and the newly identified *GmCLE27a* and *GmCLE37b* of soybean, including AtCLV3 as an outgroup. The multi-CLE domain pre-propeptides identified here cluster separately from those that were previously identified. The tree is shown with bootstrap confidence values expressed as a percentage from 1000 bootstrap replications.

3.4.5 Categorization and functional predictions of soybean CLE peptides

The function of many CLE peptides can be predicted based on sequence. The Arabidopsis CLE peptides are currently categorized into two groups: type-A affecting root and shoot meristem development, and type-B affecting vasculature development (Matsubayashi 2014). The soybean CLE peptides were assigned into different categories based on the sequence alignment, phylogenetic grouping of their pre-propeptides, and their functional roles where known. The groups were initially defined based on phylogenetic analysis, and were then further refined following examination of their CLE domain and adjacent residues. In total, seven groups (Groups I–VII) were identified (Fig. 3.5). Logo alignments (Fig. 3.6) were subsequently constructed to establish the level of conservation within the 13 amino acid CLE domain of each group, with highly conserved residues probably critical to their function.

Group I is small, consisting of only four members. It contains CLV3, CLE40, and their homeologous duplicates (Fig. 3.5). CLV3 and CLE40 are well characterized and are responsible for apical meristem regulation in the shoot and root, respectively (Grienenberger and Fletcher 2015). The CLE domain of this group is highly conserved (Fig. 3.6), particularly for amino acid residues reported to be critical for function (Song et al. 2013).

Group II contains the least conserved CLE domain of all the established groups. It is also the largest group, with 23 members, which may account for it having the lowest degree of conservation (Figs 3.5, 3.6). The group cannot be divided further with any degree of confidence using a phylogenetic approach. Interestingly, it has low conservation at residue six, which is generally considered to be critical for function, possibly having a role in enabling the CLE peptide to rotate or bend (Hastwell et al. 2015). Most of the CLE peptides in this group remain poorly characterized in any species; however, some of the

soybean CLE pre-propeptides show similarity to, and group closely with, AtCLE45 (Supplementary Fig. S3 at *JXB* online).

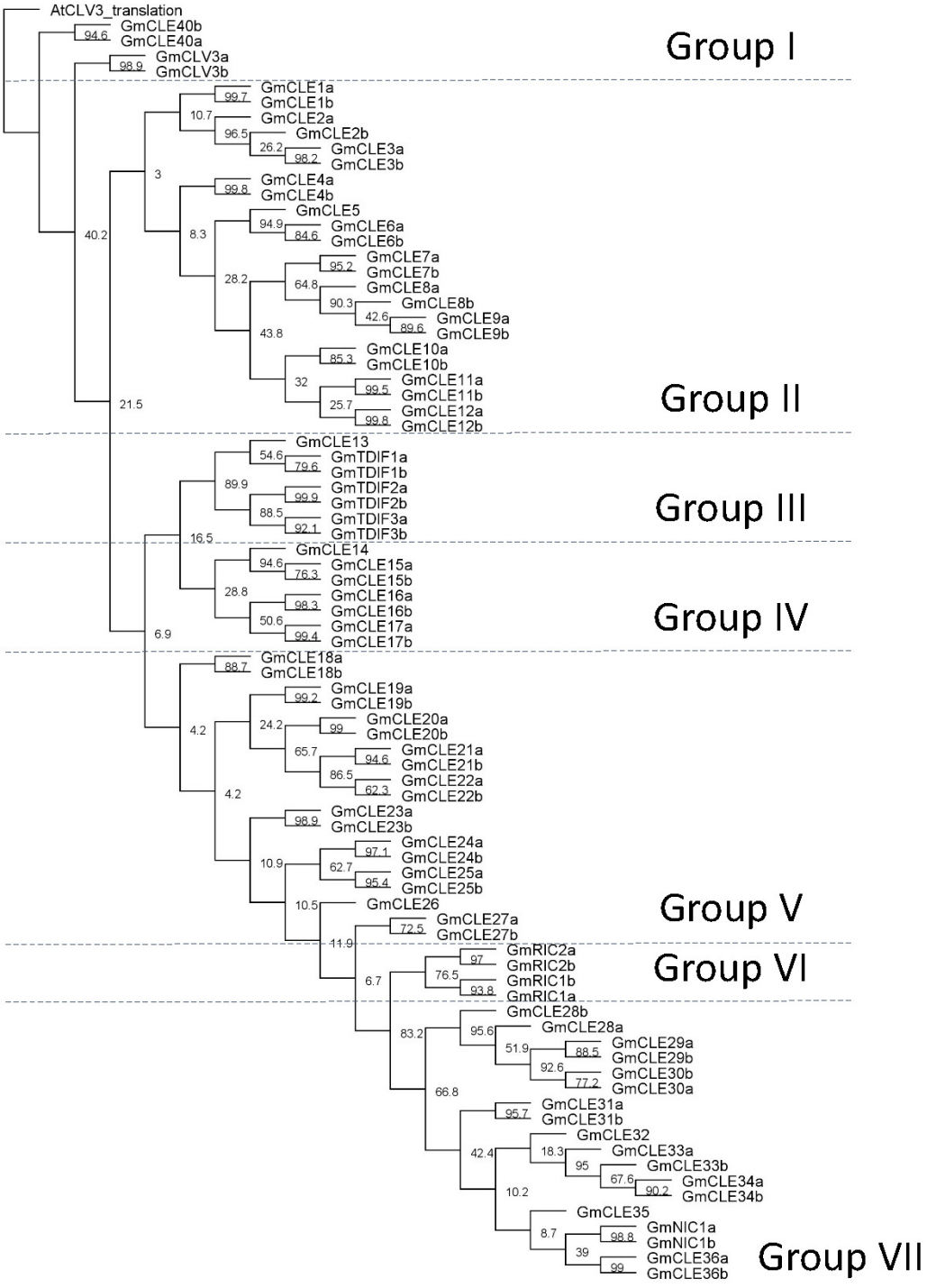


Figure 3.5 Soybean CLE pre-propeptide phylogenetic tree illustrating the seven distinct identity groups. Phylogenetic analysis was performed using the multiple sequence alignment generated with entire pre-propeptide sequences (Fig. 3.1), including AtCLV3 as an outgroup. Homeologous genes consistently cluster together with high confidence (indicated by high bootstrap values). The seven groups (Group I–VII) were assigned based on clustering in the tree, in addition to sequence similarity. The tree is shown with bootstrap confidence values expressed as a percentage from 1000 bootstrap replications.

Group III contains seven members, including the three TDIF pre-propeptides and their homeologues, in addition to one other member of unknown function that lacks a duplicate copy (Fig. 3.5). This group is orthologous to the Arabidopsis type-B CLE pre-propeptides that influence vasculature development, including AtCLE41, ACLE42, and AtCLE44 (Fig. 3.5; Supplementary Fig. S3 at *JXB* online; Matsubayashi 2014). A defining feature of this soybean group is that all of the CLE peptides begin with a histidine residue, as opposed to the classical arginine (Fig. 3.6). Interestingly, with the exception of the non-TDIF peptide (GmCLE13), the 12 amino acid CLE domain is 100% conserved. Also of note is that the members of this group are the only CLE peptides to have a serine residue at position 11, rather than the characteristic histidine (Fig. 3.6).

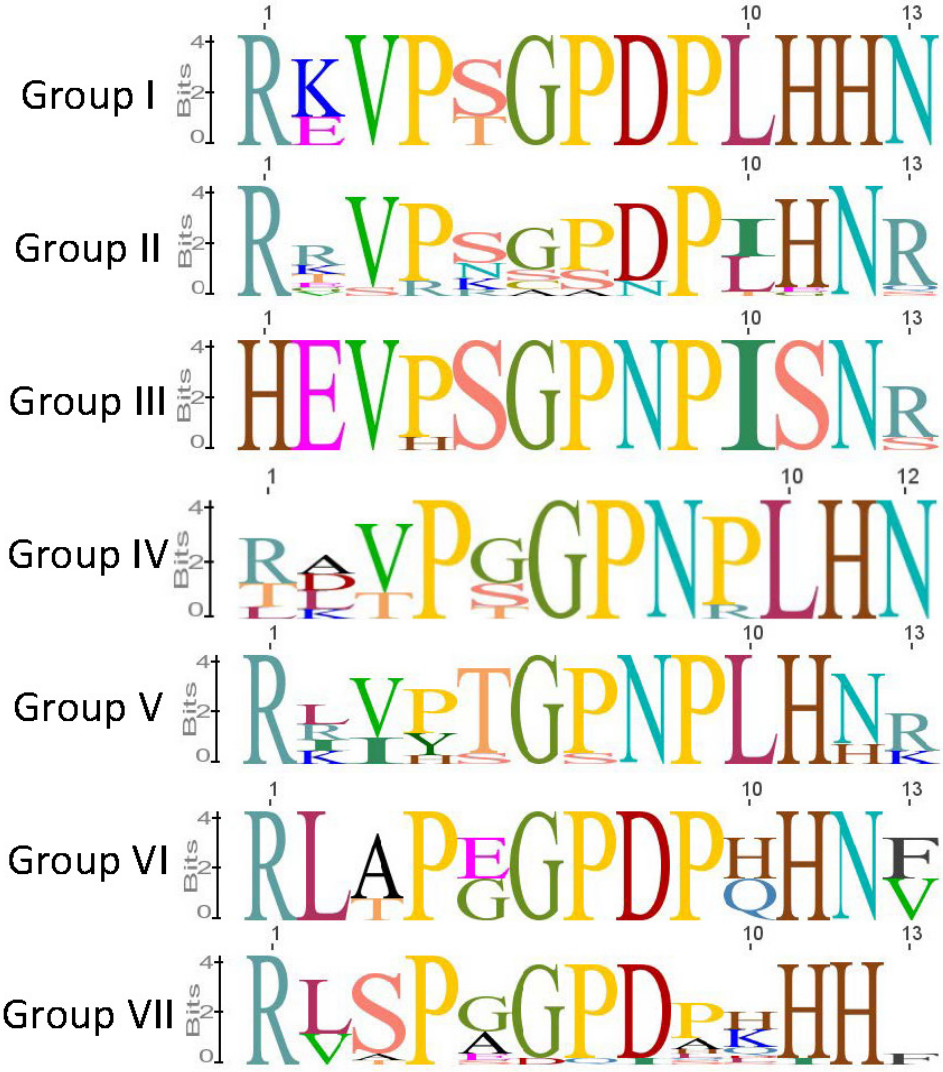


Figure 3.6 CLE domain consensus sequences from the seven soybean pre-propeptide groups. Logo diagrams illustrate the 13 amino acid CLE domain consensus sequences for soybean CLE Groups I–VII, as determined from multiple sequence alignments generated for each group. The 13th amino acid is a consensus of only those sequences that have a residue at that position. Group IV does not have any residues at that position and hence the logo diagram for this group is 12 residues only.

Group IV consists of seven members and notably does not encode any CLE peptides that are 13 amino acids in length (Fig. 3.6). It is also the group that is least conserved at residue one. The function of the group members remains poorly defined.

Group V is another large group, having 19 members (Fig. 3.5). Of the CLE peptides encoded by this group, all but one contain an acidic amino acid (glutamic acid or aspartic acid) and a lysine residue immediately preceding the CLE domain (Fig. 3.1). The CLE peptides encoded by this group also predominantly have a threonine at position 5, which is not characteristic of any of the other groups (Fig. 3.6).

Group VI is a small group consisting entirely of the rhizobia-induced CLE peptides (RICs) and their homeologous copies (Fig. 3.5). This group has been well characterized for their role in regulating legume nodule development (reviewed in Hastwell et al. 2015), including the identification of amino acid residues in the CLE domain that are critical for function (Reid et al. 2013).

Group VII consists of 18 members, and, like Group I, has two histidine residues located at positions 11 and 12 (Figs 3.5, 3.6). It contains the majority of the genes that were unpredicted in Phytozome (Table 3.1). The function of most remains unknown; however, it does include the nitrate-induced CLE peptide (NIC1a) and its homeologue, NIC1b (Reid et al. 2011*a*; referred to as NIC2 in Lim et al. 2014), that is well known for its role in controlling legume nodulation in response to the nitrogenous content of the rhizosphere (reviewed in Hastwell et al. 2015).

These groupings hold true when the common bean CLE pre-propeptides are added to the phylogenetic analysis with soybean (Supplementary Fig. S1 at *JXB*online). When Arabidopsis is also included (Supplementary Fig, S3), the groupings are still conserved generally, but are supported by lower bootstrap proportions, especially Group II. This is not surprising when dealing with >150 pre-propeptides from three different species and, even though some groups are divided further when a non-legume is included, the larger groups cannot be confidently split further based on the low bootstrap proportions. In all instances, Group III is supported by very high bootstrap proportions (>88).

A C-terminal extension is encoded by one-third of the genes identified here, spanning across the various groups, but predominantly being found in Groups I, II, and VI (Figs 3.1, 3.5). GmCLE31a and b, and GmCLE13, also contain a C-terminal extension. The presence of a predicted intron correlates slightly with the groupings, as all of the genes in

Group I contain a predicted intron, as do some in Group II, but none in Groups III–VII, with the exception of *GmCLE13*(Group III), which incidentally also contains the only CLE domain sequence divergence of its group, as noted above (Table 3.2; Figs 3.1, 3.5, 3.6).

The groupings described here could help in elucidating the function of CLE peptides where a function is yet to be assigned. Indeed, these groupings, together with genomic environment analyses, were used to identify previously unknown soybean and/or common bean orthologues of AtCLV3-, AtCLE40-, and TDIF-encoding genes, as well as likely *M. truncatula* orthologues. *AtCLV3*was the first CLE gene to be identified in any species (Fletcher et al. 1999) and has since been identified in soybean and *M. truncatula* (*GmCLV3a*, *GmCLV3b*, and *MtCLV3*; Chen et al. 2009; Wong et al. 2013). Investigations into the genomic environment and pre-propeptide sequence similarity (Fig. 3.3B) led to the identification of a CLV3 orthologue in common bean. Similar approaches were used to identify *AtCLE40* orthologues (Fig. 3.3C) in common bean and *M. truncatula*, in addition to *GmCLE40b*, the homeologue of *GmCLE40a*. Moreover, all TDIF orthologues in soybean, common bean, and *M. truncatula* were established (Fig. 3.7). In contrast, despite *AtCLE46* and *GmCLE13* sharing a high level of sequence similarity in the CLE domain, they do not show synteny to the TDIF genes, or to each other, and cluster separately (Fig. 3.7). Thus, these genes are unlikely to be true TDIF peptides.

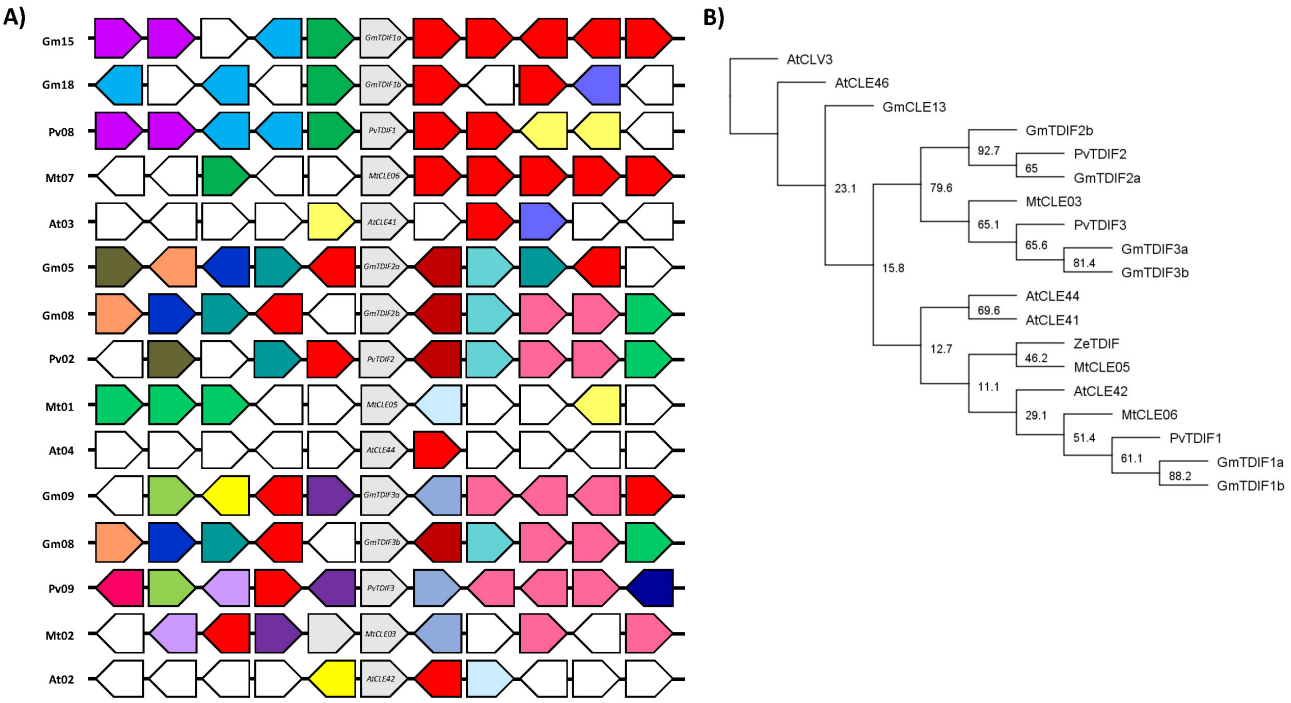


Figure 3.7 TDIF genes in soybean, common bean, *Arabidopsis*, *Zinnia elegans*, and *M. truncatula*. **(A)** Genomic environments of the TDIF-encoding genes highlight the genetic synteny between the genes identified here in soybean, common bean, and *M. truncatula* with previously characterized TDIF genes

of *A. thaliana*, *AtCLE41*, *AtCLE42*, and *AtCLE44*. TDIF-encoding genes are shown positioned centrally and shaded in grey. Species and chromosome number are indicated to the left of each genomic segment. Surrounding genes similar in putative function are indicated by the same colour and genes with unrelated putative functions are uncoloured. The direction of the arrow represents the orientation of the gene compared with that of the CLE gene. A high level of genetic synteny is shown here for each of the predicted TDIF-encoding genes, but was not found for *AtCLE46* and *GmCLE13* (data not shown), whose CLE domain begins with a histidine residue but is not a TDIF peptide. **(B)** Phylogenetic tree of TDIF-encoding pre-propeptides, including ZeTDIF, and also AtCLV3 as an outgroup. Two pre-propeptides, AtCLE46 and GmCLE13, are also included that have CLE domains beginning with a histidine residue, but are not true TDIF CLE peptides and did not group with the TDIF pre-propeptides. The tree is shown with bootstrap confidence values expressed as a percentage from 1000 bootstrap replications.

3.4.6 Expression analysis of CLE peptide-encoding genes of soybean, common bean, and Arabidopsis

A meta-analysis of the publicly available transcriptome data was conducted in soybean, common bean, and Arabidopsis (Supplementary Tables S3–S5 at *JXB*online). The transcriptomic expression of functionally characterized soybean and common bean CLE peptide-encoding genes was consistent with the literature (i.e. RICs and NIC1, Reid et al. 2011*a*; Ferguson et al. 2014). Interestingly, there were no transcriptional data available for CLV3 orthologues in soybean and common bean (Supplementary Tables S3, S4).

Trends observed in the expression of CLE peptide-encoding gene orthologues across different tissues of soybean and common bean were also consistent (Supplementary Tables S3, S4 at *JXB* online). For example: *PvCLE10*, *GmCLE10a*, and *GmCLE10b* showed varying levels of expression across all tissue types, in a similar trend; *PvCLE17* and *GmCLE17a* are expressed in all tissue types except seeds, flowers, and early pod growth; and *PvCLE19* and *GmCLE19a* show expression in all tissues except mature nodules. These three orthologous gene groups (CLE10, CLE17, and CLE19) also show high (>93) bootstrap values in the phylogenetic analyses (Supplementary Fig. S2). In contrast, *CLE24* showed different expression patterns between soybean and common bean orthologues. *GmCLE21a* and *GmCLE21b* show the same expression trends, but *PvCLE21* transcripts were only detected in the early seed development stage. In soybean, where data were available for both the *a* and *b* copy, the general trend of expression was consistent but in most cases the level or the time of expression varied. There is no consistent expression pattern between pre-propeptides belonging to soybean Groups I–

VII, but closely related peptides probably perform a similar role in different developmental tissues as with the TDIF orthologues (Supplementary Tables S3–S5; Matsubayahsi 2014).

To determine if expression trends are similar between orthologues of soybean, common bean, and Arabidopsis, and to see how orthologues clusters, a phylogenetic tree of the pre-propeptides from the three species was produced (Supplementary Fig. S3 at *JXB* online). Branches that were supported by >50 bootstrap proportions include AtCLE46 and CLE1; AtCLE21 and CLE4; AtCLE27 and CLE6; AtCLE20 and CLE23; AtCLE12 and CLE24; and the cluster containing the TDIF orthologous genes, as established previously in Fig. 3.7.

As expected, the legume orthologues show a similar expression trend for each of these branches and, in the case of AtCLE12, a similar trend was observed with GmCLE24a and PvCLE24 (Supplementary Tables S3–S5 at *JXB* online). Interestingly, AtCLE27 and AtCLE21 were not expressed in any tissues, similar to the case of their respective and related legume pre-propeptides (Supplementary Fig. S3). All the TDIF orthologues with available expression profiles show a highly similar pattern (Supplementary Tables S3–S5).

Within the meta-analysis of the transcriptomes, interesting candidates were identified as targets for future functional characterization. PvCLE29 was found only in the flower at a very high level; PvCLE24 shows very high root and nodule expression (Supplementary Table S4 at *JXB* online); and GmCLE25a is only expressed in root tissue (Supplementary Table S3).

The meta-analysis shows similar trends for orthologous genes. However, to date, only one-third of the CLE peptide-encoding genes of soybean, and less than half from Arabidopsis, are represented. It is also likely that some genes that respond to external stimuli (e.g. rhizobia for *RIC1* and *2* and nitrate for the *NIC1* orthologues) were not induced if the required treatment was not part of the study.

Feeding studies were not attempted here because the precise size and modification of each of the novel peptides is completely unknown. Although feeding unmodified or semi-modified synthetic peptides could be attempted, the peptides being fed would be designed based on prediction (in terms of both length and modifications). Furthermore, they would be applied in unnaturally high concentrations, without regard to temporal or spatial regulation, to a broad range of tissues and cell types to which they might not normally

localize. These issues would be further exacerbated in feeding studies using roots grown on agar containing high levels of sucrose and nitrate, and exposed to light. Such studies would result in an extremely high frequency of false-positive outcomes that are of little biological value. For comparison sake, an ecologist investigating the impact of wild boars on the environment would not flood a forest with hams. Indeed, it has readily been shown that CLE peptides altered from their correct modification, size, and location can induce a phenotypic effect in feeding (e.g. Fiers et al. 2005; Whitford et al. 2008; Ohyama et al. 2009; Mortier et al. 2010; Kondo et al. 2011) or site-directed mutagenesis and domain-swap studies (e.g. Ni and Clark, 2006; Song et al. 2012; Reid et al. 2013). CLE peptides unlikely to come into contact with a given receptor can be forced to bind to that receptor *in vitro*(as elegantly demonstrated by Shinohara and Matsubayashi 2015). Thus, results from peptide feeding studies may not be biologically relevant, and any phenotypic changes observed would need to be interpreted with extreme caution. For these reasons, the focus here was to use alternative approaches to help determine the role of novel peptides of unknown structure and function.

3.5 Discussion

CLE peptides are widely recognized as important contributors to plant signalling and development; however, a lot remains to be understood about these critical signal molecules. Here, this emerging field was enhanced by the discovery and categorization of the CLE peptide families of soybean and common bean, two of the world's most agriculturally important crops. A total of 84 CLE peptide-encoding genes in soybean and 44 in common bean were identified, and subsequently an array of bioinformatic approaches were conducted for comparative genomic and molecular evolution analyses. Doing so led to the identification of three pseudogenes, two multi-CLE domain-encoding genes in soybean, and a tandem gene duplication event in common bean. It also enabled the establishment of all homeologous gene copies within soybean, and orthologous copies amongst soybean, common bean, and Arabidopsis. Searches using rhizobia and mycorrhiza genomes were also performed, but revealed no CLE peptide-encoding genes in these organisms. Thus, to date, CLE peptides appear to be exclusive to plants and nematodes.

The function of most CLE peptides remains completely unknown. However, phylogenetic analyses of the entire CLE pre-propeptide families of soybean, common bean, and

Arabidopsis show that they group strongly according to their CLE domain and known/predicted function. Based on the analyses, it is demonstrated that the soybean CLE pre-propeptides (excluding multi-CLE domain-encoding genes) grouped into seven distinct categories (Groups I–VII) and that these groups are generally preserved when other species are included. This expands on the two groups reported in Arabidopsis (type-A affecting root and shoot development, and type-B affecting vasculature development; e.g. Matsubayashi 2014). The categorization approach reported here could be a useful tool for elucidating the function of unknown CLE peptides and their closely related homeologous and orthologous sequences. As an example, all known CLE peptides of similar function were found to group together (CLV3 and CLV40 formed Group I, the TDIFs formed Group III, and the RICs formed Group VI). Moreover, the groupings revealed a number of highly conserved amino acid residues present in the peptide domains of each group, which are probably central to the activity of their ligands.

The groups identified here include peptides performing a similar developmental role in a range of different tissues, as exemplified by Group III, whose Arabidopsis orthologues are known to have the same function (Matsubayashi 2014) but are expressed in a range of different tissues. This is also seen with the Group I and Group VI peptides. Given that the genes encoding the members of these groups do not show consistent expression patterns, it is possible that they too may have similar roles in different tissues. Furthermore, the transcriptome evidence presented here provides some insight into where the peptides function, as they often act in a local manner (Matsubayashi 2014). Indeed, the only known CLE peptides to act systemically are those involved in the autoregulation of nodulation signalling pathway of legumes (Hastwell et al. 2015).

The ancestral genome shared by soybean and common bean duplicated ~59 MYA and subsequently reconverged (Schmutz et al. 2010). Later, following the divergence of the two species, the soybean genome duplicated again ~13 MYA and, as a result, there are typically two soybean orthologues present for every common bean gene (Lin et al. 2010; Schmutz et al. 2014). This trend is consistent with the present findings, where common bean contains approximately half the number of CLE peptide-encoding genes as soybean. The findings are also consistent with Arabidopsis, which is reported to have only 32 CLE peptide-encoding genes (Cock and McCormick 2001), and is well known for fractionation (i.e. preferentially removing redundant and/or excess genomic information; Thomas et al. 2006). Indeed, Group VI of the soybean and common bean CLE peptide families identified here is completely absent from Arabidopsis. This category is known to be induced by

rhizobia to control legume nodulation (reviewed in Hastwell et al. 2015), suggesting that either Arabidopsis has completely lost this group, or that the legume species have gained it as a means of regulating the relationship with their symbiotic partner.

Additional methods were employed here to identify conclusively soybean and common bean orthologues of a number of key CLE peptide-encoding genes of Arabidopsis. Indeed, orthologues of *AtCLV3*, which acts in the SAM to control stem cell numbers (Gaillochet et al. 2015), were identified in common bean, and confirmed in soybean and *M. truncatula* (Chen et al. 2009; Wong et al. 2013). Interestingly, it is also shown that *MtCLV3* encodes three CLE peptide domains, but only one is translated due to the presence of an intron. Orthologues of *AtCLE40*, which acts in the RAM to control stem cell numbers (Hobe et al. 2003; Sharma et al. 2003; Stahl et al. 2009), were also identified here in these same three legume species. This includes the homeologous copy of *GmCLE40a*, called *GmCLE40b*, which is unlikely to produce a functional product due to a naturally occurring mutation that truncates the pre-propeptide prior to the CLE domain. Orthologues of the three TDIF CLE peptide-encoding genes of Arabidopsis, which act throughout the plant in vascular differentiation (Grienenberger and Fletcher 2015), were also identified here, including six genes in soybean, three in common bean, and three in *M. truncatula*. The predicted TDIF-encoding genes (together with one other soybean gene of unknown function) make up Group III of the CLE pre-propeptide family. A number of additional Arabidopsis orthologue candidates were also identified throughout the other various CLE peptide groups defined here.

Genome-wide searches to identify CLE peptide-encoding genes in legumes have been conducted previously using soybean, *M. truncatula*, and *L. japonicus*(Cock and McCormick 2001; Oelkers et al. 2008; Okamoto et al. 2009; Mortier et al. 2010, 2011; Lim et al. 2011), with a few additional genes also identified in common bean (Oelkers et al. 2008; Ferguson et al. 2014). However, many of these studies were limited by the technology and bioinformatic resources available at the time. Recent bioinformatic advances were capitalized on here to identify, and subsequently characterize, categorize, and compare thoroughly, the CLE peptide families of soybean and common bean. This also enabled unification of the nomenclature for these species, taking into account the duplicated nature of the soybean genome and the presence of orthologous genes amongst the two species.

Taken together, this research helped to assemble the complete CLE peptide families of two agriculturally important legume species, categorized them into groups to provide

insight into their structure and function, identified key orthologues existing amongst them and Arabidopsis, and used transcriptional evidence to help elucidate their localization and activity. This represents one of the most in-depth studies conducted within and between any CLE peptide family to date. Future work to establish unequivocally the function of these critical peptides, identify their binding partners, and determine the precise structural modifications of their mature ligands is now needed to enhance further the understanding of these novel hormones in regulating plant development.

3.6 Acknowledgements

This work was funded by the Hermon Slade Foundation, and the Australian Research Council Discovery Project grants (DP130103084 and DP130102266). The Fellowship Fund Inc. is also thanked for provision of a Molly-Budtz Olsen PhD fellowship to AHH. We would like to thank Alina Tollenaere, Dongxue Li, Ong Cu, and Candice Jones for technical assistance, and Dr Dugald Reid for helping with preliminary investigations.

3.7 References

Araya T, Miyamoto M, Wibowo J, Suzuki A, Kojima S, Tsuchiya YN, Sawa S, Fukuda H, Von Wirén N, Takahashi H. 2014. CLE-CLAVATA1 peptide-receptor signaling module regulates the expansion of plant root systems in a nitrogen-dependent manner. *Proceedings of the National Academy of Sciences* **111**(5): 2029-2034.

Chen SK, Kurdyukov S, Kereszt A, Wang XD, Gresshoff PM, Rose R. 2009. The association of homeobox gene expression with stem cell formation and morphogenesis in cultured *Medicago truncatula*. *Planta* **230**(4): 827-840.

Chen S, Lang P, Chronis D, Zhang S, De Jong WS, Mitchum MG, Wang X. 2015. *In planta* processing and glycosylation of a nematode CLAVATA3/ENDOSPERM SURROUNDING REGION-like effector and its interaction with a host CLAVATA2-like receptor to promote parasitism. *Plant Physiology* **167**(1): 262-272.

Cock JM, McCormick S. 2001. A large family of genes that share homology with *CLAVATA3*. *Plant Physiology* **126**(3): 939-942.

Depuydt S, Rodriguez-Villalon A, Santuari L, Wyser-Rmili C, Ragni L, Hardtke CS. 2013. Suppression of Arabidopsis protophloem differentiation and root meristem

growth by CLE45 requires the receptor-like kinase BAM3. *Proceedings of the National Academy of Sciences* **110**(17): 7074-7079.

Endo S, Shinohara H, Matsubayashi Y, Fukuda H. 2013. A novel pollen-pistil interaction conferring high-temperature tolerance during reproduction via CLE45 signaling. *Current Biology* **23**(17): 1670-1676.

Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE, Gresshoff PM. 2010. Molecular analysis of legume nodule development and autoregulation. *Journal of Integrative Plant Biology* **52**(1): 61-76.

Ferguson BJ, Li D, Hastwell AH, Reid DE, Li Y, Jackson SA, Gresshoff PM. 2014. The soybean (*Glycine max*) nodulation-suppressive CLE peptide, GmRIC1, functions interspecifically in common white bean (*Phaseolus vulgaris*), but not in a supernodulating line mutated in the receptor PvNARK. *Plant Biotechnology Journal* **12**(8): 1085-1097.

Ferguson BJ, Mathesius U. 2014. Phytohormone regulation of legume-rhizobia interactions. *Journal of Chemical Ecology* **40**(7): 770-790.

Fiers M, Golemiec E, Xu J, van der Geest L, Heidstra R, Stiekema W, Liu CM. 2005. The 14–amino acid CLV3, CLE19, and CLE40 peptides trigger consumption of the root meristem in Arabidopsis through a CLAVATA2-dependent pathway. *The Plant Cell* **17**(9): 2542-2553.

Fiume E, Fletcher JC. 2012. Regulation of Arabidopsis embryo and endosperm development by the polypeptide signaling molecule CLE8. *The Plant Cell* **24**(3): 1000-1012.

Fletcher JC, Brand U, Running MP, Simon R, Meyerowitz EM. 1999. Signaling of cell fate decisions by CLAVATA3 in Arabidopsis shoot meristems. *Science* **283**(5409): 1911-1914.

Fujisawa T, Okamoto S, Katayama T, Nakao M, Yoshimura H, Kajiya-Kanegae H, Yamamoto S, Yano C, Yanaka Y, Maita H. 2014. CyanoBase and RhizoBase: databases of manually curated annotations for cyanobacterial and rhizobial genomes. *Nucleic Acids Research* **42**(D1): D666-D670.

Funayama-Noguchi S, Noguchi K, Yoshida C, Kawaguchi M. 2011. Two CLE genes are induced by phosphate in roots of *Lotus japonicus*. *Journal of Plant Research* **124**(1): 155-163.

Gaillochet C, Daum G, Lohmann JU. 2015. O Cell, Where Art Thou? The mechanisms of shoot meristem patterning. *Current Opinion in Plant Biology* **23**: 91-97.

Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins MR, Appel RD, Bairoch A. 2005. Protein identification and analysis tools on the ExPASy server. In: Walker J, ed. The proteomics protocols handbook. Clifton, NJ: Humana Press, 571–607.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**(D1): D1178-D1186.

Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. 2010. A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Research* **38**: W695-W699.

Grienenberger E, Fletcher JC. 2015. Polypeptide signaling molecules in plant development. *Current Opinion in Plant Biology* **23**: 8-14.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**(5): 696-704.

Handa Y, Nishide H, Takeda N, Suzuki Y, Kawaguchi M, Saito K. 2015. RNA-seq transcriptional profiling of an arbuscular mycorrhiza provides insights into regulated and coordinated gene expression in *Lotus japonicus* and *Rhizophagus irregularis*. *Plant and Cell Physiology* **56**(8): 1490-1511.

Hastwell AH, Gresshoff PM, Ferguson BJ. 2015. The structure and activity of nodulation-suppressing CLE peptide hormones of legumes. *Functional Plant Biology* **42**(3): 229-238.

Hirakawa Y, Kondo Y, Fukuda H. 2010. TDIF peptide signaling regulates vascular stem cell proliferation via the *WOX4* homeobox gene in Arabidopsis. *The Plant Cell* **22**(8): 2618-2629.

Hirakawa Y, Kondo Y, Fukuda H. 2011. Establishment and maintenance of vascular cell communities through local signaling. *Current Opinion in Plant Biology* **14**(1): 17-23.

Hobe M, Müller R, Grünewald M, Brand U, Simon R. 2003. Loss of CLE40, a protein functionally equivalent to the stem cell restricting signal CLV3, enhances root waving in Arabidopsis. *Development Genes and Evolution* **213**(8): 371-381.

Ito Y, Nakanomyo I, Motose H, Iwamoto K, Sawa S, Dohmae N, Fukuda H. 2006. Dodeca-CLE peptides as suppressors of plant stem cell differentiation. *Science* **313**(5788): 842-845.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**(12): 1647-1649.

Kinoshita A, Nakamura Y, Sasaki E, Kyozuka J, Fukuda H, Sawa S. 2007. Gain-of-function phenotypes of chemically synthetic CLAVATA3/ESR-related (CLE) peptides in *Arabidopsis thaliana* and *Oryza sativa*. *Plant and Cell Physiology* **48**(12): 1821-1825.

Kondo T, Yokomine K, Nakagawa A, Sakagami Y. 2011. Analogs of the CLV3 peptide: synthesis and structure–activity relationships focused on proline residues. *Plant and Cell Physiology* **52**(1): 30-36.

Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**(1): 105-132.

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* **40**(D1): D1202-D1210.

Lavin M, Herendeen PS, Wojciechowski MF. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Systematic Biology* **54**(4): 575-594.

Libault M, Farmer A, Brechenmacher L, Drnevich J, Langley RJ, Bilgin DD, Radwan O, Neece DJ, Clough SJ, May GD. 2010a. Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. *Plant Physiology* **152**(2): 541-552.

Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, He J, Xu D, May G, Stacey G. 2010b. An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *The Plant Journal* **63**(1): 86-99.

Lin J-Y, Stupar RM, Hans C, Hyten DL, Jackson SA. 2010. Structural and functional divergence of a 1-Mb duplicated region in the soybean (*Glycine max*) genome and comparison to an orthologous region from *Phaseolus vulgaris*. *The Plant Cell* **22**(8): 2545-2561.

Lim CW, Lee YW, Hwang CH. 2011. Soybean nodule-enhanced CLE peptides in roots act as signals in GmNARK-mediated nodulation suppression. *Plant and Cell Physiology* **52**(9): 1613-1627.

Lim CW, Lee YW, Lee SC, Hwang CH. 2014. Nitrate inhibits soybean nodulation by regulating expression of CLE genes. *Plant Science* **229**: 1-9.

Matsubayashi Y. 2014. Posttranslationally modified small-peptide signals in plants. *Annual Review of Plant Biology* **65**: 385-413.

McClean PE, Lavin M, Gepts P, Jackson SA 2008. *Phaseolus vulgaris*: a diploid model for soybean. *Genetics and Genomics of Soybean*: Vol. 2. New York: Springer, 55–76.

McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. 2013. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Research* **41**(W1): W597-W600.

Meng L, Buchanan BB, Feldman LJ, Luan S. 2012. A putative nuclear CLE-like (CLEL) peptide precursor regulates root growth in Arabidopsis. *Molecular Plant* **5**(4): 955-957.

Mitchum MG, Hussey RS, Baum TJ, Wang X, Elling AA, Wubben M, Davis EL. 2013. Nematode effector proteins: an emerging paradigm of parasitism. *New Phytologist* **199**(4): 879-894.

Mortier V, Den Herder G, Whitford R, Van de Velde W, Rombauts S, D'haeseleer K, Holsters M, Goormachtig S. 2010. CLE peptides control *Medicago truncatula* nodulation locally and systemically. *Plant Physiology* **153**(1): 222-237.

Mortier V, De Wever E, Vuylsteke M, Holsters M, Goormachtig S. 2012. Nodule numbers are governed by interaction between CLE peptides and cytokinin signaling. *The Plant Journal* **70**(3): 367-376.

Mortier V, Fenta BA, Martens C, Rombauts S, Holsters M, Kunert K, Goormachtig S. 2011. Search for nodulation-related CLE genes in the genome of *Glycine max*. *Journal of Experimental Botany* **62**(8): 2571-2583.

Ni J, Clark SE. 2006. Evidence for functional conservation, sufficiency, and proteolytic processing of the CLAVATA3 CLE domain. *Plant Physiology* **140**(2): 726-733.

Ni J, Guo Y, Jin H, Hartsell J, Clark SE. 2011. Characterization of a CLE processing activity. *Plant Molecular Biology* **75**(1-2): 67-75.

O'Rourke JA, Iniguez LP, Fu F, Bucciarelli B, Miller SS, Jackson SA, McClean PE, Li J, Dai X, Zhao PX. 2014. An RNA-Seq based gene expression atlas of the common bean. *BMC Genomics* **15**(1): 866.

Oelkers K, Goffard N, Weiller GF, Gresshoff PM, Mathesius U, Frickey T. 2008. Bioinformatic analysis of the CLE signaling peptide family. *BMC Plant Biology* **8**(1): 1.

Ohyama K, Shinohara H, Ogawa-Ohnishi M, Matsubayashi Y. 2009. A glycopeptide regulating stem cell fate in *Arabidopsis thaliana*. *Nature Chemical Biology* **5**(8): 578-580.

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**(1): 67-77.

Okamoto S, Shinohara H, Mori T, Matsubayashi Y, Kawaguchi M. 2013. Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. *Nature Communications* **4**: 2191.

Olsen AN, Skriver K. 2003. Ligand mimicry? Plant-parasitic nematode polypeptide with similarity to CLAVATA3. *Cell* **108**: 489-500.

Opsahl-Ferstad HG, Deunff EL, Dumas C, Rogowsky PM. 1997. *ZmEsr*, a novel endosperm-specific gene expressed in a restricted region around the maize embryo. *The Plant Journal* **12**(1): 235-246.

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L. 2007. The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Research* **35**: D883-D887.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**(10): 785-786.

Reid DE, Ferguson BJ, Gresshoff PM. 2011a. Inoculation-and nitrate-induced CLE peptides of soybean control NARK-dependent nodule formation. *Molecular Plant-Microbe Interactions* **24**(5): 606-618.

Reid DE, Ferguson BJ, Hayashi S, Lin Y-H, Gresshoff PM. 2011b. Molecular mechanisms controlling legume autoregulation of nodulation. *Annals of Botany* **108**(5): 789-795.

Reid DE, Li D, Ferguson BJ, Gresshoff PM. 2013. Structure–function analysis of the Gm RIC1 signal peptide and CLE domain required for nodulation control in soybean. *Journal of Experimental Botany* **64**(6): 1575-1585.

Replogle A, Wang J, Bleckmann A, Hussey RS, Baum TJ, Sawa S, Davis EL, Wang X, Simon R, Mitchum MG. 2011. Nematode CLE signaling in Arabidopsis requires CLAVATA2 and CORYNE. *The Plant Journal* **65**(3): 430-440.

Rodriguez-Villalon A, Gujas B, Kang YH, Breda AS, Cattaneo P, Depuydt S, Hardtke CS. 2014. Molecular genetic framework for protophloem formation. *Proceedings of the National Academy of Sciences* **111**(31): 11551-11556.

Rojo E, Sharma VK, Kovaleva V, Raikhel NV, Fletcher JC. 2002. CLV3 is localized to the extracellular space, where it activates the Arabidopsis CLAVATA stem cell signaling pathway. *The Plant Cell* **14**(5): 969-977.

Roulin A, Auer PL, Libault M, Schlueter J, Farmer A, May G, Stacey G, Doerge RW, Jackson SA. 2013. The fate of duplicated genes in a polyploid plant genome. *The Plant Journal* **73**(1): 143-153.

Sawa S, Kinoshita A, Nakanomyo I, Fukuda H. 2006. CLV3/ESR-related (CLE) peptides as intercellular signaling molecules in plants. *The Chemical Record* **6**(6): 303-310.

Scheible W-R, Morcuende R, Czechowski T, Fritz C, Osuna D, Palacios-Rojas N, Schindelasch D, Thimm O, Udvardi MK, Stitt M. 2004. Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of Arabidopsis in response to nitrogen. *Plant Physiology* **136**(1): 2483-2499.

Severin AJ, Woody JL, Bolon Y-T, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE. 2010. RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. *BMC plant biology* **10**(1): 160.

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU. 2005. A gene expression map of *Arabidopsis thaliana* development. *Nature Genetics* **37**(5): 501-506.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J. 2010. Genome sequence of the palaeopolyploid soybean. N*ature* **463**(7278): 178-183.

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C. 2014. A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics* **46**(7): 707-713.

Sharma VK, Ramirez J, Fletcher JC. 2003. The Arabidopsis CLV3-like (CLE) genes are expressed in diverse tissues and encode secreted proteins. *Plant Molecular Biology* **51**(3): 415-425.

Shinohara H, Matsubayashi Y. 2013. Chemical synthesis of Arabidopsis CLV3 glycopeptide reveals the impact of hydroxyproline arabinosylation on peptide conformation and activity. *Plant and Cell Physiology* **54**(3): 369-374.

Shinohara H, Matsubayashi Y. 2015. Reevaluation of the CLV3-receptor interaction in the shoot apical meristem: dissection of the CLV3 signaling pathway from a direct ligand-binding point of view. *The Plant Journal* **82**(2): 328-336.

Shinohara H, Moriyama Y, Ohyama K, Matsubayashi Y. 2012. Biochemical mapping of a ligand-binding domain within Arabidopsis BAM1 reveals diversified ligand recognition mechanisms of plant LRR-RKs. *The Plant Journal* **70**(5): 845-854.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7**(1): 539.

Song X-F, Guo P, Ren S-C, Xu T-T, Liu C-M. 2013. Antagonistic peptide technology for functional dissection of CLV3/ESR genes in Arabidopsis. *Plant Physiology* **161**(3): 1076-1085.

Song X-F, Yu D-L, Xu T-T, Ren S-C, Guo P, Liu C-M. 2012. Contributions of individual amino acid residues to the endogenous CLV3 function in shoot apical meristem maintenance in Arabidopsis. *Molecular Plant* **5**(2): 515-523.

Stahl Y, Wink RH, Ingram GC, Simon R. 2009. A signaling module controlling the stem cell niche in Arabidopsis root meristems. *Current Biology* **19**(11): 909-914.

Strabala TJ, Phillips L, West M, Stanbra L. 2014. Bioinformatic and phylogenetic analysis of the CLAVATA3/EMBRYO-SURROUNDING REGION (CLE) and the CLE-LIKE signal peptide genes in the Pinophyta. *BMC Plant Biology* **14**(1): 47.

Thomas BC, Pedersen B, Freeling M. 2006. Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome research* **16**(7): 934-946.

van der Graaff E, Laux T, Rensing SA. 2009. The WUS homeobox-containing (WOX) protein family. *Genome biology* **10**(12): 1.

Wang X, Mitchum MG, Gao B, Li C, Diab H, Baum TJ, Hussey RS, Davis EL. 2005. A parasitism gene from a plant-parasitic nematode with function similar to CLAVATA3/ESR (CLE) of *Arabidopsis thaliana*. *Molecular Plant Pathology* **6**(2): 187-191.

Whitford R, Fernandez A, De Groodt R, Ortega E, Hilson P. 2008. Plant CLE peptides from two distinct functional classes synergistically induce division of vascular cells. *Proceedings of the National Academy of Sciences* **105**(47): 18625-18630.

Wong CE, Singh MB, Bhalla PL. 2013. Spatial expression of *CLAVATA3* in the shoot apical meristem suggests it is not a stem cell marker in soybean. *Journal of Experimental Botany* **64**(18): 5641-5649.

Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H. 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**(7378): 520-524.

Chapter 4

CLE peptide-encoding gene families in *Medicago truncatula* and *Lotus japonicus*, compared with those of soybean, common bean and Arabidopsis

Preface

This chapter has been published in Scientific Reports (2017, doi:10.1038/s41598-017-09296-w). This article identifies and characterises the CLE peptide-encoding genes in two model legumes, *Medicago truncatula* and *Lotus japonicus*. As part of this study, novel candidates were identified as having potential roles in plant development aspects, including nodulation control, with some published and preliminary investigations complementing the bioinformatic findings from this study (de Bang et al. 2017; Harding, Hastwell, Gresshoff and Ferguson, unpublished). Additional supplementary tables and figures may be found online at https://www.nature.com/articles/s41598-017-09296-w.

4.1 Abstract

CLE peptide hormones are critical regulators of many cell proliferation and differentiation mechanisms in plants. These 12-13 amino acid glycosylated peptides play vital roles in a diverse range of plant tissues, including the shoot, root and vasculature. CLE peptides are also involved in controlling legume nodulation. Here, the entire family of CLE peptide-encoding genes was identified in *Medicago truncatula* (52) and *Lotus japonicus* (53), including pseudogenes and non-functional sequences that were identified. An array of bioinformatic techniques were used to compare and contrast these complete CLE peptide-encoding gene families with those of fellow legumes, *Glycine max* and *Phaseolus vulgaris*, in addition to the model plant *Arabidopsis thaliana*. This approach provided insight into the evolution of CLE peptide families and enabled us to establish putative *M. truncatula* and *L. japonicus* orthologues. This includes orthologues of nodulation-suppressing CLE peptides and *AtCLE40* that controls the stem cell population of the root apical meristem. A transcriptional meta-analysis was also conducted to help elucidate the function of the CLE peptide family members. Collectively, our analyses considerably increased the number of annotated CLE peptides in the model legume species, *M. truncatula* and *L*. *japonicus*, and substantially enhanced the knowledgebase of this critical class of peptide hormones.

4.2 Introduction

CLAVATA3/Endosperm Surrounding Region-related (CLE) peptides belong to a class of cysteine poor, post-translationally modified peptides that are derived from a prepropeptide (Tavormina et al. 2015; Hastwell et al. 2015a and 2015b). The mature CLE peptide is 12 to 13 amino acids long and those that have been structurally confirmed all possess a tri-arabinose moiety attached to a highly conserved hydroxylated central proline residue (Ohyama et al. 2009; Okamoto et al. 2013; Chen et al. 2015). They act as hormone-like signals (Ferguson and Mathesius 2014) and are perceived by class XI leucine-rich repeat receptor kinases (Endo et al. 2014). They are also unique to plants, with the exception of CLE peptide-encoding genes of the cyst-knot nematode (Olsen and Skriver 2003), which were likely acquired from plants via horizontal gene transfer (Chen et al. 2015; Replogle et al. 2011). CLE peptides have roles in regulating stem cell populations of various plant organs (Gaillochet and Lohmann 2015; Greb and Lohmann 2016). Prominent examples include CLAVATA3 (CLV3) in the shoot apical meristem (Clark et al. 1995; Gaillochet et al. 2015; Somssich et al. 2016), *AtCLE40* in the root apical meristem (Hobe et al. 2003; Sharma et al. 2003; Berckmans and Simon 2016), a number of legume-specific CLE peptides that suppress nodule organogenesis (Hastwell et al. 2015b; Reid e al. 2011), and a sub-class of highly conserved CLE peptides that regulate vascular differentiation (Sawa et al. 2006; Ito et al. 2006; Hirakawa et al. 2010; Etchells et al. 2016; De Rybel et al. 2016). Those of the cyst-knot nematode are thought to have a role in establishing the pathogen's feeding site (Mitchum et al. 2013).

Medicago truncatula and *Lotus japonicus* are model legume species that offer a number of molecular advantages to understanding aspects of legume development, as well as microbial and fungal symbioses (Verdier et al. 2013). However, only a few CLE peptide-encoding genes have been functionally characterised in these species to date. This includes *LjCLE-RS1*, *LjCLE-RS2*, *LjCLE-RS3*, *MtCLE12* and *MtCLE13*, which are involved in nodulation regulation (Okamoto et al. 2009; Mortier et al. 2010; Okamoto et al. 2013; Hastwell et al. 2015b; Nishida et al. 2016). Others include *LjCLE7*, *LjCLE15*, *LjCLE19* *LjCLE20*, *LjCLE24* and *LjCLE29*, that are up-regulated in response to phosphate and/or mycorrhizae (Funayama-Noguchi et al. 2011; Handa e al. 2015); and *MtCLV3* (Chen at al. 2009) and *LjCLV3* (Okamoto et al. 2009; Okamoto et al. 2011), the orthologues of the most thoroughly characterised CLE peptide-encoding gene, *AtCLV3* (Somssich et al. 2016). In *M. truncatula*, the likely orthologues of the Treachery Element Inhibitory Factor

(TDIF) encoding genes, *AtCLE41*, *AtCLE42* and *AtCLE44* (Etchells et al. 2016; De Rybel et al. 2016), have also been identified (Hastwell et al. 2015a).

Recent genomic and bioinformatic advances allow for the identification of entire peptide families. This is extremely helpful for comparable genomic studies and for advancing the important functional characterisation of individual peptide members. Here, we used a genome-wide approach to identify the complete CLE peptide-encoding gene families of *M. truncatula* and *L. japonicus*. Comparative bioinformatic approaches were used to assist in identifying orthologous genes between these, and other plant species, as well as in the categorisation and functional characterisation of these critical peptide-encoding genes.

4.3 Results

4.3.1 Identification of CLE peptide-encoding genes in *L. japonicus* and *M. truncatula*

A thorough genome-wide search of the *M. truncatula* and *L. japonicus* genomes was conducted to identify the complete CLE peptide-encoding gene families of these species. Multiple BLAST searches identified 52 and 53 CLE peptide-encoding genes in each of the two species respectively (Figs 4.1–4.3, Table 4.1). Initial BLAST and TBLASTN queries used sequences of known soybean and *A. thaliana* CLE peptide-encoding genes and prepropeptides (Hastwell et al. 2015a) to ensure all genes of interest were captured. The resulting identified sequences were verified and false-positives removed from further analyses. Additional CLE peptide-encoding genes were identified by BLAST and TBLASTN reciprocal searches of the *M. truncatula* and *L. japonicus* genomes using the sequences identified in the initial searches. A number of the genes identified are reported here for the first time, with the nomenclature of the newly discovered genes consistent with previously identified CLE peptide-encoding genes (Figs 4.1–4.3, Table 4.1). A recent study published after our searches were conducted included 20 *M. truncatula* CLE peptide-encoding genes (Goad et al. 2016), but no nomenclature was given as species-specific analyses were not conducted. A complete listing of all CLE peptide encoding gene family members from *M. truncatula* and *L. japonicus* is provided in Supplementary Table S1.

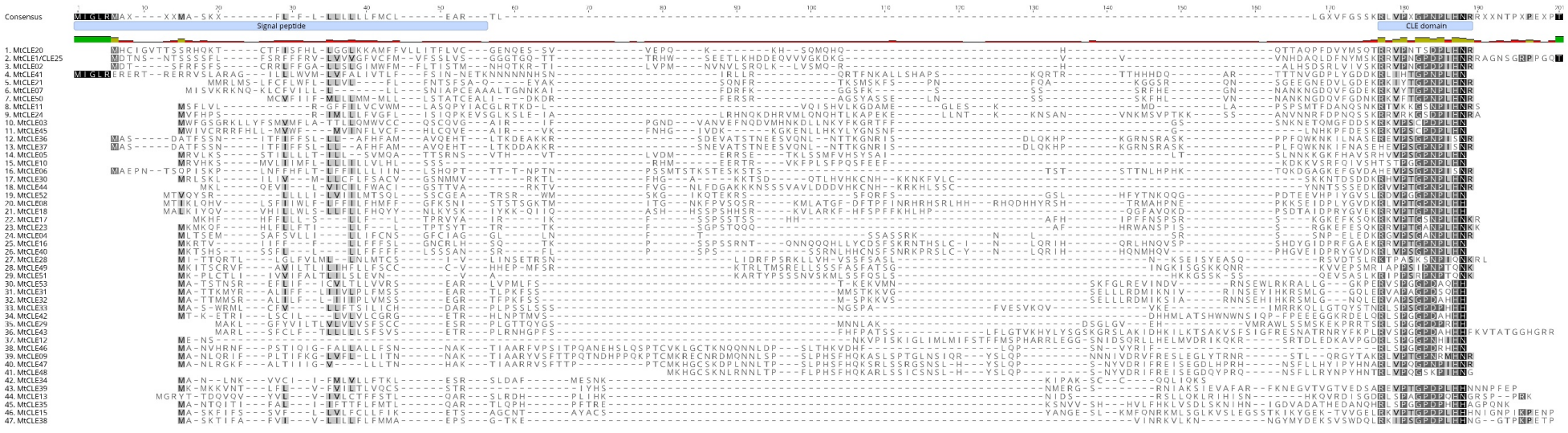


Figure 4.1 Multiple sequence alignment of *Medicago truncatula* CLE prepropeptides. The sequences show high similarity, as indicated by darker shading, in the signal peptide and CLE domains. Not shown are the multi-CLE domain containing prepropeptides (MtCLE14, MtCLE26, MtCLE27 and MtCLE22, see Fig. 4.3) and MtCLE19, which has a premature stop codon very early in the prepropeptide (see Fig. 4.4). MtCLE34 is a likely pseudogene without a functional CLE domain. The signal peptide approximate location and CLE domain is shown on the consensus sequence.

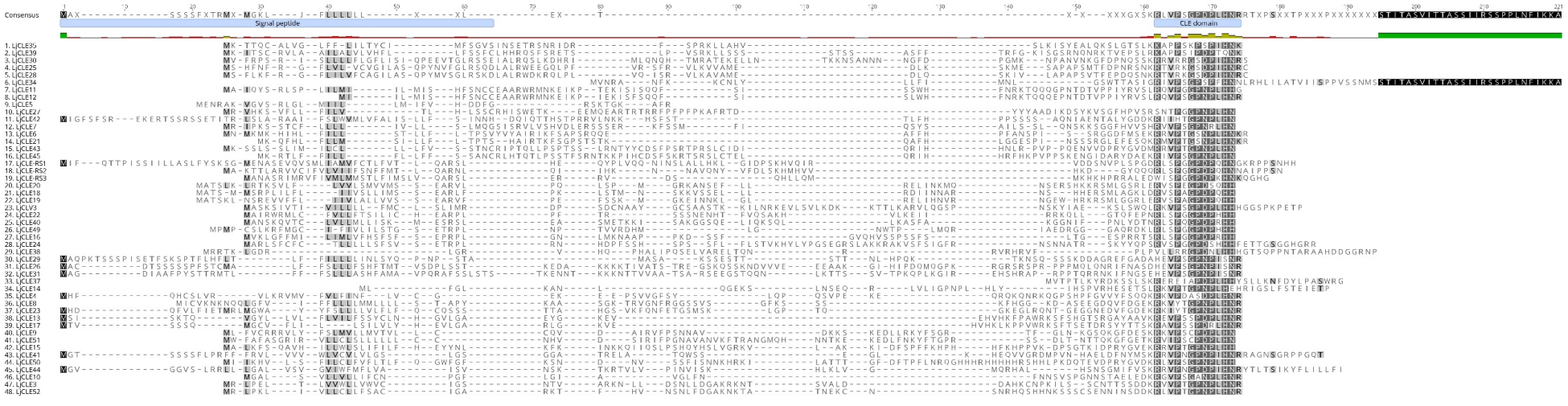


Figure 4.2 Multiple sequence alignment of *Lotus japonicus* CLE prepropeptides. As with the *M. truncatula* sequences (Fig. 4.1), the *L. japonicus* sequences show high similarity in the signal peptide and CLE domains, as indicated by darker shading. Not shown are the multi-CLE domain containing prepropeptides (LjCLE32, LjCLE33, LjCLE46 and LjCLE47; see Fig. 4.3) and LjCLE48, the truncated *L. japonicus* AtCLE40 orthologue as it shows very little amino acid conservation. LjCLE5 is a likely pseudogene without a functional CLE domain (see Fig. 4.5). The signal peptide approximate location and CLE domain is shown on the consensus sequence.

Table 4.1 Name, ID and various features of CLE genes in *Medicago truncatula* and *Lotus japonicus*.

Name	Phytozome V11 ID (Mtv4)	Pre-propeptide length	Chromosome location	Orientation	Predicted intron	SP cleavage site[a]
MtCLE1/CLE25	Medtr5g037140	110	chr5:16209190..16209492	forward	No	39
MtCLE02	Medtr6g009390	90	chr6:2758371..2758643	reverse	No	34
MtCLE03	Medtr1g110820	99	chr1:50033208..50033507	forward	No	31
MtCLE04	Medtr5g014860	66	chr5:5053422..5053622	reverse	No	26
MtCLE05	Medtr1g100733	84	chr1:45667039..45667293	forward	No	25
MtCLE06	Medtr7g058790	99	chr7:21150139..21150438	reverse	No	35
MtCLE07	Medtr7g089320	85	chr7:34939800..34940057	forward	No	28
MtCLE08	Medtr8g076990	120	chr8:32679901..32680263	reverse	No	27
MtCLE09	Medtr7g084110	137	chr7:32430490..32430903	reverse	No	31
MtCLE10	Medtr6g054925	74	chr6:19620161..19620385	forward	No	27
MtCLE11	Medtr3g037730	83	chr3:13874060..13874311	reverse	No	27
MtCLE12	Medtr4g079630	81	chr4:30800344..30800589	forward	No	29
MtCLE13	Medtr4g079610	84	chr4:30793797..30794051	forward	No	27
MtCLE14	Medtr7g084100	221	chr7:32428499..32429164	reverse	No	28
MtCLE15	Medtr2g087170[b]	100	chr2:36639989..36640517	forward	Yes	26
MtCLE16	Medtr5g043830	101	chr5:19252630..19252935	reverse	No	20
MtCLE17	Medtr5g085990	72	chr5:37176921..37177139	reverse	No	21
MtCLE18	Medtr1g093800	104	chr1:42088638..42088952	forward	No	27
MtCLE19	unannotated	[c]	chr7:46832505..46832810	forward	No	ND
MtCLE20	Medtr1g018700	91	chr1:5449791..5450066	forward	No	43
MtCLE21	Medtr5g089080	84	chr5:38716369..38716623	forward	No	27
MtCLE22	Medtr2g087180	181	chr2:36645611..36646743	reverse	Yes	26
MtCLE23	Medtr3g080060	72	chr3:36207269..36207487	forward	No	24
MtCLE24	Medtr5g040640	108	chr5:17857512..17857838	reverse	No	27
MtCLE26	unannotated	117	chr1:27528684..27529037	forward	No	22
MtCLE27	Medtr1g062850	116	chr1:27531440..27531790	forward	No	22
MtCLE28	Medtr1g106920	89	chr1:48060869..48061138	forward	No	23

Name	Phytozome V11 ID (Mtv4)	Pre-propeptide length	Chromosome location	Orientation	Predicted intron	SP cleavage site[a]
MtCLE29	Medtr2g015445	78	chr2:4575877..4576113	forward	No	24
MtCLE30	Medtr2g038665	81	chr2:16905712..15905957	reverse	No	22
MtCLE31	Medtr2g078130	85	chr2:32437536..32437793	reverse	No	26
MtCLE32	Medtr2g078140	83	chr2:32444505..32444756	reverse	No	22
MtCLE33	Medtr2g078160	75	chr2:32460085..32460212	reverse	No	22
MtCLE34	Medtr2g091120	49	chr2:39236945..39237094	forward	No	23
MtCLE35	Medtr2g091125	92	chr2:39246810..39247088	forward	No	24
MtCLE36	Medtr2g437780	108	chr2:14915100..14915427	forward	No	27
MtCLE37	Medtr2g437800	108	chr2:14923591..14923917	forward	No	27
MtCLE38	Medtr4g051618	75	chr4:18583841..18584193	reverse	Yes	22
MtCLE39	Medtr4g066100	82	chr4:24906117..24906560	reverse	Yes	21
MtCLE40	Medtr4g082920	92	chr4:32235471..32235749	reverse	No	25
MtCLE41	Medtr4g084520	114	chr4:32892006..32892350	reverse	No	32
MtCLE42	Medtr4g087850	74	chr4:34563502..34563726	reverse	No	23
MtCLE43	unannotated	117	chr4:34572155..34572508	reverse	No	22
MtCLE44	Medtr4g126940	89	chr4:52604820..52605086	reverse	No	19
MtCLE45	Medtr5g056935	84	chr5:23427594..23427848	reverse	No	32
MtCLE46	Medtr7g084130	98	chr7:32437149..32437445	reverse	No	27
MtCLE47	unannotated	126	chr7:32590148..32590528	reverse	No	28
MtCLE48	unannotated	87	chr7:32594246..32594509	reverse	No	ND
MtCLE49	Medtr7g093050	92	chr7:36960868..36961146	reverse	No	27
MtCLE50	Medtr7g094080	81	chr7:37442023..37442268	forward	No	24
MtCLE51	Medtr8g042980	78	chr8:16651803..16652039	forward	No	20
MtCLE52	Medtr8g096970	89	chr8:40711987..40712256	reverse	No	25
MtCLE53	Medtr8g463700	82	chr8:22459174..22469422	forward	No	26
LjCLE3	Lj4g3v2140240.1	81	chr4:29668322..29668567	forward	No	19
LjCLE4	Lj5g3v0296280.1	87	chr5:2776799..2777062	reverse	No	28
LjCLE5	Unannotated	37	chr0:191639..191752	reverse	No	14
LjCLE6	Lj2g3v2904560.1	77	chr2:37869993..37870226	reverse	No	28

Name	Phytozome V11 ID (Mtv4)	Pre-propeptide length	Chromosome location	Orientation	Predicted intron	SP cleavage site[a]
LjCLE7	Lj5g3v2013980.1	85	chr5:28436402..28436659	forward	No	26
LjCLE8	Lj1g3v4106880.1	90	chr1:48730186..48730458	forward	No	29
LjCLE9	Lj2g3v1155200.1	84	chr2:18195792..18196046	reverse	No	30
LjCLE10	Lj2g3v1984000.1	58	chr2:28901709..28901885	reverse	No	20
LjCLE11	Lj4g3v2917660.1	91	chr4:38847122..38847397	reverse	No	28
LjCLE12	unannotated	78	chr4:38846082..38846318	reverse	No	ND
LjCLE13	Lj5g3v1494620.1	77	chr5:21653524..21653754	reverse	No	31
LjCLE14	Lj3g3v1261020.1	71	chr3:16238530..16238742	forward	No	ND
LjCLE15	Lj5g3v1789230.1	104	chr5:25374552..25374863	forward	No	27
LjCLE16	Lj6g3v1996000.1	75	chr6:23228716..23228940	reverse	No	21
LjCLE17	Lj1g3v4931750.1	74	chr1:60060788..60061012	forward	No	27
LjCLE18	Lj3g3v1063710.1	80	chr3:14363293..14363535	reverse	No	26
LjCLE19	Lj3g3v0428680.1	84	chr3:4038516..4038770	reverse	No	28
LjCLE20	Lj3g3v0428740.1	83	chr3:4052954..4053205	reverse	No	27
LjCLE21	Lj6g3v1055570.1	73	chr6:12069820..12070041	reverse	No	22
LjCLE22	Lj0g3v0114139.1	76	chr0:49962922..49963152	forward	No	23
LjCLE23	Lj0g3v0005899.1	96	chr0:2220612..2220902	reverse	No	35
LjCLE24	Lj4g3v0496580.1	110	chr4:8347397..8347729	forward	No	22
LjCLE25	Lj4g3v1635250.1	94	chr4:24032504..24032788	reverse	No	27
LjCLE26	Lj4g3v0189810.1	122	chr4:2377271..2377630	forward	No	33
LjCLE27	Lj2g3v0276540.1	86	chr2:4685035..4685295	forward	No	21
LjCLE28	Lj1g3v0492090.1	95	chr1:6477403..6477690	forward	No	20
LjCLE29	Lj2g3v1389560.1	99	chr2:22031058..22031357	forward	No	40
LjCLE30	Lj2g3v1277900.1	109	chr2:20511088..20511417	forward	No	23
LjCLE31	Lj6g3v1415960.1	124	chr6:16797110..16797364	reverse	No	31
LjCLE32	unannotated	274	chr1:45492391..45493215	reverse	No	27
LjCLE33	Lj3g3v1314940.1	347	chr3:17115763..17116806	forward	No	ND
LjCLE34	Lj3g3v2248290b	89	chr3:126894445..126894714	forward	No	26
LjCLE35	unannotated	79	chr5:30013235..30013474	forward	No	19

Name	Phytozome V11 ID (Mtv4)	Pre-propeptide length	Chromosome location	Orientation	Predicted intron	SP cleavage site[a]
LjCLE37	unannotated	44	chr6:16475826..16475960	reverse	No	ND
LjCLE38	Lj1g3v4241120.1	77	chr1:50124894..50125127	reverse	No	ND
LjCLE39	Lj1g3v4317570.1	93	chr1:50801119..50801398	reverse	No	29
LjCLE40	Lj3g3v2848710.1	80	chr3:35016355..35016597	forward	No	22
LjCLE41	Lj2g3v1354640.1	98	chr2:21357987..21358244	forward	No	32
LjCLE42	Lj4g3v0643890.1	111	chr4:10320058..10320393	forward	No	41
LjCLE43	unannotated	96	chr0:78808582..78808872	forward	No	24
LjCLE44	Lj2g3v1022600.3	100	chr2:16174090..16174392	reverse	No	29
LjCLE45	Lj2g3v1265080.1	96	chr2:20113738..20114028	forward	No	21
LjCLE46	Lj3g3v1314910.1	285	chr3:17104803..17105660	reverse	No	27
LjCLE47	Lj3g3v1314920.1	250	chr3:17108302..17109054	reverse	No	27
LjCLE48	unannotated	[c]	chr3:40213173..40213683	forward	Yesc	ND
LjCLE49	Lj4g3v0496650.1	74	chr4:08364934..08365729	forward	No	26
LjCLE50	Lj4g3v1785920.1	102	chr4:25243942..25244250	reverse	No	25
LjCLE51	Lj5g3v2193950.1	98	chr5:31824885..23948823	forward	No	30
LjCLE52	Lj6g3v1280830.1	82	chr6:15475896..15476144	reverse	No	19
LjCLE-RS1	Lj0g3v0000559.1	116	chr0:240796..241074	reverse	No	23
LjCLE-RS2	Lj3g3v2848800.1	81	chr3:35039780..35040025	forward	No	28
LjCLE-RS3	Lj3g3v2848810.1	72	chr3:35048895..35049113	forward	No	28
LjCLV3	Lj3g3v1239970.1	105	chr3:16182605..16182777	forward	Yes	24

[a]Signal peptide cleaved after noted residue. [b]Unannotated transcript variant. [c]Likely untranscribed pseudogene. ND, not detected.

Additional CLE peptide-encoding genes in both *L. japonicus* and *M. truncatula* were identified that contain multiple CLE domains; some of which are also reported here for the first time. These multi-CLE peptide domain encoding genes include *LjCLE32*, *LjCLE33, LjCLE46* and *LjCLE47* in *L. japonicus*; and *MtCLE14*, *MtCLE22*, *MtCLE26* and *MtCLE27* in *M. truncatula* (Fig. 4.3). *LjCLE32* and *LjCLE33* encode eight and nine putative CLE peptides respectively; *MtCLE22* encodes four putative CLE peptides; *MtCLE26* and *MtCLE27*encode three putative CLE peptides; whereas all others contain seven putative CLE peptide domains (Fig. 4.3a; Supplementary Table S1). Interestingly, these multi-CLE domain containing genes contain repeating motifs of 24 to 35 amino acids, with each motif having a consistent length within their respective prepropeptide, with the sole exception of LjCLE33 which has varying motif lengths (Supplementary Table S2).

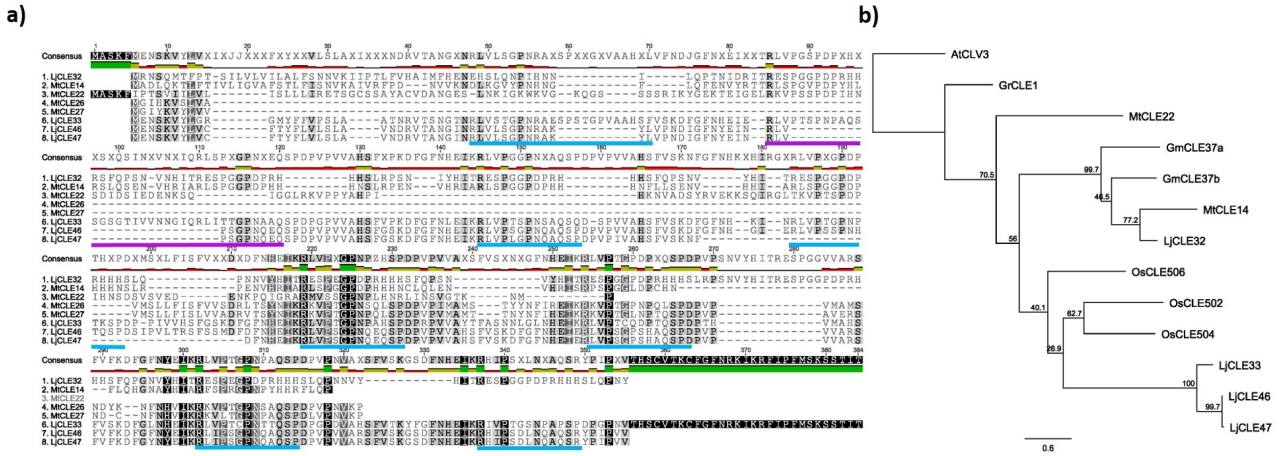


Figure 4.3 Multi-CLE domain prepropeptides. **(a)** Multiple sequence alignment of four *Lotus japonicus*(LjCLE32, LjCLE33, LjCLE46 and LjCLE47) and four *Medicago truncatula* (MtCLE14, MtCLE26, MtCLE27 and MtCLE22) multi-CLE domain containing prepropeptides (See Supplementary Table S3). Putative CLE domains are located above the blue and purple underlined regions. LjCLE21, LJCLE33 and MtCLE14 also have a second CLE domain present above the purple underlined region. **(b)** Phylogenetic tree of known multi-CLE domain prepropeptides in *L. japonicus*, *M. truncatula*, *Glycine max*, *Oryza sativa* and potato cysts nematode (*Globodera rostochiensis*), including AtCLV3 as an outgroup. The tree is shown with bootstrap confidence values as a percentage of 1,000 bootstraps.

Pseudogenes were also identified in both the *L. japonicus* and *M. truncatula* genomes. These genes include mutations where the CLE domain is not translated in frame, likely resulting in a non-functional gene. This includes the pseudogenes *MtCLE34*, which is annotated within the *M. truncatula* genome (Fig. 4.1, Table 4.1; Supplementary Fig. S1) and *MtCLE19* (Fig. 4.4). In *L. japonicus*, LjCLE5 (Figs 4.2 and 4.5, Table 4.1) and LjCLE48 are also unlikely to be functional (Fig. 4.6). These pseudogenes, and the genes containing

multiple CLE-domains, were excluded from the sequence characterisation studies detailed below because they fail to align well with the more typical single-CLE domain sequences.

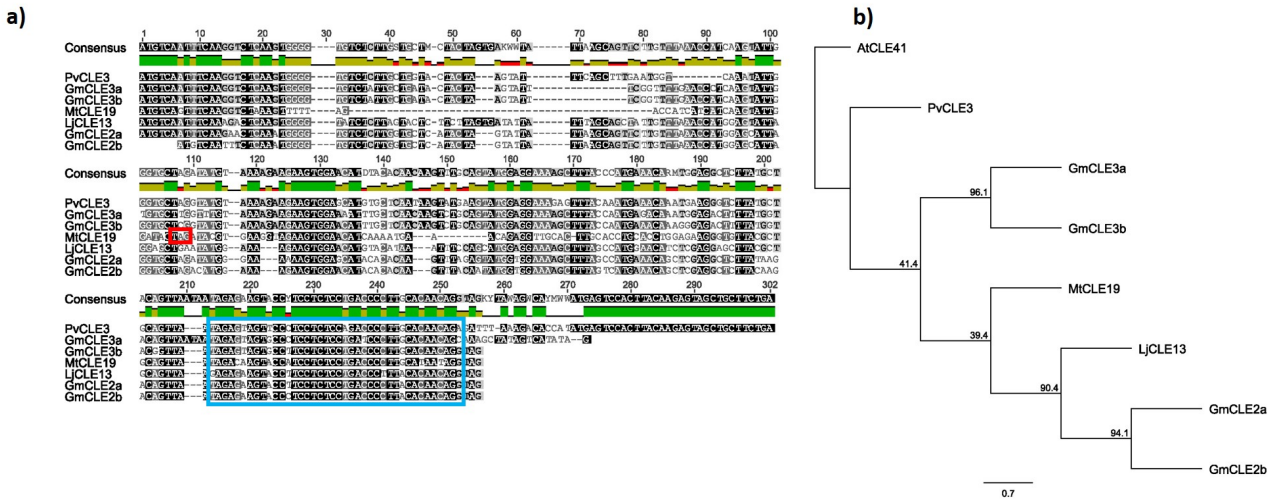


Figure 4.4 Genomic sequence characterisation of *MtCLE19*, the likely non-functional *M. truncatula* orthologue of *GmCLE2a*, *GmCLE2b*, and *LjCLE13*. **(a)** Multiple sequence alignment demonstrating that *MtCLE19* exhibits high similarity to *GmCLE2a*, *GmCLE2b*, and *LjCLE13*, with slightly less similarity to *GmCLE3a*, *GmCLE3b* and *PvCLE3*. The red box indicates a premature stop codon and the blue box indicates the CLE domain. Grey nucleotides are semi-conserved and black nucleotides are 100% conserved. **(b)** Phylogenetic tree with bootstrap confidence values expressed as a percentage of 1,000 bootstrap replications, using *AtCLE41* as an outgroup.

A BLAST search of the *L. japonicus* genome with the *LjCLE34* nucleotide sequence (first reported by Okamoto et al. (2009), identified two possible genes having two synonymous nucleotide changes that result in identical prepropeptides. These genes are located at chr3:27855838..27856107 and chr0:126894445..126894714, and interestingly, both are found within a larger predicted protein. It therefore appears that these two genes arose as a transposable element and subsequent duplication event, or they are the result of a genome sequencing error. Interestingly, the CLE domain of LjCLE34 is not located at the C-terminus of the prepropeptide but towards the centre, similar to that of AtCLE18, which has a C-terminal CLE-Like/Root Growth Factor/GOLVEN (CLEL/RGF/GLV) domain in addition to a CLE domain (Meng et al. 2012). LjCLE34 shares some homology at the C-terminus with AtCLE18 which includes the region of the CLEL/RGF/GLV domain (Supplementary Fig. S2).

CLE peptide-encoding genes of *M. truncatula* and *L. japonicus* are located across all chromosomes, with the greatest number located on chromosome two of *M. truncatula* (eleven) and chromosome three of *L. japonicus* (thirteen) (Table 4.1). There are five CLE

peptide-encoding genes of *L. japonicus* currently located on unassigned scaffolds (Table 4.1). The CLE prepropeptides of both species vary in length, with the average single-CLE domain prepropeptide being 88 residues in *L. japonicus* and 91 residues in *M. truncatula*. The multi-CLE domain prepropeptides of both species range from 116 to 347 amino acids.

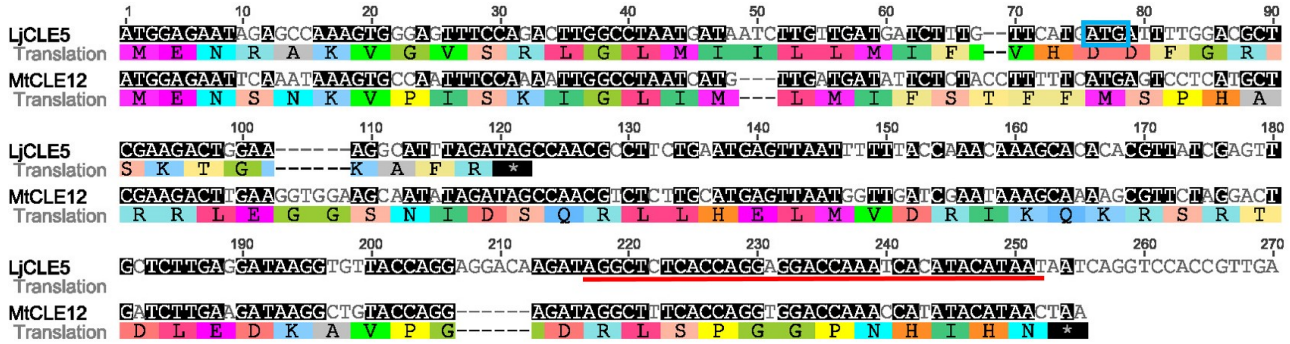


Figure 4.5 Sequence alignment of *LjCLE5* and *MtCLE12*. The *LjCLE5* translation start site corresponding to that of *MtCLE12* results in a truncated prepropeptide (denoted by the asterisk shaded in black). The blue box represents an alternative start codon that results in the CLE domain region of LjCLE5 (underlined in red) being translated in frame. Black highlighted nucleic acid bases are 100% conserved

Some CLE peptide-encoding genes appear directly in tandem within the genome. For example, on chromosome 2 of *M. truncatula*, *MtCLE31* is 6.7 Kb upstream of *MtCLE32*, which itself is 15.3 Kb upstream of *MtCLE33*. Also on chromosome 2, *MtCLE34* is 9.6 Kb upstream of *MtCLE35* and *MtCLE36* is 6.7 Kb upstream of *MtCLE37*. On chromosome 7, *MtCLE14*, *MtCLE09* and *MtCLE46* are all within 9 Kb, and *MtCLE47* is 3.7 Kb upstream of *MtCLE48*. On chromosome 4, *MtCLE12* and *MtCLE13*are not directly in tandem, but are only 6.3 Kb apart (Table 4.1). On chromosome 3 of *L. japonicus*, *LjCLE46*is 2.6 Kb apart from *LjCLE47*, which is 6.7 Kb upstream of *LjCLE33*. Also on chromosome 3, *LjCLE40*, *LjCLE-RS2* and *LjCLE-RS3* are within 24 Kb, and although not directly in tandem, *LjCLE19* and *LjCLE20* are only 14.2 Kb apart. On chromosome 4, *LjCLE11* and *LjCLE12* are only 0.8 Kb apart (Table 4.1). Interestingly, the genes appearing directly in tandem within the *L. japonicus* genome share >50% amino acid sequence similarity, while only some of the tandem gene pairs in *M. truncatula* exhibit more than a 50% level of similarity (Supplementary Table S3).

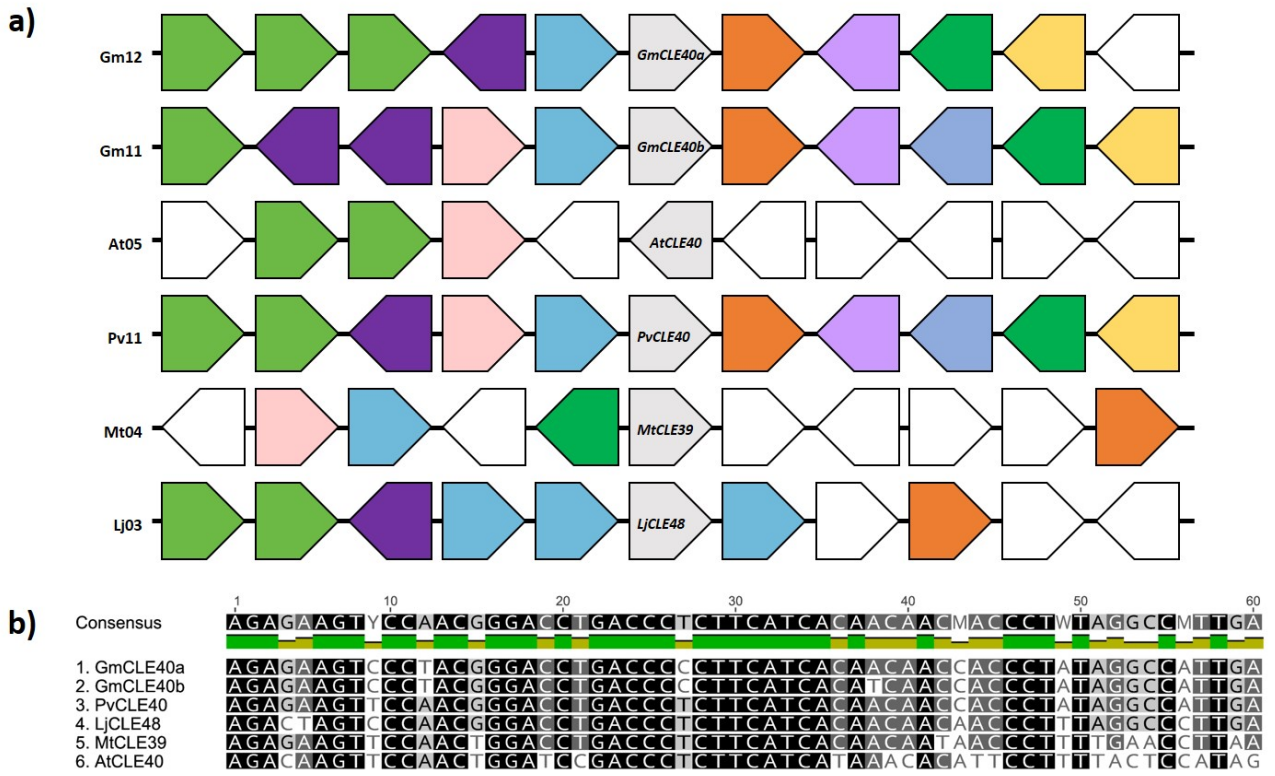


Figure 4.6 *AtCLE40* and orthologues in *Medicago truncatula*, *Phaseolus vulgaris*, and *Glycine max*, in addition to the truncated orthologue *in Lotus japonicus*, LjCLE48. **(a)** The genomic environment of each shows strong synteny. Arrows represent individual genes and their transcriptional direction in relation to *CLE40*. Similar colours represent genes from the same family, and are typically orthologous. **(b)** A multiple sequence alignment of the CLE40 domain coding region. Shading represents conservation amongst nucleotides with grey nucleotides semi-conserved and black nucleotides 100% conserved.

4.3.2 Identification of orthologous CLE peptide sequences

To identify gene orthologues of the *M. truncatula*and *L. japonicus* CLE prepropeptides, multiple sequence alignments were generated. Most orthologues were present in a 1:1 ratio between the two species (Supplementary Fig. S3). When no orthologue was evident, further BLAST searches were conducted in an attempt to identify one. In some instances, this yielded additional CLE peptide-encoding genes. Subsequent multiple sequence alignments with the CLE prepropeptides of *M. truncatula*, *L. japonicus*, soybean, common bean and *A. thaliana* were constructed (data not shown) and used to identify additional CLE peptide-encoding genes. All orthologous sequences identified are shown in Figs 4.1 and 4.2.

A multiple sequence alignment of the prepropeptides of *M. truncatula*, *L. japonicus*, common bean and *A. thaliana* was used to construct a phylogenetic tree (Supplementary

Fig. S3). Similar phylogenetic trees have been constructed using only the CLE domain of the prepropeptides; however, this domain is highly conserved and only 12-14 amino acids long, and hence alignments and trees constructed using only the conserved motif can be less informative. In contrast, the tree constructed here, using the entire prepropeptide sequences, allows for the identification of conserved residues within other domains that may relate to cleavage and other important facets of post-translational modification (Hastwell et al. 2015b).

4.3.3 Characterisation of M. truncatula and L. japonicus CLE prepropeptides

The domain structure of all CLE prepropeptides includes a hydrophobic signal peptide near the N-terminus, followed by a large variable region and a short but highly conserved CLE domain (with a multi-CLE domain occasionally present) and a small number (11 in *L. japonicus* and 8 in *M. truncatula*) that have a short C-terminal extension of unknown function (Figs 4.1 and 4.2; Hastwell et al. 2015b). The amino acid composition of all known CLE prepropeptides, across legume and non-legume species, is typically rich in lysine and serine, and poor in tyrosine, cysteine and tryptophan, with the latter being poorly represented in all plant proteins (Hastwell et al. 2015a). The CLE prepropeptides of *M. truncatula* and *L. japonicus* fit this amino acid profile (Supplementary Table S4). The CLE domain represents the functional peptide ligand, which is post-translationally cleaved and modified to 13 amino acids in AtCLV3 and LjCLE-RS1 (Ohyama et al. 2009; Shinohara et al. 2012; Okamoto et al. 2013; Chen et al. 2015). A total of 66% (*L. japonicus*) and 61% (*M. truncatula*) of the prepropeptides have an amino acid at the 13th residue, with the remaining having a stop codon at position 13, and thus being only 12 amino acids long. In both species, the amino acid most commonly found at position 13 is arginine (Figs 4.1 and 4.2, Supplementary Fig. S4).

An arginine residue is found at the start of 83% of *L. japonicus* and 87% of *M. truncatula* CLE domains. Although less common, a number of CLE domains also begin with a histidine, and this is conserved between orthologues of different species. Three of the four peptides beginning with a histidine in *A. thaliana* are Tracheary Differentiation Inhibitory Factors (TDIF) that are involved in vascular differentiation (Grienenberger and Fletcher 2015). *L. japonicus* and *M. truncatula* each have three CLE peptides beginning with a histidine (LjCLE26, LjCLE29 and LjCLE31, and MtCLE05, MtCLE06 and MtCLE37) that appear orthologous to the TDIF factors. However, they do not appear to have an

orthologue of the functionally unrelated fourth CLE peptide of Arabidopsis to begin with a histidine, AtCLE46, and its putative soybean orthologue, GmCLE13 (Hastwell et al. 2015a).

The most highly conserved CLE domain residues of *M. truncatula* are arginine at position one, glycine at position six and histidine at position 11, with all three present in 87% of the peptides (Fig. 4.1). Interestingly, the most conserved CLE domain residue of *L. japonicus* is histidine at position 11 (91%), with only three sequences having a serine at this position and one sequence having a glutamine (Fig. 4.2). Residues 1, 4, 6, 7, 9 and 11 are also highly conserved (>82%) in the CLE domain of both species (Figs 4.1 and 4.2, Supplementary Fig. S4). These residues are all considered critical for function except for the proline at position nine (Reid et al. 2013).

Outside of the CLE domain there is little conservation within the *L. japonicus* and *M. truncatula* CLE prepropeptide families (Figs 4.1 and 4.2). However, the signal peptide, which is predicted to either export the entire prepropeptide or the cleaved propeptide outside of the cell (Lim et al. 2011; Tavormina et al. 2015), contains a typical hydrophobic motif consisting of predominantly leucine and isoleucine (Figs 4.1 and 4.2). The size of the predicted signal peptide ranges from 19 to 43 residues (Table 4.1). Additionally, the truncated LjCLE5 prepropeptide has a predicted signal peptide cleavage site between residues 14 and 15 (Table 4.1).

Hastwell et al.(2015a) classified the CLE prepropeptides of soybean and common bean into seven distinct Groups (I to VII). The prepropeptides within each group show sequence conservation within and outside of the CLE domain. Based on the phylogenetic tree of the prepropeptides in *L. japonicus*, *M. truncatula*, *A. thaliana* and *P. vulgaris*, these groups remain conserved (Supplementary Fig. S3, Supplementary Table S5). This is especially evident with the Group VI CLE prepropeptides, which function in nodulation regulation, and Group III CLE prepropeptides, which show high sequence conservation with the Arabidopsis TDIF peptides, AtCLE41, AtCLE42 and AtCLE44 (Supplementary Fig. S3, Supplementary Table S5).

4.3.4 Identification of CLE40

A well characterised peptide, AtCLE40, has been shown to act as the root paralogue of AtCLV3 to regulate the stem cell population of the root apical meristem (Hobe et al. 2003;

Sharma et al. 2003; Berckmans and Simon 2016). Putative orthologues of *AtCLE40* have been identified in *M. truncatula*, *P. vulgaris* and *G. max* (*MtCLE39*, *PvCLE40*, *GmCLE40a* and *GmCLE40b*; Hastwell et al. 2015a). Interestingly, our BLAST searches using the *L. japonicus* genome failed to identify a CLE40 orthologue. However, a region on chromosome 3 (chr3:40213173..40213683) exhibits a very high level of sequence similarity to these *CLE40*orthologues, in addition to having a similar genomic environment to them (Fig. 4.6). All previously identified *CLV3* and *CLE40* orthologues contain two introns. The putative *L. japonicus CLE40* orthologue, identified here as *LjCLE48*, contains conserved predicted intron boundaries for the second intron, which correspond to the *CLE40* orthologues, but there are no predicted boundary sites for the first intron. Given this critical change at the 5′ end of *LjCLE48*, it appears unlikely that the resulting prepropeptide would produce a functional peptide product. This may suggest that another CLE peptide has evolved to perform the function of CLE40 in *L. japonicus*.

4.3.5 Nodulation CLE peptides

CLE genes in Group VI of soybean and common bean are known to respond to symbiotic bacteria, collectively called rhizobia, and act to control legume nodulation. The rhizobia-induced nodulation-suppressing CLE peptide encoding genes of *L. japonicus* and *M. truncatula*, known as LjCLE-RS1, LjCLE-RS2, LjCLE-RS3, MtCLE12 and MtCLE13 (Okamoto et al. 2009; Mortier et al. 2010; Reid et al. 2011; Ferguson et al. 2014; Nishida et al. 2016), cluster with these Group VI members of soybean and common bean (Hastwell et al. 2015a). Interestingly, two additional CLE prepropeptides of unknown function, called MtCLE35 and LjCLE5, also group closely (Supplementary Fig. S3). Okamoto et al. (2009) noted that LjCLE5 did not have a predicted signal peptide and that no expression could be detected. However, upstream of the previously predicted LjCLE5 start codon is another possible methionine (Fig. 4.5). The sequence following this alternative start codon corresponds closely with that of MtCLE12 (71.1% similarity), but the translation would result in a truncated protein prior to the CLE domain. Signal peptide prediction using SignalP (www.cbs.dtu.dk/services/SignalP/) suggests that there is a possible cleavage site at position 30 of the longer (but non-functional) LjCLE5. Interestingly, MtCLE35 contains the consensus sequence TLQAR, which is consistent with the nodulation-suppressing CLE peptides, whereas LjCLE5 does not. The functional analysis of MtCLE35 would be of great interest to the nodulation field.

In addition to having rhizobia-induced CLE peptides, soybean has an additional nitrate-induced CLE peptide, GmNIC1a, which acts locally to supress nodulation (Reid et al. 2011b). To date, no orthologue of GmNIC1a has been reported in *L. japonicus* or *M. truncatula*. Here, we used *GmNIC1a* and a BLAST search of the *L. japonicus* and *M. truncatula* genomes to reveal likely orthologous candidates (Supplementary Fig. S3). In soybean and common bean, *NIC1* and *RIC1* are located tandemly within the genome (Reid et al. 2011b; Ferguson et al. 2014). In *L. japonicus*, the putative NIC1 and RIC1 orthologues (*LjCLE40* and *LjCLE-RS2*, respectively) appear in tandem with LjCLE-RS3 and are approximately 24 kb apart on chromosome 3. Interestingly, LjCLE40 was also recently found to be induced by rhizobia inoculation (Nishida et al. 2016). In *M. truncatula*, the predicted orthologue of *NIC1* is *MtCLE34*, which is located tandemly on chromosome 2 with *MtCLE35*. However, a C > T mutation at base 148 of *MtCLE34* results in a premature stop codon and thus the translated product of this gene is likely non-functional. Further investigations are required to determine if the product is indeed truncated.

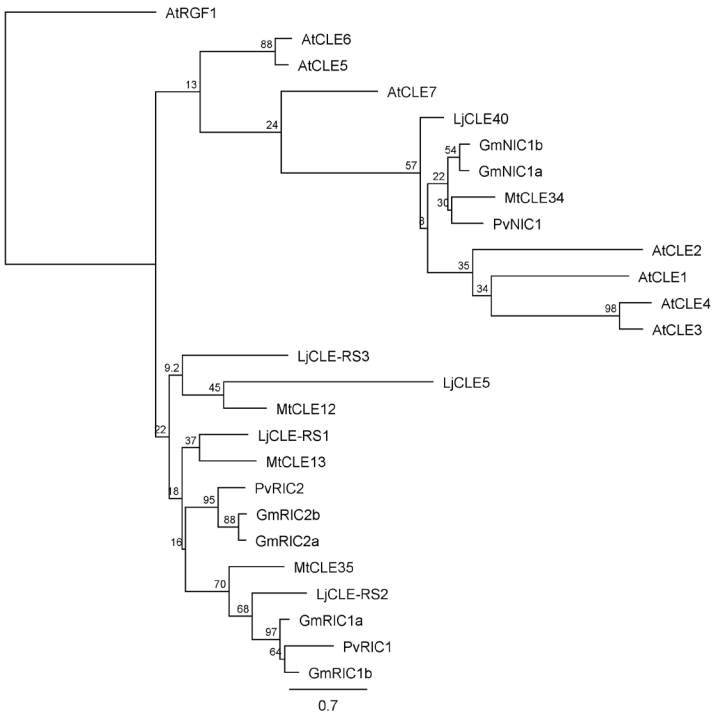


Figure 4.7 Phylogenetic tree of known legume nitrate-induced CLE peptides, rhizobia-induced CLE peptides, including two likely orthologous identified here in addition to *Arabidopsis thaliana*AtCLE1-7, which are most similar to these legume-specific CLE peptides. Bootstrap confidence values displayed are expressed as a percentage of 1,000 bootstrap replications, using AtRGF1 as an outgroup.

The legume nodulation CLE peptides are most similar to AtCLE1-7 of *A. thaliana*, however no direct orthologues have been identified as *A. thaliana* lacks the ability to form a

symbiotic relationship with rhizobia or arbuscular mycorrhizae (Hastwell et al. 2015b). A targeted phylogenetic analysis was utilised here to investigate whether there are specific *A. thaliana* CLE peptides within AtCLE1-7 that are more closely linked with the nodulation CLE peptides of *M. truncatula*, *L. japonicus*, *P. vulgaris* and *G. max* (Fig. 4.7). As expected, the rhizobia-induced CLE peptides form a distinct branch from the nitrate-induced CLE peptides of legumes, and not surprisingly, the *A. thaliana* CLE peptides AtCLE1-7 group closer to these nitrate-induced sequences. This finding further supports the distinction of Group VI made by Hastwell et al. (2015).

4.3.6 Expression of CLE peptide-encoding genes of *M. truncatula* and *L. japonicus*

It would be of little biological relevance to apply the peptides identified here to plants without first understanding their structural modifications and location of synthesis. We therefore used an *in-silico* approach to further assist in the functional characterisation of these genes. Publicly available transcriptome databases of *M. truncatula* and *L. japonicus* were used to collect expression data of the CLE peptide-encoding genes. A meta-analysis was performed to determine if putative orthologues identified by sequence characterisation and phylogenetic analyses exhibited similar expression patterns (Tables 4.2 and 4.3). Some similarity was seen between the putative orthologues, but the number of currently annotated CLE-peptide encoding genes limited a more detailed analysis.

A number of putative orthologues identified in the phylogenetic tree (Supplementary Fig. S3) showed similar expression trends across tissues, such as *PvCLE25* (Hastwell et al. 2015a) and *MtCLE08*, which were both expressed in the root, nodules and stem (Table 4.2). LjCLE15 is expressed highest in the stem with lower expression levels found across all other tissue types and genes that group closely, *MtCLE18* and *PvCLE24*, are expressed in both the stem and root, whereas *AtCLE12*, which also groups closely is only found in the root (Tables 4.2 and 4.3). *MtCLE17* shares a similar expression pattern to *PvCLE23*, *GmCLE23a* and *GmCLE23b* (Hastwell et al. 2015b), being expressed across all tissue types except in seeds, with *MtCLE17* also having notable higher expression in flowers than that of its putative orthologues, which shows little expression in the flower tissue (Table 4.2). *MtCLE12* and *MtCLE13* are currently the only functionally characterised *M. truncatula* CLE peptide-encoding genes, and the transcriptomic data for both genes is consistent with the literature (Mortier et al. 2010), being expressed in the nodules at different stages of development.

Table 4.2 Normalised *Medicago truncatula* CLE peptide-encoding gene expression displayed as log2-transformed values (5.75 = 54.1 fold). The colour scale is independent for each gene.

	Seed 10d	Seed 12d	Seed 16d	Seed 20d	Seed 24d	Seed 36d	Flower	Pod	Vegetative bud	Petiole	Stem	Leaf with petiolules	Nodule 4d	Nodule 10d	Nodule 14d	Nodule mature (4w)	Root	Non-inoculated root
MtCLE18	-2.71	-2.35	-2.47	-1.29	-2.21	-1.64	-1.01	0.17	0.95	1.79	3.34	0.03	0.58	0.1	-0.42	0.29	1.26	2.11
MtCLE22	0.01	-0.15	0.85	1.13	1.82	1.31	-0.96	2.56	0.68	-0.32	0.3	0.15	0.19	0.85	-0.2	0.08	-1.48	-0.94
MtCLE13	-0.62	0.12	-0.3	-0.23	-1.24	0.51	-0.03	-0.35	-1.07	-1.08	-1.06	0	4.78	5.09	4.42	5.16	-0.1	0.07
MtCLE12	-0.34	0.58	-0.53	0.74	0.96	-0.82	-0.22	-0.05	-0.34	-0.77	-2.83	0.7	5.75	5.39	5.24	5.67	-0.79	-0.65
MtCLE04	0.48	0.49	-0.65	0.45	-0.01	-0.72	-0.4	-0.27	-0.11	-0.33	-0.34	-0.35	2.57	2.31	3.22	2.14	-0.78	-0.4
MtCLE01/25	-1.04	-1.5	-0.96	-0.01	0.05	-0.88	1.98	0.11	0.56	0.18	0.74	-0.01	0.42	-0.87	-1.34	-1.18	0.24	0.43
MtCLE17	-2.06	-2.57	-2.85	-0.98	-0.3	-0.94	1.41	-0.15	-0.17	0.17	-0.2	0.41	1.99	1	1.04	0.16	1.49	2.25
MtCLE02	-2.49	-0.44	0.89	0.37	0.17	-1.91	0.61	0.17	1.04	-0.98	-0.5	-1.34	0.07	-0.22	-0.58	0.25	0.87	0.77
MtCLE06	-4.6	-2.06	-0.46	-0.86	-2.51	-3.69	0.72	1.34	0.7	1.45	2.06	-2.04	0.23	0.12	-0.04	-0.45	0.25	0.8
MtCLE08	-2.72	-2.11	-1.42	0.07	-2.38	-1.81	-0.57	-1.92	-1.2	1.68	2.29	-2.37	3.36	2.51	2.98	2.87	2.07	3.14

Min Max

Table 4.3 Normalised *Lotus japonicus* CLE peptide-encoding gene expression displayed as displayed as log2-transformed values (1.96 = 3.9 fold). The colour scale is independent for each gene.

	L.japonicus Gifu salt	L.japonicus MG20 salt	Seed 10d	Seed 12d	Seed 14d	Seed 16d	Seed 20d	Flower	flower 13w 5mM nitrate	Pod 10d	Pod 14d	Pod 20d
LjCLE15	0.06	-0.28	-1.80	-2.15	-1.91	-1.90	-1.76	-1.44	-0.14	-1.66	-1.42	-1.72
LjCLE34	0.20	0.21	-0.04	0.19	-0.02	0.09	0.01	-0.01	0.09	-0.24	-0.11	-0.03
LjCLE38	0.24	-0.49	-0.85	-1.02	-1.78	-1.93	-2.82	-2.93	1.61	-1.84	-2.53	-2.65

	Leaf	Leaf 6w 5mM nitrate	Petiole	Stem	Stem 6w 5mM Nitrate	Shoot 3w uninocul	Shoot 3w inocul3	Root	Root tip 3w uninocul	Root 3w 5mM nitrate	Root 6w 5mM nitrate	Root 3w nodC inocul1
LjCLE15	-1.19	-0.52	0.75	1.96	0.73	-0.26	-0.52	0.17	-0.35	0.10	0.13	0.21
LjCLE34	0.12	0.01	0.10	0.36	0.10	0.12	0.06	0.63	-0.26	0.03	-0.09	-0.10
LjCLE38	0.51	0.54	-1.41	-1.38	-0.80	1.94	1.80	-0.47	-2.21	-1.11	-1.53	-0.73

	Rroot 3w inocul1	Root 3w inocul3	Nodule 3w inocul14	Nodule 3w inocul21	Root 0h (uninocul root)	Nodule 3w inocul	Root + nodule 3w inocul7	Rroot + nodule 3w inocul21	Root ZON 3w uninocul	Rroot ZON 3w inocul1 (Nod_factor)	Root ZON 3w inocul1	Root 4dpi mycorrhized	Root 28dpi mycorrhized
LjCLE15	0.00	0.10	-0.41	-0.56	-0.34	-1.78	-0.16	-0.13	0.29	0.40	-0.09	0.08	-0.18
LjCLE34	-0.10	0.23	-0.18	-0.20	0.77	0.01	-0.08	0.01	-0.31	-0.76	-0.16	0.23	-0.21
LjCLE38	-1.67	-0.70	-1.80	-1.83	0.32	-1.55	-0.94	-0.66	-0.63	-1.20	-0.90	0.12	-0.67

Min – Max

In contrast, some CLE peptide-encoding gene orthologues did not exhibit similar expression patterns within the transcriptomes according to the tissues and treatments available. *PvTDIF1*, *GmTDIF1a*and *GmTDIF1b* show high levels of expression across the different tissues (Hastwell et al. 2015a), with high root expression being of particular importance, as it is the only TDIF peptide-encoding gene to exhibit expression in the root. Their putative orthologues, *AtCLE41* and *AtCLE44* are also expressed in the root, in addition to other tissue types tested (Hastwell et al. 2015a), and *M. truncatula* orthologue, *MtCLE06*, shows no expression in the seeds and is only lowly expressed in the root. *PvCLE29* was noted by Hastwell et al. (2015a) to have very high expression only in the flower. The putative orthologue *LjCLE19*, has previously been shown to respond in the root to phosphate treatment (Funayama-Noguchi et al. 2011) and more recently mycorrhizae colonization (Handa et al. 2015), which is also not consistent with the expression of *PvCLE29* (Hastwell et al. 2015a).

4.4 Discussion

The importance of peptides in plant development is becoming increasingly evident with an extensive number of peptides and peptide families being discovered (Tavormina et al. 2015). CLE peptides are no exception, with confirmed roles in meristematic tissue maintenance, and abiotic and biotic responses; however, the precise function of most is yet to be elucidated. To assist in the discovery of novel CLE peptide functions, the entire CLE peptide family of two model legumes, *M. truncatula* and *L. japonicus*, was identified here. Our analyses increased the number of annotated CLE peptides from 24 to 52 in *M. truncatula* and from 44 to 53 in *L. japonicus*. These were subjected to a range of comparative bioinformatics analyses to create a resource that can be utilised for further reverse-genetics-based functional characterisation. Additionally, six multi CLE domain-encoding genes and a number of pseudogenes were identified across the two species.

The phylogenetic analysis conducted using entire families of CLE prepropeptides of *M. truncatula*, *L. japonicus*, *A. thaliana* and *P. vulgaris* shows strong groupings between those having a similar CLE domain and a known or predicted function. The gene clusters identified here are generally conserved with those identified by Hastwell et al.(2015a), which were divided into seven groups (Group I – VII).

M. truncatula and *L. japonicus* have a similar sized genome (500 Mbp) and share a common ancestor ~37-38 MYA, which is more recent than their shared ancestry with *P. vulgaris* (~45-59 MYA) (Choi et al. 2004). The number of CLE peptide-encoding genes present (52 and 53 respectively), is consistent with the number in the *P. vulgaris* genome, 46, and is roughly half that of *G. max*, which has 84 (Hastwell et al. 2015a) due to a more recent (~13 MYA) whole genome duplication event (Schmutz et al. 2010).

The number of CLE peptide-encoding genes in the legumes is higher than that of *A. thaliana*, which has 32. This is predominately due to the absence of CLE peptide-encoding genes involved in symbioses between rhizobia (Group VI) or mycorrhizae (Delaux et al. 2014; Hastwell et al. 2015a; Handa et al. 2015). The symbioses formed by legumes enable them to acquire nutrients that would otherwise be unavailable (Smith and Smith 2011; Gresshoff et al. 2015). Nodulation control pathways are well characterised in *M. truncatula* and *L. japonicus*, beginning with the production of a CLE peptide (Ferguson et al. 2010; Reid et al. 2011b; Hastwell et al. 2015b). However, a separate nitrate-regulated nodulation pathway identified in *G. max* has not yet been established in these two species. Here, a putative orthologue of *GmNIC1* and *PvNIC1*, which responds to the level of nitrate in the rhizosphere to inhibit nodulation (Reid et al. 2011a; Hastwell et al. 2015b; Ferguson et al. 2014), has been identified in *M. truncatula*. However, *MtCLE34* is likely to be non-functional as a result of a truncation before the CLE domain. The putative orthologue in *L. japonicus*, *LjCLE5*, which has not yet been detected in gene expression studies, is likely to be non-functional as a result of a naturally-occurring insertion/deletion mutation. Further analysis is also needed to determine if MtCLE35 has a functional role in nodulation and if another gene in *L. japonicus* has gained the ability to regulate nodulation in response to nitrogen. Indeed, the latter is hinted towards by the ability of *LjLCE-RS1* to be induced by both rhizobia and nitrate to control nodule numbers (Okamoto et al. 2009; Hastwell et al. 2015b).

Although *A. thaliana* does not enter into a symbiosis with either rhizobia or mycorrhizae, its genome contains orthologues to known symbiosis genes, such as *AtPOLLUX* (Delaux et al. 2013). However, our work indicates that no CLE peptide-encoding genes have yet been identified that show homology or synteny to the rhizobia-induced CLE peptides. It would be of interest to determine if such CLE peptide encoding genes previously existed, or exist but have been overlooked in *A. thaliana* due to being highly divergent from the symbiosis CLE peptides in legumes and other species.

Recent advances in genome sequencing, bioinformatics resources and the identification of entire CLE peptide families of soybean, common bean and Arabidopsis, have been utilised to capture the entire CLE peptide-encoding gene families of two important model legume species, *M. truncatula* and *L. japonicus*. Further characterisation of these CLE peptide-encoding genes revealed orthologues amongst the species, many of which appear functional, with some likely to be pseudogenes. The identification and genetic characterisation of these genes will benefit future studies aimed at functionally characterising these integral molecular components of plant meristem formation and maintenance.

4.5 Methods

4.5.1 Gene identification

Candidate CLE peptide-encoding genes were identified in *L. japonicus* and *M. truncatula* using TBLASTN searches with known all CLE prepropeptides of *G. max* (Hastwell et al. 2015a), *P. vulgaris* (Hastwell et al. 2015a) and *A. thaliana* (Cock and McCormick 2001). The *M. truncatula Mt4.0v1* genome was searched in Phytozome (https://phytozome.jgi.doe.gov/; Young et al. 2011; Goodstein et al. 2012) and the *L. japonicus v3.0* genome was searched in Lotus Base (https://lotus.au.dk/). Initial searches were conducted with E-value = 10. The results were manually validated for the presence of a CLE peptide-encoding gene in an open reading frame. Orthologues were also identified using TBLASTN of newly identified CLE prepropeptide sequences where clear orthologous were not identified between *M. truncatula* and *L. japonicus*, using E-value = 1.

Hidden Markov Models (HMMs) were generated for *M. truncatula* and *L. japonicus* CLEs individually, using all full length prepropeptide sequences as input into HMMER3, respectively (www.hmmer.org). Next, based on the generated HMMs, jackHMMER (www.hmmer.org) was applied to iteratively search for CLE sequences in *M. truncatula* and *L. japonicus* protein databases using a bit score of 50.

4.5.2 Phylogenetic analysis

Multiple sequence alignments were constructed as outlined in Hastwell et al. (2015a). Manual adjustments were made to some predicted sequences, particularly in regards to

their start codon, based on similarity to duplicate genes, clustering genes, and/or likely orthologous genes. Multiple sequence alignments constructed without truncated or likely non-functional CLE prepropeptides were used to generate phylogenetic trees. The trees were constructed using methods described in Hastwell et al. (2015a) using 1,000 bootstrap replications in all cases, except for the tree constructed using the entire families of *L. japonicus*, *M. truncatula*, *A. thaliana* and *P. vulgaris* CLE peptides, which used 100 bootstrap replications. Where orthologues were not apparent, the genomes of *L. japonicus* and *M. truncatula* were re-searched in an attempt to identify a possible orthologue.

4.5.3 Sequence chracterisation

The presence of a signal peptide encoding domain and putative signal peptide cleavage site of the CLE prepropeptides was identified using SignalP (http://www.cbs.dtu.dk/services/SignalP/; Petersen et al. 2011). If no signal peptide was detected, the sequence was manually examined for an up- or downstream methionine, which could be the likely start codon. The modified sequence was re-entered into SignalP and a signal peptide was detected in most instances. Possible intron boundary sites were identified using the NetPlantGene Server (http://www.cbs.dtu.dk/services/NetPGene/; Hebsgaard et al. 1996) and the nucleotide splice sites and resulting prepropeptides were compared with orthologous sequences. Sequence logo graphs of the CLE domain were generated using multiple sequence alignments in Geneious Pro v10.0.2 (Kearse et al. 2012).

Genomic environments were established using five up- and down-stream annotated genes in Phytozome and Lotus Base (https://phytozome.jgi.doe.gov/; https://lotus.au.dk/; Young et al. 2011; Goodstein et al. 2012). Orthologues of individual genes within the genomic environment lacking functional family annotations were identified using BLAST within and between the two databases.

4.5.4 *M. truncatula* and *L. japonicus* transcriptome meta-analysis

The meta-analysis of the normalised transcriptome data was done using publicly available data sets located on the Medicago eFP browser (http://bar.utoronto.ca/efpmedicago/; Benedito et al. 2008; Young et al. 2011; Righetti et al. 2015) and the *Medicago truncatula* Gene Expression Atlas (http://mtgea.noble.org/v3/; Benedito et al. 2008; He et al. 2009) for

M. truncatula, and The *Lotus japonicus* Gene Expression Atlas (http://ljgea.noble.org/v2/; Verdier et al. 2013) for *L. japonicus*.

4.6 Acknowledgements

This work was funded by the Hermon Slade Foundation, and Australian Research Council Discovery Project grants (DP130103084 and DP130102266). The Fellowship Fund Inc. is also thanked for provision of a Molly-Budtz Olsen PhD Fellowship to AHH. TCDB was funded by the European Research Council via a Global Postdoc Fellowship under the Horizon 2020 Marie Sklodowska-Curie Action (Grant no. 659251). We would also like to thank Wolf R. Scheible and Michael K. Udvardi at the Samuel Roberts Noble Foundation, OK, USA for their assistance with the study.

4.7 References

Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Wandrey M, Verdier J, Zuber H, Ott T. 2008. A gene expression atlas of the model legume *Medicago truncatula*. *The Plant Journal* **55**(3): 504-513.

Berckmans B, Simon R. 2016. A feed-forward regulation sets cell fates in roots. *Trends in Plant Science* **21**(5): 373-375.

Chen SK, Kurdyukov S, Kereszt A, Wang XD, Gresshoff PM, Rose R. 2009. The association of homeobox gene expression with stem cell formation and morphogenesis in cultured *Medicago truncatula*. *Planta* **230**(4): 827-840.

Chen S, Lang P, Chronis D, Zhang S, De Jong WS, Mitchum MG, Wang X. 2015. *In planta* processing and glycosylation of a nematode CLAVATA3/ENDOSPERM SURROUNDING REGION-like effector and its interaction with a host CLAVATA2-like receptor to promote parasitism. *Plant Physiology* **167**(1): 262-272.

Choi H-K, Mun J-H, Kim D-J, Zhu H, Baek J-M, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB. 2004. Estimating genome conservation between crop and model legume species. *Proceedings of the National Academy of Sciences of the United States of America* **101**(43): 15289-15294.

Clark SE, Running MP, Meyerowitz EM. 1995. CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. *Development* **121**(7): 2057-2067.

Cock JM, McCormick S. 2001. A large family of genes that share homology with *CLAVATA3*. *Plant Physiology* **126**(3): 939-942.

De Rybel B, Mähönen AP, Helariutta Y, Weijers D. 2016. Plant vascular development: from early specification to differentiation. *Nature Reviews Molecular Cell Biology* **17**(1): 30-40.

Delaux P-M, Séjalon-Delmas N, Bécard G, Ané J-M. 2013. Evolution of the plant–microbe symbiotic 'toolkit'. *Trends in Plant Science* **18**(6): 298-304.

Delaux P-M, Varala K, Edger PP, Coruzzi GM, Pires JC, Ané J-M. 2014. Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *PLoS Genetics* **10**(7): e1004487.

Endo S, Shinohara H, Matsubayashi Y, Fukuda H. 2013. A novel pollen-pistil interaction conferring high-temperature tolerance during reproduction via CLE45 signaling. *Current Biology* **23**(17): 1670-1676.

Etchells JP, Smit ME, Gaudinier A, Williams CJ, Brady SM. 2016. A brief history of the TDIF-PXY signalling module: balancing meristem identity and differentiation during vascular development. *New Phytologist* **209**(2): 474-484.

Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE, Gresshoff PM. 2010. Molecular analysis of legume nodule development and autoregulation. *Journal of Integrative Plant Biology* **52**(1): 61-76.

Ferguson BJ, Li D, Hastwell AH, Reid DE, Li Y, Jackson SA, Gresshoff PM. 2014. The soybean (*Glycine max*) nodulation-suppressive CLE peptide, GmRIC1, functions interspecifically in common white bean (*Phaseolus vulgaris*), but not in a supernodulating line mutated in the receptor PvNARK. *Plant Biotechnology Journal* **12**(8): 1085-1097.

Ferguson BJ, Mathesius U. 2014. Phytohormone regulation of legume-rhizobia interactions. *Journal of Chemical Ecology* **40**(7): 770-790.

Funayama-Noguchi S, Noguchi K, Yoshida C, Kawaguchi M. 2011. Two CLE genes are induced by phosphate in roots of *Lotus japonicus*. *Journal of Plant Research* **124**(1): 155-163.

Gaillochet C, Daum G, Lohmann JU. 2015. O Cell, Where Art Thou? The mechanisms of shoot meristem patterning. *Current Opinion in Plant Biology* **23**: 91-97.

Gaillochet C, Lohmann JU. 2015. The never-ending story: from pluripotency to plant developmental plasticity. *Development* **142**(13): 2237-2249.

Greb T, Lohmann JU. 2016. Plant Stem Cells. *Current Biology* **26**(17): R816-R821.

Gresshoff PM, Hayashi S, Biswas B, Mirzaei S, Indrasumunar A, Reid D, Samuel S, Tollenaere A, van Hameren B, Hastwell AH, Scott P, Ferguson BJ 2015. The value of biodiversity in legume symbiotic nitrogen fixation and nodulation for biofuel and food production. *Journal of Plant Physiology* **172**: 128-136.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**(D1): D1178-D1186.

Grienenberger E, Fletcher JC. 2015. Polypeptide signaling molecules in plant development. *Current Opinion in Plant Biology* **23**: 8-14.

Handa Y, Nishide H, Takeda N, Suzuki Y, Kawaguchi M, Saito K. 2015. RNA-seq transcriptional profiling of an arbuscular mycorrhiza provides insights into regulated and coordinated gene expression in *Lotus japonicus* and *Rhizophagus irregularis*. *Plant and Cell Physiology* **56**(8): 1490-1511.

Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouzé P, Brunak S. 1996. Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. *Nucleic Acids Research* **24**(17): 3439-3452.

Hastwell AH, Gresshoff PM, Ferguson BJ. 2015a. Genome-wide annotation and characterization of CLAVATA/ESR (CLE) peptide hormones of soybean (Glycine max) and common bean (Phaseolus vulgaris), and their orthologues of *Arabidopsis thaliana*. *Journal of Experimental Botany* **66**(17): 5271-5287.

Hastwell AH, Gresshoff PM, Ferguson BJ. 2015b. The structure and activity of nodulation-suppressing CLE peptide hormones of legumes. *Functional Plant Biology* **42**(3): 229-238.

He J, Benedito VA, Wang M, Murray JD, Zhao PX, Tang Y, Udvardi MK. 2009. The *Medicago truncatula* gene expression atlas web server. *BMC Bioinformatics* **10**(1): 441.

Hirakawa Y, Kondo Y, Fukuda H. 2010. TDIF peptide signaling regulates vascular stem cell proliferation via the *WOX4* homeobox gene in Arabidopsis. *The Plant Cell* **22**(8): 2618-2629.

Hobe M, Müller R, Grünewald M, Brand U, Simon R. 2003. Loss of CLE40, a protein functionally equivalent to the stem cell restricting signal CLV3, enhances root waving in Arabidopsis. *Development Genes and Evolution* **213**(8): 371-381.

Ito Y, Nakanomyo I, Motose H, Iwamoto K, Sawa S, Dohmae N, Fukuda H. 2006. Dodeca-CLE peptides as suppressors of plant stem cell differentiation. *Science* **313**(5788): 842-845.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**(12): 1647-1649.

Lim CW, Lee YW, Hwang CH. 2011. Soybean nodule-enhanced CLE peptides in roots act as signals in GmNARK-mediated nodulation suppression. *Plant and Cell Physiology* **52**(9): 1613-1627.

Meng L, Buchanan BB, Feldman LJ, Luan S. 2012. A putative nuclear CLE-like (CLEL) peptide precursor regulates root growth in Arabidopsis. *Molecular Plant* **5**(4): 955-957.

Mitchum MG, Hussey RS, Baum TJ, Wang X, Elling AA, Wubben M, Davis EL. 2013. Nematode effector proteins: an emerging paradigm of parasitism. *New Phytologist* **199**(4): 879-894.

Mortier V, Fenta BA, Martens C, Rombauts S, Holsters M, Kunert K, Goormachtig S. 2011. Search for nodulation-related CLE genes in the genome of *Glycine max*. *Journal of Experimental Botany* **62**(8): 2571-2583.

Nishida H, Handa Y, Tanaka S, Suzaki T, Kawaguchi M. 2016. Expression of the *CLE-RS3* gene suppresses root nodulation in *Lotus japonicus*. *Journal of Plant Research* **129**(5): 909-919.

Ohyama K, Shinohara H, Ogawa-Ohnishi M, Matsubayashi Y. 2009. A glycopeptide regulating stem cell fate in *Arabidopsis thaliana*. *Nature Chemical Biology* **5**(8): 578-580.

Okamoto S, Nakagawa T, Kawaguchi M. 2011. Expression and functional analysis of a CLV3-like gene in the model legume *Lotus japonicus*. *Plant and Cell Physiology* **52**(7): 1211-1221.

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**(1): 67-77.

Okamoto S, Shinohara H, Mori T, Matsubayashi Y, Kawaguchi M. 2013. Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. *Nature Communications* **4**: 2191.

Olsen AN, Skriver K. 2003. Ligand mimicry? Plant-parasitic nematode polypeptide with similarity to CLAVATA3. *Cell* **108**: 489-500.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**(10): 785-786.

Reid DE, Ferguson BJ, Gresshoff PM. 2011a. Inoculation-and nitrate-induced CLE peptides of soybean control NARK-dependent nodule formation. *Molecular Plant-Microbe Interactions* **24**(5): 606-618.

Reid DE, Ferguson BJ, Hayashi S, Lin Y-H, Gresshoff PM. 2011b. Molecular mechanisms controlling legume autoregulation of nodulation. *Annals of Botany* **108**(5): 789-795.

Reid DE, Li D, Ferguson BJ, Gresshoff PM. 2013. Structure–function analysis of the Gm RIC1 signal peptide and CLE domain required for nodulation control in soybean. *Journal of Experimental Botany* **64**(6): 1575-1585.

Replogle A, Wang J, Bleckmann A, Hussey RS, Baum TJ, Sawa S, Davis EL, Wang X, Simon R, Mitchum MG. 2011. Nematode CLE signaling in Arabidopsis requires CLAVATA2 and CORYNE. *The Plant Journal* **65**(3): 430-440.

Righetti K, Vu JL, Pelletier S, Vu BL, Glaab E, Lalanne D, Pasha A, Patel RV, Provart NJ, Verdier J. 2015. Inference of longevity-related genes from a robust coexpression network of seed maturation identifies regulators linking seed storability to biotic defense-related pathways. *The Plant Cell* **27**(10): 2692-2708.

Sawa S, Kinoshita A, Nakanomyo I, Fukuda H. 2006. CLV3/ESR-related (CLE) peptides as intercellular signaling molecules in plants. *The Chemical Record* **6**(6): 303-310.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J. 2010. Genome sequence of the palaeopolyploid soybean. N*ature* **463**(7278): 178-183.

Sharma VK, Ramirez J, Fletcher JC. 2003. The Arabidopsis CLV3-like (CLE) genes are expressed in diverse tissues and encode secreted proteins. *Plant Molecular Biology* **51**(3): 415-425.

Shinohara H, Moriyama Y, Ohyama K, Matsubayashi Y. 2012. Biochemical mapping of a ligand-binding domain within Arabidopsis BAM1 reveals diversified ligand recognition mechanisms of plant LRR-RKs. *The Plant Journal* **70**(5): 845-854.

Smith FA, Smith SE. 2011. What is the significance of the arbuscular mycorrhizal colonisation of many economically important crop plants? *Plant and Soil* **348**(1-2): 63.

Somssich M, Je BI, Simon R, Jackson D. 2016. CLAVATA-WUSCHEL signaling in the shoot meristem. *Development* **143**(18): 3238-3248.

Tavormina P, De Coninck B, Nikonorova N, De Smet I, Cammue BP. 2015. The plant peptidome: an expanding repertoire of structural features and biological functions. *The Plant Cell* **27**(8): 2095-2118.

Verdier J, Torres-Jerez I, Wang M, Andriankaja A, Allen SN, He J, Tang Y, Murray JD, Udvardi MK. 2013. Establishment of the *Lotus japonicus* Gene Expression Atlas (LjGEA) and its use to explore legume seed maturation. *The Plant Journal* **74**(2): 351-362.

Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H. 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**(7378): 520-524.

Chapter 5

Arabinosylation modulates the growth-regulating activity of the peptide hormone CLE40a from soybean

Preface

This journal article has been published in Cell Chemical Biology (2015, http://dx.doi.org/10.1016/j.chembiol.2017.08.014). This article describes the synthesises of an amino acid building block that contains the triarabiniose moiety found in CLE peptides. The building block was used to generate the GmCLE40a peptide and a novel feeding method was established to determine if the arabinosylated variant of the peptide exhibited higher activity than an unmodified variant. The candidate peptide was identified from previous studies in Chapter 3 (Hastwell et al. 2015a). Additional supplementary information and figures may be found online at http://www.cell.com/cell-chemical-biology/fulltext/S2451-9456(17)30312-4.

5.1 Summary

Small post-translationally modified peptide hormones mediate crucial developmental and regulatory processes in plants. CLAVATA/ENDOSPERM-SURROUNDING REGION (CLE) genes are found throughout the plant kingdom and encode for 12–13 amino acid peptides that must often undergo post-translational proline hydroxylation and glycosylation with *O*-β1,2-triarabinose moieties before they become functional. Apart from a few recent examples, a detailed understanding of the structure and function of most CLE hormones is yet to be uncovered. This is mainly owing to difficulties in isolating mature homogeneously modified CLE peptides from natural plant sources. In this study, we describe the efficient synthesis of a synthetic Ara*f*$_3$Hyp glycosylamino acid building block that was used to access a hitherto uninvestigated CLE hormone from soybean called GmCLE40a. Through the development and implementation of a novel *in vivo* root growth assay, we show that the synthetic triarabinosylated glycopeptide suppresses primary root growth in this important crop species.

5.2 Introduction

CLAVATA/Endosperm-Surrounding Region (CLE) genes were first discovered in *Arabidopsis thaliana* in 1999 and have since been identified throughout the plant kingdom (Fletcher et al. 1999, Hastwell et al. 2015a, Oelkers et al. 2008, Okamoto et al. 2009, Strabala et al. 2014, Zhang et al. 2014, Hastwell et al. 2017). CLE genes encode short (12–13 amino acid) peptide hormones with up to three highly conserved proline residues (Hastwell et al. 2015b, Kucukoglu and Nilsson 2015). To date, only a handful of mature functional CLE peptides have been isolated and structurally characterized. For those that have been isolated, prolines 4 and 7 are almost invariably post-translationally hydroxylated to form *trans*-4-hydroxy-L-hydroxyproline (Hyp) residues (Ito et al. 2006, Kondo et al. 2006, Ohyama et al. 2009, Shinohara et al. 2012). In addition, Hyp-7 can be side-chain *O*-glycosylated with the plant specific β1,2-linked tri-L-arabinofuranosyl oligomer to form a central triarabinosylated Hyp (Ara*f*$_3$Hyp) residue (Ohyama et al. 2009). In recent reports, the Ara*f*$_3$Hyp motif has proven to be crucial for the activity of several plant hormones (Okamoto et al. 2013, Shinohara and Matsubayashi 2013, Xu et al. 2015).

Triarabinosylated CLE glycopeptides mediate diverse developmental processes in plants (Kucukoglu and Nilsson 2015). For example, the first triarabinosylated CLE glycopeptide to be identified in Arabidopsis, *At*CLV3, is responsible for the negative regulation of stem cell differentiation in the shoot apical meristem, which gives rise to the above-ground features of the plant (Clark et al. 1993, Clark et al. 1995, Clark et al. 1997, Ogawa et al. 2008, Schoof et al. 2000). The recently discovered orthologs of AtCLV3, SlCLV3, and SlCLE9, fulfill the same function in tomato and have been identified as key factors that control the size and number of fruiting organs (Xu et al. 2015). These latter two peptides represent the only two mature functional CLE hormones to be structurally characterized in a commercially important crop species, highlighting a hitherto unknown molecular basis for the artificial selection of tomato cultivars with better fruiting yields.

The legume (Fabaceae) family contains several important crop and pasture species, including soybean, pea, common bean, mung bean, clover, cowpea, alfalfa, chickpea, lentil, and peanut. Recently, a systemic signal responsible for the negative regulation of root nodule production in *Lotus japonicus* was identified as a triarabinosylated CLE glycopeptide, with a structure analogous to that of CLV3 (Okamoto et al. 2013). While several other CLE peptide-encoding genes have been identified in legume species, the

mature structures and functions of these CLE hormones have yet to be elucidated (Oelkers et al. 2008, Hastwell et al. 2015a, Hastwell et al. 2015b, Hastwell et al. 2017). Hence, probing the structure and function of CLE hormones from the world's most economically significant legume crop species, soybean, is of considerable interest. Recently, the complete family of soybean CLE peptides was identified and genetically characterized (Hastwell et al. 2015a). This includes the ortholog of AtCLE40, a CLV3-related peptide that maintains essential stem cell homeostasis in the root apical meristem (RAM) (Greb and Lohmann 2016, Yamaguchi et al. 2016). Soybean is a paleopolyploid, having undergone genome-wide duplication events roughly 53 and 19 million years ago, followed by the process of diploidization (Schmutz et al. 2010). Consequently, most genes in the soybean genome are duplicated; however, in some instances, one of the gene copies has subsequently undergone genetic variation and/or loss. This is the case with the soybean hormone CLE40, with GmCLE40a maintaining a typical CLE prepropeptide sequence, whereas its homeologous duplicate, GmCLE40b, has a nonsense mutation upstream of the CLE peptide domain that likely renders its ligand un-transcribed and functionless (Hastwell et al. 2015a).

Modifying root architecture via key developmental factors is viewed as a pivotal step in enhancing agricultural sustainability and food security (Meister et al. 2014). CLE40 represents a logical molecular component to evaluate for this purpose due to its central role in root organogenesis and in light of the aforementioned findings with the functionally related CLV3 in enhancing fruit development in tomato. Importantly, CLE40 contains the amino acid motif present in AtCLV3 that is required for arabinosylation (Ohyama et al. 2009), which we proposed would be vital for optimum biological activity (Shinohara and Matsubayashi 2013). However, to date, investigation of the effect of arabinosylation on CLE40 in any species has not been possible due to the difficulty in isolating homogeneously modified hormone in sufficient quantity and purity for biological study. This is owing to the enzymatic nature of the post-ribosomal hydroxylation and glycosylation events, which means that peptide hormones are produced as complex mixtures containing both the triarabinosylated and unglycosylated variants and varying proline hydroxylation patterns (Matsubayashi 2014). Both the inherent challenges associated with purifying such complex mixtures and the low concentration of the hormone in plant tissues mean that chemical synthesis is currently the only viable means to obtain useful quantities of homogeneous triarabinosylated plant glycopeptides for biological study. As such, synthetic methods that enable access to peptides and proteins bearing the Ara*f*$_3$Hyp modification

are of enormous interest to advance the field of plant molecular biology and are a key focus of the work reported here.

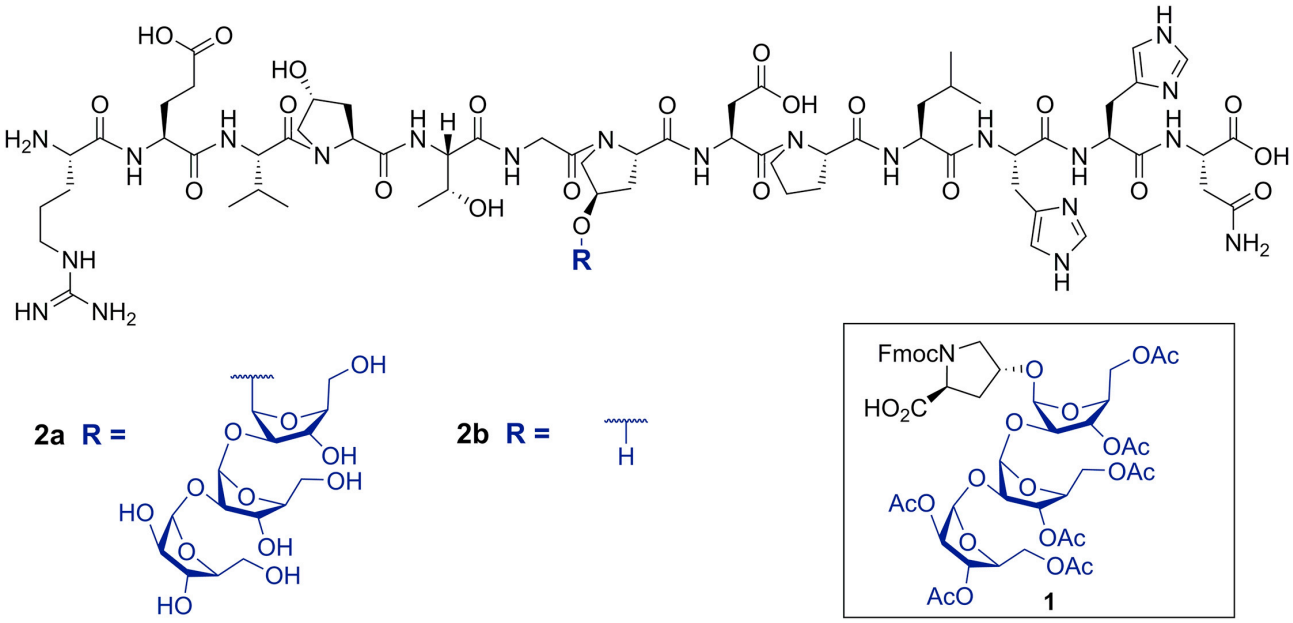


Figure 5.1 Structure of the suitably protected Ara*f*$_3$Hyp building block target **1** (in box) and triarabinosylated and unglycosylated *Gm*CLE40a **2a** and **2b**

The synthesis of Ara*f*$_3$Hyp is complicated by the presence of contiguous β-arabinofuranosidic linkages which, even in a non-contiguous setting, are difficult to construct stereoselectively (Lowary, 2003, Yin and Lowary 2001). Although several chemical methodologies have been developed for the direct construction of β-arabinofuranosidic linkages (Crich et al. 2007, Gadikota et al. 2003, Ishiwata et al. 2006, Lee et al. 2005, Li and Singh 2001, Liu et al. 2013, Zhu et al. 2006), such methodologies do not provide complete stereoselectivity when applied to the construction of plant-derived glycans containing Ara*f*-β-Hyp and contiguous Ara*f*-β1,2-Ara*f* linkages, such as those present in Ara*f*$_3$Hyp (Kaeothip and Boons 2013, Kaeothip et al. 2013, Xie and Taylor 2010). Recently, the Matsubayashi (Shinohara and Matsubayashi 2013) and Ito (Kaeothip et al. 2013) groups reported highly stereoselective routes to an Fmoc-protected and peracetylated Ara*f*$_3$Hyp glycosylamino acid building block **1** (in box, Figure 5.1) and subsequent incorporation into the AtCLV3 peptide through Fmoc-strategy solid-phase peptide synthesis (Fmoc-SPPS). Ito and co-workers have also reported the use of the Ara*f*$_3$Hyp building block for the synthesis of a multiply glycosylated cell-wall extensin fragment, which is also known to natively bear the triarabinose moiety (Ishiwata et al. 2014). Key to both synthetic strategies was the use of an intramolecular aglycone delivery (IAD) strategy, mediated by *p*-methoxybenzyl (PMB) ether (Désiré and Prandi 1999) and

naphthylmethyl (NAP) ether (Ishiwata et al. 2008) auxiliaries, respectively, to construct each β-arabinofuranosidic linkage with complete stereoselectivity.

Here, we report a highly efficient synthesis of Ara*f*$_3$Hyp glycosylamino acid building block **1** using a modified NAP ether-mediated-IAD (NAP-IAD) strategy (Kaeothip et al. 2013) and its use for the chemical synthesis of the soybean triarabinosylated CLE glycopeptide, GmCLE40a **2a** (Figure 5.1). We also report the functional characterization of **2a** through biological evaluation and demonstrate the importance of the central Ara*f*$_3$Hyp moiety through comparison with the synthetic Hyp7 unglycosylated variant **2b**.

5.3 Results and Discussion

5.3.1 Synthesis of triarabinosylated GmCLE40a glycopeptide

In order to access sufficient quantities of the homogeneous triarabinosylated GmCLE40a glycopeptide **2a**, we sought a highly stereoselective route to the Fmoc-protected and peracetylated Ara*f*$_3$Hyp building block **1** using an IAD-based synthetic strategy. One major drawback of the IAD route reported previously was the low yielding formation of the two sterically crowded Ara*f*-β1,2-Ara*f* glycosidic linkages when a conventional donor-mediated tethering step was applied to form the crucial mixed acetal intermediates (Kaeothip et al. 2013, Shinohara and Matsubayashi 2013). Ito and co-workers addressed this issue through recourse to an acceptor-mediated tethering approach, requiring installation of the NAP ether on each acceptor prior to tethering (Kaeothip et al. 2013). While this provided the requisite mixed acetal intermediates in high yield, it added extra linear steps to the synthesis for auxiliary installation and protecting group manipulation. To overcome these issues, we proposed a modification of the NAP-IAD strategy involving less sterically demanding *O*-protecting groups, such as acetyl groups, with a view to improving the yields for donor-mediated tethering steps, thereby enabling access to all three β-arabinofuranosidic linkages in Ara*f*$_3$Hyp using a single 2-*O*-NAP-derived arabinofuranosyl donor.

To this end, we first synthesized thioglycoside donor **3** bearing 3,5-di-*O*-acetyl protection and a 2-*O*-NAP ether auxiliary for IAD (Figure 5.2, see Methods S1 for donor preparation). Donor **3** was tethered to acceptor, Fmoc-Hyp-OBn (1 equiv.), under the promotion of 2,3-dichloro-5,6-dicyano-1,4-benzoquinone (DDQ), affording the bench stable mixed acetal **4**,

which was taken forward without purification. The crude acetal **4** was unreactive when treated under conventional promotion conditions (MeOTf [4 equiv.]/2,4,6-tri-*tert*-butylpyrimidine [TTBP, 5 equiv.]) but smoothly underwent IAD in the presence of MeOTf/Me_2S_2 (4 equiv.)/TTBP (5 equiv.) to give the di-*O*-acetylated-Ara*f*-β-Hyp monoglycoside **5** in 56% overall yield after acidolytic workup (10% trifluoroacetic acid [TFA] in $CHCl_3$). This overall yield was comparable with that obtained in a control experiment with flash chromatographic purification of the intermediate mixed acetal. Importantly, the Ara*f*-β-Hyp linkage was formed with complete stereoselectivity by virtue of the IAD transformation.

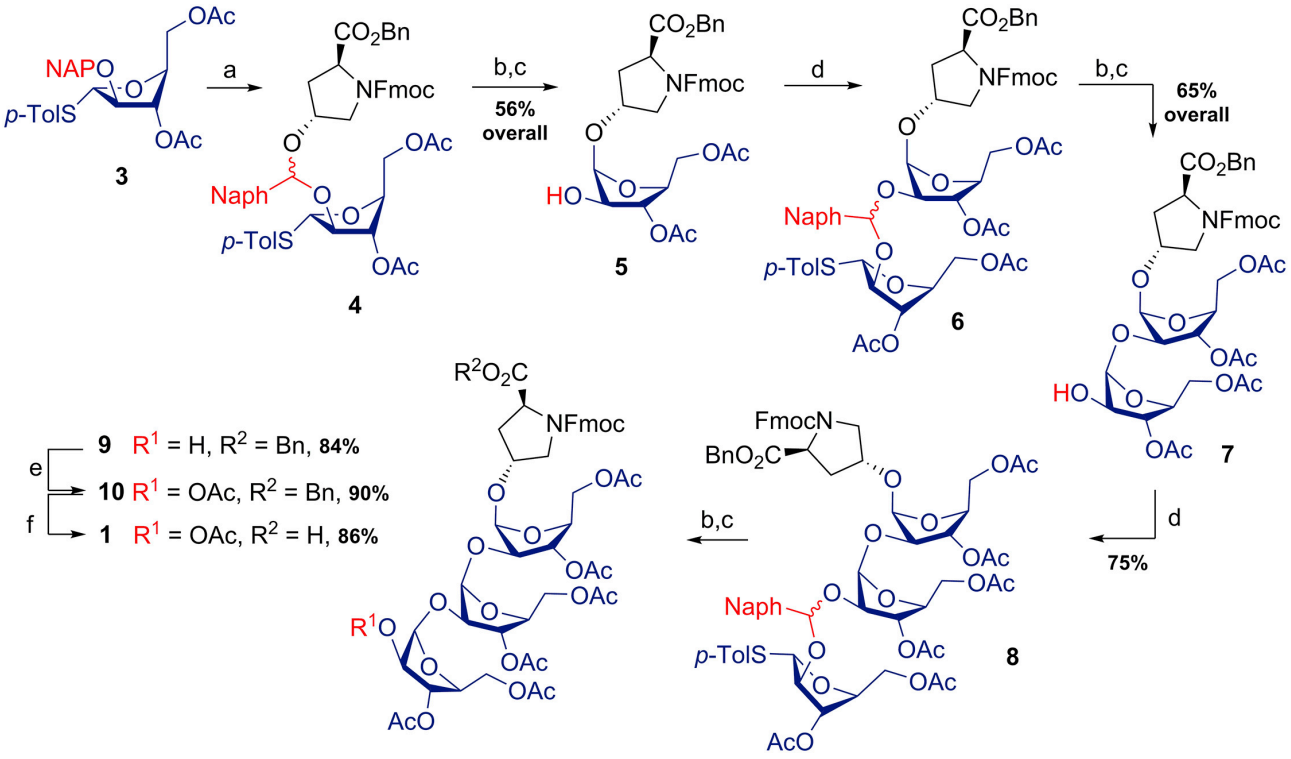


Figure 5.2 Synthesis of glycosylamino acid building block **1**.

Reaction conditions: (a) Fmoc-Hyp-OBn, DDQ, 4 Å molecular sieves, CH_2Cl_2, room temperature (rt), 18 hr; (b) Me_2S_2, MeOTf, TTBP, 4 Å molecular sieves, CH_2Cl_2, rt, 12–18 hr; (c) TFA:$CHCl_3$ 1:9 v/v, rt, 30 min; (d) **3**, DDQ, CH_2Cl_2, rt, 18 hr; (e) Ac_2O, pyridine, rt, 16 hr; (f) Et_3SiH, 10% Pd/C, MeOH, 2 hr.

Next, we attempted the tethering of NAP ether **3** to acceptor **5** to provide the requisite mixed acetal for formation of the first challenging Ara*f*-β1,2-Ara*f* linkage. Pleasingly, mixed acetal **6** was afforded in an excellent isolated yield of 85% using only 1.05 equiv. of NAP ether **3** when the reaction was tested on a small scale. As discussed above, this particular tethering step has been reported to be low yielding in both previously reported syntheses of the Ara*f*$_3$Hyp building block **1** (Kaeothip et al. 2013, Shinohara and Matsubayashi 2013).

We attribute the improvement in tethering yield to the sterically unencumbering acetyl-protected donor and acceptor combination. In a scale-up of the tethering procedure, the mixed acetal **6** was used directly in the next IAD step without purification to afford diglycoside acceptor **7** in 65% overall yield with exclusively β-stereoselectivity. An excess of donor **3** (2 equiv.) was employed to push the next tethering reaction to completion to afford mixed acetal **8** in 75% yield after purification by flash column chromatography. Subsequent rearrangement gave triglycoside **9**, containing all three β-arabinofuranosidic linkages, in 85% yield. To complete the synthesis of **1**, treatment of **9** with Ac_2O in pyridine provided the peracetylated glycoside **10**, which was subjected to a mild transfer hydrogenation procedure using Et_3SiH and Pd/C catalyst to remove the benzyl ester without affecting the Fmoc protecting group (Mandal and McMurray 2007). This provided the desired glycosylamino acid building block **1** bearing the Ara*f*$_3$Hyp moiety in 86% yield. Overall the Ara*f*$_3$Hyp building block **1** was assembled in 18% yield over 11 steps, which significantly improves on the best synthesis previously reported (Kaeothip et al. 2013) using the longer acceptor-mediated NAP-IAD strategy (12% over 15 steps).

With the suitably protected Ara*f*$_3$Hyp building block **1** in hand, the synthesis of the target triarabinosylated CLE glycopeptide **2a** could now commence (Figure 5.3). Toward this end, 2-chlorotrityl chloride resin was first loaded with Fmoc-His(Trt)-OH, followed by conventional Fmoc-SPPS conditions to afford resin-bound hexapeptide **11**. The resin-bound hexapeptide **11** was subsequently Fmoc-deprotected and treated with a coupling mixture containing 1.2 equiv. of the Ara*f*$_3$Hyp glycosylamino acid building block **1**, 1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU), 1-hydroxy-7-azabenzotriazole (HOAt), and *i*Pr_2NEt. Under these conditions, the glycosylamino acid was incorporated quantitatively (as judged by liquid chromatography-mass spectrometry analysis of a test cleavage) to provide resin-bound glycoheptapeptide **12**. Subsequent extension using conventional Fmoc-SPPS gave the resin-bound glycotridecapeptide, which was Fmoc-deprotected and acidolytically released from the resin with concomitant global deprotection of the side-chain protecting groups from the amino acids. The carbohydrate moiety was next deacetylated in solution using NaOMe in MeOH before purification by reversed-phase high-performance liquid chromatography using 0.1% TFA in the eluent. After lyophilization, the triarabinosylated GmCLE40a glycopeptide **2a** was isolated in an excellent overall yield of 22% over the 26 linear steps (~94% per step).

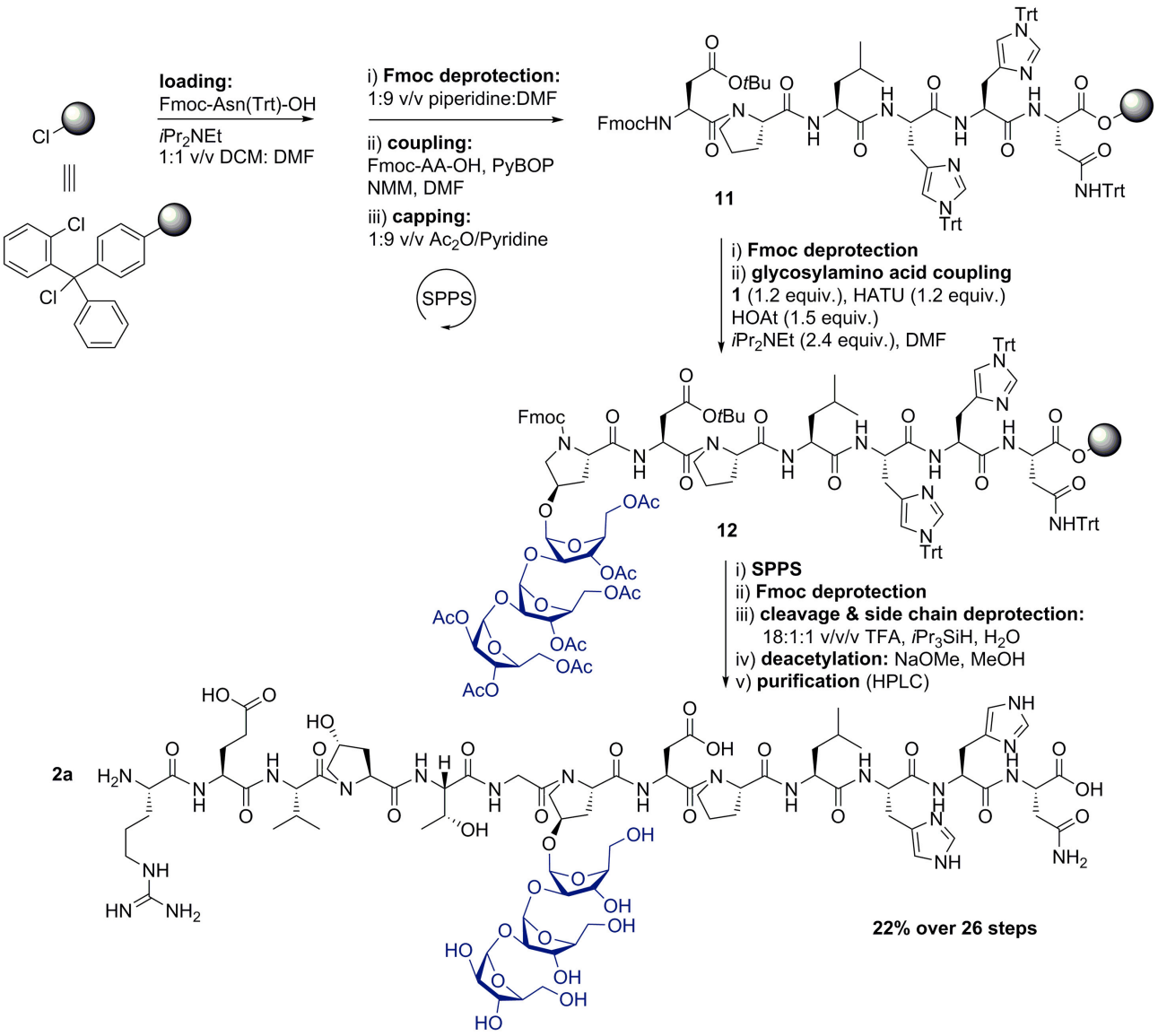


Figure 5.3 Synthesis of triarabinosylated *Gm*CLE40a peptide **2a**.

5.3.2 Functional characterization of the GmCLE40a glycopeptide hormone

Only six putative CLE40 orthologs have been reported to date (in Arabidopsis, soybean, common bean, rice, *Medicago truncatula* and *Lotus japonicus*) (Hastwell et al. 2015a, Hastwell et al. 2017, Hobe et al. 2003, Kinoshita et al. 2007, Sharma et al. 2003). To better understand the signal provided by the *GmCLE40a*glycopeptide and establish amino acid conservation within the peptide ligand, BLAST searches were conducted across a range of plant species. A total of 26 putative CLE40 orthologs (including GmCLE40b) were identified across 21 different species (Figure 5.4A). Each contain two introns, consistent with the CLV3 and CLE40 encoding genes of Arabidopsis, soybean, and common bean. Interestingly, those from Arabidopsis and the Brassicaceae family form a distinct branch with a high bootstrap value (94.2), and those identified in the monocot species group

within the same clade as CLV3, but on a distinct and well supported (96.9) branch (Figure 5.4C). This includes OsFCP1 (also known as OsCLE402). The CLE domain of the orthologs is highly conserved, with only four residues showing less than 90% pairwise identity (positions 2, 7, 9, and 13; Figures 5.4A and 5.4B). The amino acid residue at position 7 is a proline in 65% (including GmCLE40a) with a serine in 31% (including AtCLE40, and orthologs from other species within the Brassicaceae family). Importantly, proline and serine are both residues that can be subjected to *O*-glycosylation (Van den Steen et al. 1998), and all of the CLE40 orthologs identified contain the motif for arabinosylation.

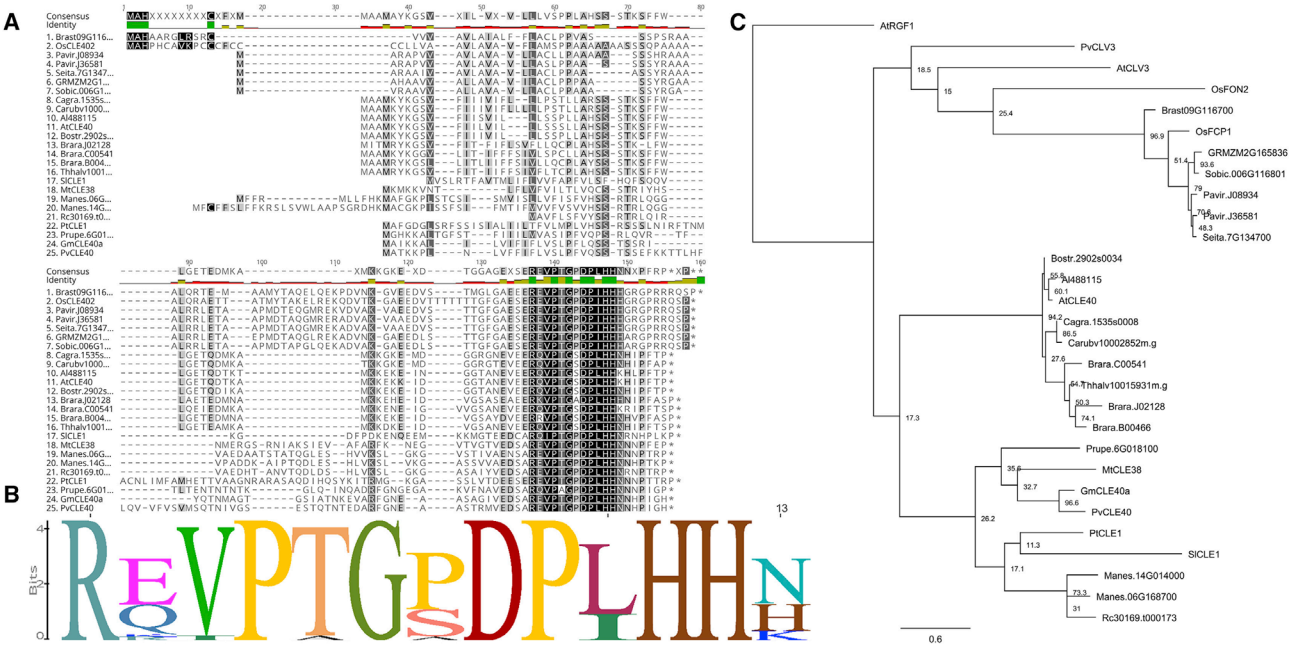


Figure 5.4 CLE40 Orthologs in various species. **(A)** Multiple sequence alignment of CLE40 prepropeptides showing highest conservation between orthologs in the CLE domain at positions 136–148. Outside of the CLE domain, sequence conservation is typically seen between closely related species. **(B)** Sequence logo diagram representing amino acid conservation in the CLE domain of the CLE40 orthologs. **(C)** Phylogenetic tree of the CLE40 orthologs, along with some *At*CLV3 orthologs which group separately, and *At*RGF1 as an outgroup. CLE40 orthologs from monocots group within the same clade as CLV3 orthologs, but on a distinct branch with OsFCP1, which is known to function in the root apical meristem of rice. GmCLE40b has been excluded as it is truncated before the CLE domain.

5.3.3 Development of a bioassay to assess GmCLE40a activity in soybean

To evaluate GmCLE40a activity, a novel bioassay was developed to quantify the effect of the peptide on soybean root growth. Initially, *Agrobacterium*-mediated soybean hairy root

transformation was carried out to establish where *GmCLE40a* is transcriptionally active, and hence where to apply synthetic peptide hormones in the bioassay. These studies involved driving GUS reporter gene expression with the 2.5 kB promotor region located directly upstream of *GmCLE40a*. GUS expression was observed in the apical region of the root tip (Figure 5.5), which is in agreement with the peptide's role in regulating the stem cell population of the RAM, and consistent with the expression pattern of *AtCLE40* (Stahl et al. 2009). Based on this, different concentrations of the GmCLE40a glycopeptide **2a** or its unglycosylated variant **2b**were applied every 12 hr to the tap root tip of wild-type soybean seedlings, and the length of the root was subsequently recorded. Specialized growth pouches were modified and used to enable precision feeding of the peptides and to record the development of the tap root in a non-destructive and repetitious manner (Figure S1). It is important to note that while most feeding studies broadly apply peptides to the entire root/plant, precision feeding is a highly localized technique, minimizing unwanted and biologically irrelevant responses.

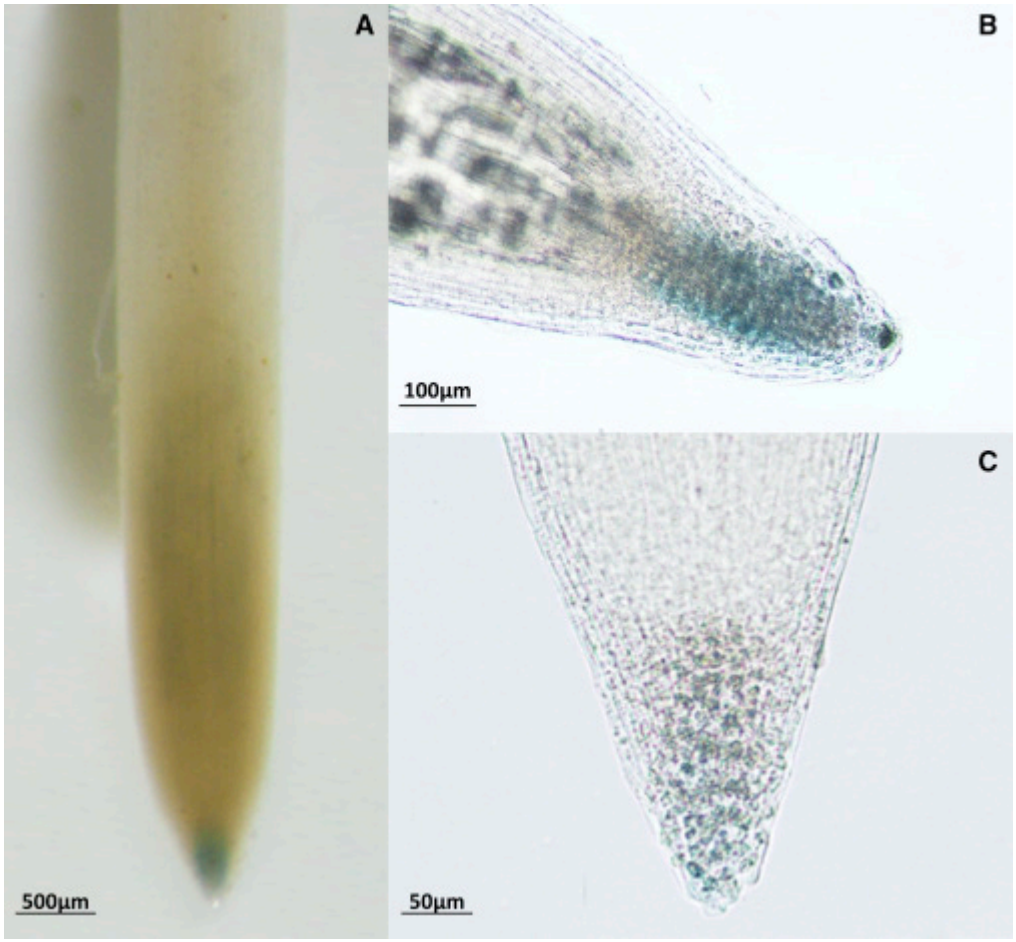


Figure 5.6 Expression pattern of pro*GmCLE40a::GUS* in 2-week-old soybean hairy roots. **(A)** and **(B)** show activation of the *GmCLE40a* promoter in the apical region of the tap root. **(C)** exemplifies its activity in lateral roots.

5.3.4 Triarabinosylated GmCLE40a glycopeptide possesses potent root growth inhibition in soybean

To determine the biological activity of GmCLE40a variants, soybean seedlings treated with different concentrations of the triarabinosylated GmCLE40a glycopeptide **2a**, the unglycosylated GmCLE40a variant **2b**, or water (control) were used in the root growth bioassay. Tap root lengths were measured throughout the experiment to establish the effect of the treatments on growth over time (Figure 5.6).

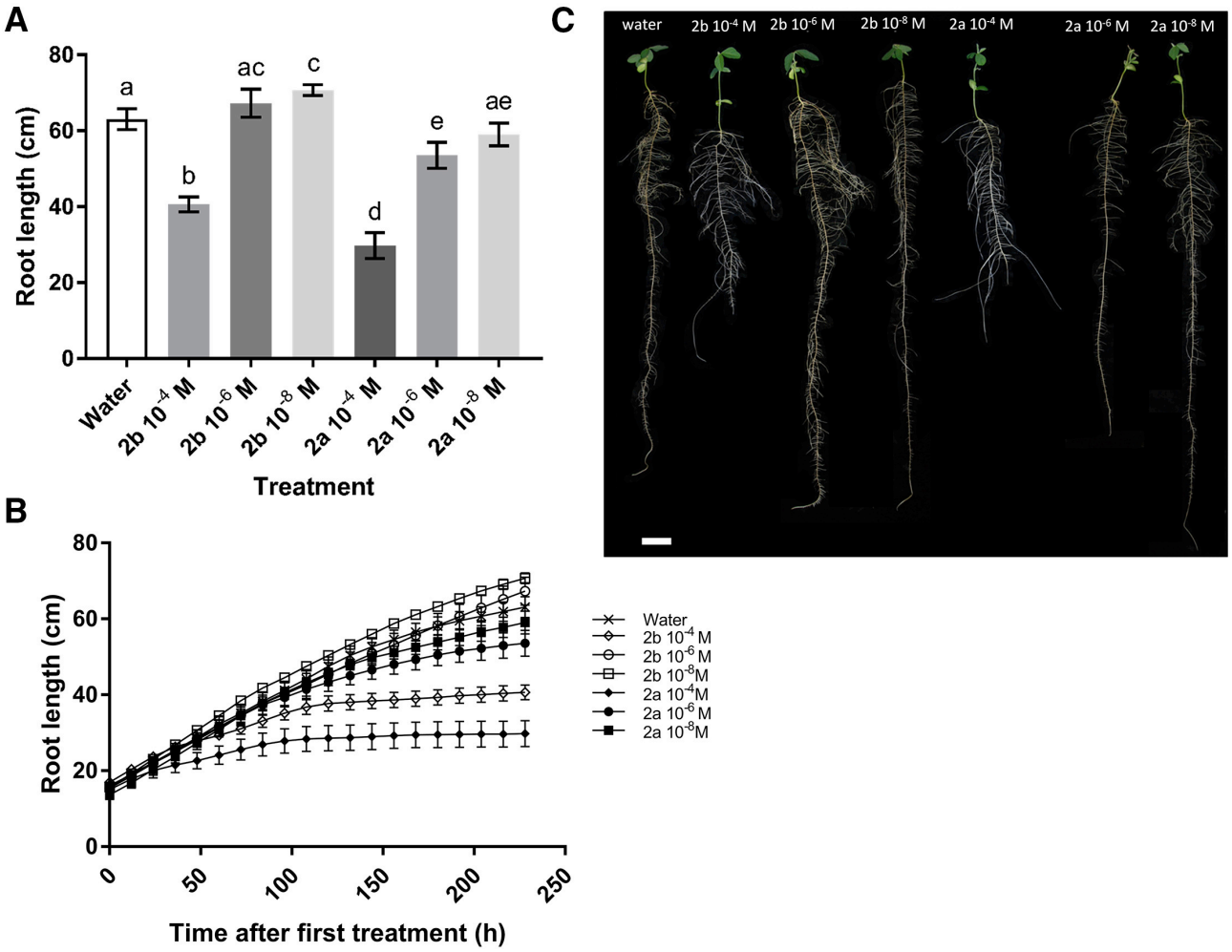


Figure 5.6 Soybean root growth following treatment with 10^{-8} to 10^{-4} M *Gm*CLE40a glycopeptide **2a** and peptide **2b**. The tip of the tap root was treated directly every 12 hr for a total of 228 hr. **(A)** Total tap root length after 228 hr of treatment. Different letters above the bars represent significant statistical differences (Student's t test, $p \leq 0.05$). **(B)** Total length of the tap root recorded throughout the experiment. Some error bars are not presented as they appear smaller than the icons displayed. **(C)** Fifteen day-old soybean plants following treatment for 228 hr. n = 9–15 plants per treatment. Error bars represent the SEM. Scale bar, 5.5 cm.

Compared with the control treatment, application of the triarabinosylated GmCLE40a glycopeptide **2a** significantly reduced root growth at concentrations of 10^{-4} ($p < 0.0001$) and 10^{-6} M ($p \leq 0.05$; Figure 5.6). In contrast, GmCLE40a peptide **2b**, containing an unmodified Hyp residue, only inhibited root growth when applied at 10^{-4} M ($p \leq 0.05$). A significant difference in root length was also observed between the glycosylated **2a** and unglycosylated peptide **2b** treatments, with the former suppressing root growth significantly more than the latter at both 10^{-4} and 10^{-6} M ($p \leq 0.05$; Figure 5.6). Moreover, for each concentration tested the growth rate of the root was significantly reduced by arabinosylated GmCLE40a **2a** compared with **2b** (Figure 5.7). These results demonstrate that arabinosylated GmCLE40a is significantly more potent than the hydroxylated version of the peptide at reducing root growth. This indicates an important role for the carbohydrate moiety in CLE40 activity, possibly acting directly in perception and/or protection of the peptide ligand from peptidase breakdown.

Suppression of soybean root growth by GmCLE40a application is consistent with *A. thaliana* studies using semi- or unmodified AtCLE40 (Fiers et al. 2005). Excess levels of AtCLE40 caused by overexpression can also significantly reduce root growth, and intriguingly, so can reduced levels caused by genetic mutation (Hobe et al. 2003). This suggests that homeostasis is required for optimum root growth, where either elevated or reduced levels of the peptide can prevent maximum growth. Interestingly, application of hydroxylated GmCLE40a at 10^{-8} M led to a mild yet significant enhancement of soybean root growth compared with the arabinosylated glycopeptide and water control treatments ($p \leq 0.05$; Figure 5.2A). To the best of our knowledge, this is the first report of such an increase and it is tempting to speculate that certain CLE40 analogs have the potential to enhance root growth, which would have tremendous commercial and agricultural potential. It should be noted that this reversed effect of promoting or inhibiting plant development when exogenously applying different concentrations of a plant hormone has previously been reported for gibberellin, brassinosteroids, auxin, and cytokinin (e.g., Ferguson and Mathesius, 2014, Hayashi et al. 2014, Wei and Li, 2016).

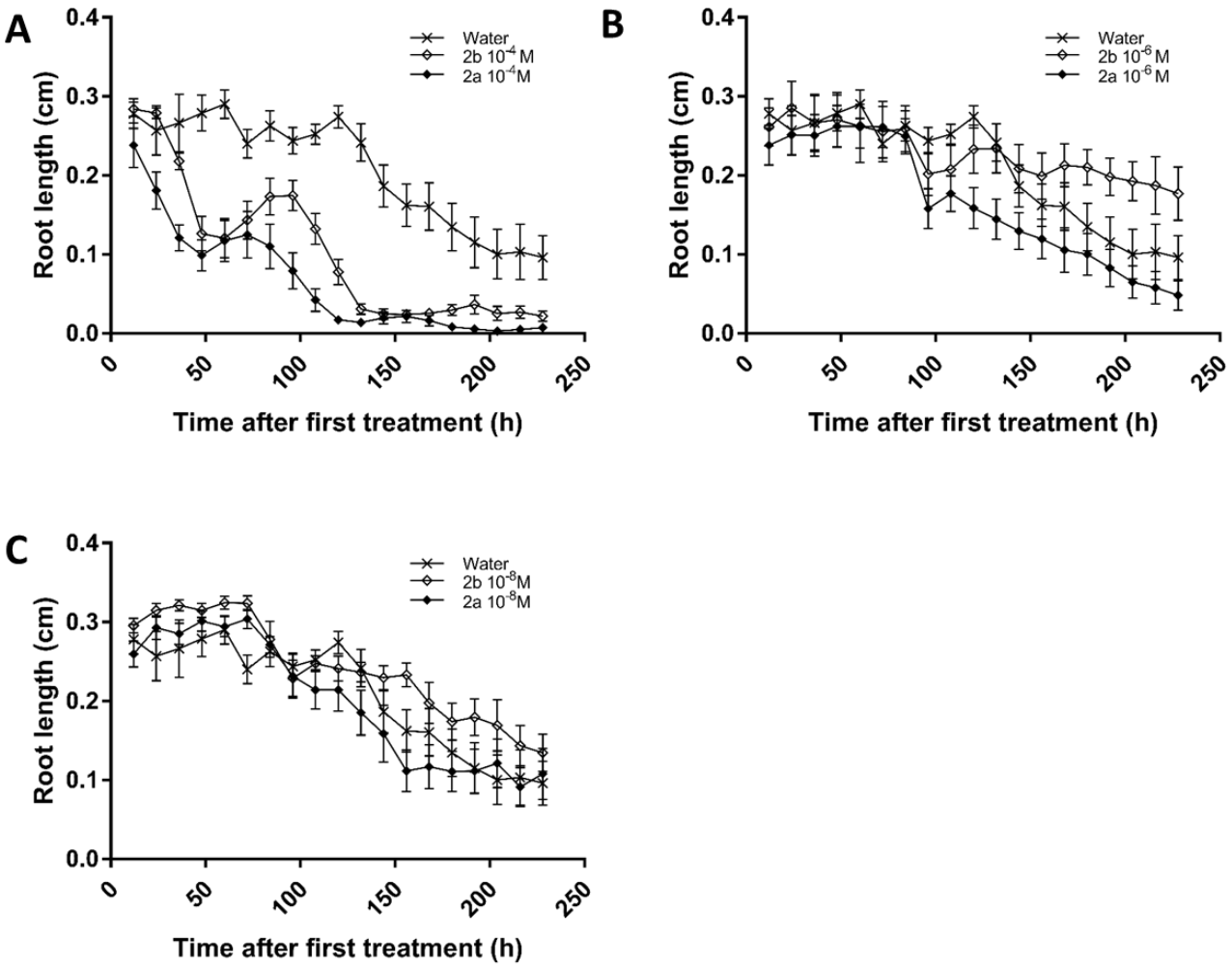


Figure 5.7 Rate of tap root growth of soybean plants treated every 12 hr with *Gm*CLE40a variants, including glycopeptide **2a,** peptide **2b**, and water control. (A) 10^{-4} M, (B) 10^{-6} M, and (C) 10^{-8} M. Some error bars are not presented as they appear smaller than the icons displayed. n = 9–15 plants per treatment. Error bars represent the SEM.

5.3.5 NMR conformational analysis of CLE40a (glyco)peptides

Using nuclear magnetic resonance (NMR) and computational techniques, Shinohara and Matsubayashi (2013)) have previously shown that the central triarabinose of the CLE40a ortholog, AtCLV3, causes the C terminus to bend away from the glycan through a kink at the conserved Gly-6 residue. Based on these data, the authors proposed that the triarabinose moiety on CLV3 may be crucial for maintaining the correct conformation of the peptide ligand for receptor binding and the downstream biological activity. In order to determine whether the same conformational effect was operative in glycosylated CLE40a, we conducted comparative 2D nuclear Overhauser spectroscopy (NOESY) analysis of CLE40a (glyco)peptides **2a** and **2b**. Using the homonuclear total correlation spectroscopy,

double quantum filtered correlation spectroscopy and NOESY spectra, we made full ^{1}H resonance assignments for both the peptide and sugar portions of the two molecules at 278 K and 298 K. A number of nuclear Overhauser effects (NOEs) were observed, particularly at the lower temperature, that were consistent with the peptides displaying significant conformational preferences. For example, HN-HN(*i*,*i* + 1) NOEs were observed for the residue pairs T5-G6, L10-H11, and H11-H12. As exemplified by the plot of Hα chemical shifts for both **2a** and **2b** (Figure S2), there were no significant chemical shift changes between the glycosylated and non-glycosylated forms of CLE40a, other than changes in the side chain of HyP7 that are expected from addition of the trisaccharide unit. This, together with the observation that very similar NOE patterns were observed for the two peptides (Figure S3), strongly suggests that glycosylation of CLE40a does not have a significant effect on the conformational preferences of the peptide.

Given that triarabinosylation of CLE40a does not provide the same conformational changes to the underlying peptide backbone as observed for AtCLV3, it can be deduced that alteration to the conformation or shape of the peptide is not responsible for the increased inhibition of root growth observed for **2a**, compared to the unmodified Hyp-containing peptide **2b**. As such, it is possible that the improved activity is owing to improved interaction with the putative receptor through H-bonding interactions with the carbohydrate moiety. Alternatively, the triarabinose unit may provide improved proteolytic stability of the peptide hormone that would enhance the half-life, and therefore activity, of the hormone. Studies to address these possibilities will be the subject of future work in our laboratories.

In summary, we have developed an efficient synthetic route to a suitably protected glycosylamino acid building block bearing Ara*f*$_3$Hyp, a post-translational modification that has recently emerged as a common feature of plant peptide hormones and proteins. The building block was used to access a homogeneous arabinosylated GmCLE40a glycopeptide, which functions to control the stem cell population of the RAM of plants. Moreover, a novel and highly effective bioassay was developed to evaluate the peptide's activity in relation to root growth. Findings from this work demonstrate that the Ara*f*$_3$Hyp residue significantly enhances GmCLE40a root growth suppressive activity. This raises a pertinent and fundamental question relating to the use of plant peptide variants in application and binding studies, which are often performed using only semi- or unmodified variants to ascertain peptide perception and function. Whether the Ara*f*$_3$Hyp residue of CLE40 is optimal for receptor binding, enhanced ligand stability, localization, or some

other aspect that promotes the peptide's activity is now of great interest to determine. Moreover, the Ara*f*$_3$Hyp residue prepared here can now be used to synthesize additional plant proteins and peptide signals of interest, with a focus on those that could potentially benefit crop development and yields. Studies toward this end will be the subject of future work in our laboratories.

5.4 Significance

The ubiquity and diverse functionality of CLE hormones make them important research targets in the study of plant development. Most CLE hormones isolated to date are triarabinosylated and possess little or no biological activity without this critical post-translational modification. However, the isolation of mature functional CLE hormones is not always possible, meaning that structural and functional characterization is dependent on access to the homogeneous triarabinosylated isoform through chemical synthesis. This paper outlines a more straightforward and higher yielding route to the Ara*f*$_3$Hyp glycosylamino acid and demonstrates its utility through chemical synthesis and functional characterization of soybean CLE40a using a novel root growth bioassay. The tools presented here should assist with the functional characterization of new CLE hormones and other triarabinosylated plant glycopeptides.

5.5 Star methods

5.5.1 Experimental model and subject details

Wild type soybean, *Glycine max* [L.] Merr. cv. Bragg, was used in this study. For experiments using pouches, chlorine gas sterilized seeds were germinated for 2 days in Grade 3 sterilized vermiculite and autoclaved Milli-Q® water. Germinated seedlings having a straight radicle of a similar length (2-3 cm) were transplanted to modified CYG germination pouches (Mega International, Newport, MN, USA) (Hayashi et al. 2012). The pouch length was increased as required to prevent roots from reaching the bottom. Pouches were watered with autoclaved Milli-Q® water, making sure to avoid excess water build-up or drying out of the filter paper.

Seedlings grown for genetic transformation were first ethanol sterilized (Ferguson et al. 2014) and germinated in Grade 2 vermiculite for 4 days prior to *A. rhizogenes*stab-

inoculation (Ferguson et al. 2014, Kereszt et al. 2007, Lin et al. 2010). Three days after inoculation, additional vermiculite was added to cover the wound site and this was covered with cellophane wrap to enhance humidity and promote transgenic hairy-root growth. Plants were watered every three days, alternating between water and B&D nutrient solution containing 1mM KNO_3 (Broughton and Dilworth 1971). Two weeks after inoculation, hairy roots were harvested for histochemical beta-glucuronidase (GUS) staining.

All plants were grown under 16:8 day:night conditions. For peptide feeding, plants were grown at 28°C:25°C respectively in a E-75L1 growth chamber (Percival Scientific, Perry, IA, USA); and for hairy-root transformation, 25°C:22°C respectively in a TPG-1260-TH growth chamber (Thermoline, Wetherill Park, NSW, Australia).

E. coli XL1-Blue was cultured at 37°C overnight LB with 50 µg/ml kanamycin and *Agrobacterium rhizogenes* K599 cultured for genetic transformation of soybean was grown at 28°C on Solid LB medium with 50 µg/ml rifampicin and 100 µg/ml ampicillin as described in Reid et al. (2011)).

5.5.2 General synthetic and analytical procedures

Commercial materials, including solvents were used as received unless otherwise noted. Anhydrous MeOH, DMF and CH_2Cl_2 were obtained from a PURE SOLV^{TM}solvent dispensing unit. Solution-phase reactions were carried out under an atmosphere of dry nitrogen or argon.

Flash column chromatography was performed using 230–400 mesh Kieselgel 60 silica eluting with gradients as specified. Analytical thin layer chromatography (TLC) was performed on commercially prepared silica plates (Merck Kieselgel 60 0.25 mm F254). Compounds were visualized using UV at 254 nm and 5% H_2SO_4 in ethanol charring solution.

^{1}H NMR, ^{13}C NMR, DEPT-135 and 2D NMR spectra were recorded at 300 K using a Bruker DRX500, DRX400 or AVANCE300 spectrometer. Chemical shifts are reported in parts per million (ppm) and are referenced to solvent residual signals: $CDCl_3$ δ 7.26 [^{1}H], and δ 77.16 [^{13}C]; and D_2O δ 4.79 [^{1}H]. ^{1}H NMR data is reported as chemical shift, multiplicity, relative integral, coupling constant, and assignment where possible. Signal

assignments and regiochemical information were obtained through standard 2D experiments (HSQC, HMBC and phase-sensitive COSY). Glycosylamino acid ^{1}H NMR signal assignments marked with the superscripts ‘ and “ indicate signals corresponding to the central arabinoside (Ara*f*-β1,2-**Ara*f*-**β1,2-Ara*f*-β-Hyp) and terminal arabinoside (**Ara*f*-**β1,2-Ara*f*-β1,2-Ara*f*-β-Hyp), respectively. Unmarked ^{1}H NMR sugar signal assignments refer to the reducing terminal arabinoside (Ara*f*-β1,2-Ara*f*-β1,2-**Ara*f*-β-Hyp).

High resolution ESI+ mass spectra were measured on a Bruker–Daltonics Apex Ultra 7.0T Fourier transform mass spectrometer (FTICR). Infrared (IR) absorption spectra were recorded on a Bruker ALPHA Spectrometer with Attenuated Total Reflection (ATR) capability. Compounds were deposited as films on the ATR plate *via* a CH_2Cl_2 solution. Optical rotations were recorded at ambient temperature (293K) on a Perkin–Elmer 341 polarimeter at 589 nm (sodium D line) with a cell path length of 1 dm, and the concentrations are reported in g/100 mL.

UPLC chromatograms and low resolution ESI mass spectra were obtained on a Shimadzu NexeraX2 UPLC equipped with a SPD-M30A diode array detector and a LCMS-2020 ESI mass spectrometer operating in positive ion mode.

Preparative reverse-phase HPLC was performed using a Waters 600 Multisolvent Delivery System and pump with Waters 486 Tuneable absorbance detector operating at 214 nm. Analytical reverse-phase HPLC was performed on a Waters 2695 separations module equipped with a 2996 DAD detector operating at 214 nm.

5.5.3 Synthesis of thioglycoside donor

Thioglycoside donor 3 was prepared in 8 steps from l-arabinose. Please see Methods S1 for detailed synthetic methods and characterization data for 3 and synthetic intermediates.

5.5.4 General procedure for β-arabinofuranosylation via IAD

5.5.4.1 Mixed acetal formation

A mixture of acceptor Fmoc-Hyp-OBn, **5** or **7** (1.0 equiv.), donor **3** (1-2 equiv.) and activated powdered 4Å molecular sieves (1 g.mmol^{-1} of acceptor) in anhydrous CH_2Cl_2(20 mL.mmol^{-1} of acceptor, 50 mM) was stirred for 2 h at rt before addition of DDQ (2.0-2.5

equiv.) in a single portion. The resulting dark green-blue reaction mixture was stirred at rt under argon for 18 h and then cautiously (CO_2 evolution) treated with *ca.* 5 volume equivalents of sat. aq. $NaHCO_3$ solution. The biphasic mixture was vigorously stirred until complete hydrolysis of DDQ, as indicated by almost complete decolourization of the organic layer (*ca.* 20 min). The mixture was filtered through celite, and the celite pad was washed with additional CH_2Cl_2. The organic layer was separated, and the red aqueous layer was extracted with additional equivalents of CH_2Cl_2. The combined organic extracts were dried over $MgSO_4$, filtered and concentrated under reduced pressure, affording crude mixed acetal as a mixture of diastereoisomers (see Figure S4 for exemplar UPLC data), which was either purified by silica gel column chromatography or used directly in the next step without further purification.

5.5.4.2 IAD

A mixture of mixed acetal **4**, **6** or **8** (1.0 equiv.), TTBP (4.0 equiv), Me_2S_2 (4.0 equiv.) and activated powdered 4Å molecular sieves (2.5 g.mmol^{-1} of acetal) in anhydrous CH_2Cl_2 (100 mL.mmol^{-1} of acetal, 10 mM) was stirred at rt for 2 h before addition of MeOTf (4.0 equiv.). The mixture was stirred at rt for 12-18 h and then treated with sat. aq. $NaHCO_3$ solution (*ca.* 0.5 volume equiv.) and vigorously stirred for an additional 30 min. After filtering the mixture through celite, the organic layer was removed and concentrated. The residue, containing mostly naphthylidene adducts (see Figures S4–S6 for exemplar UPLC data), was dissolved in TFA:$CHCl_3$ (1:9 v/v), and stirred at rt for 30 min before co-evaporation with toluene. The residue was purified by silica gel chromatography (eluent as specified), affording β-arabinofuranoside as a white foam.

IAD reactions could be monitored by UPLC-MS using the following procedure. Analytical samples were diluted with MeCN, filtered, and eluted with a linear gradient of 50 to 100% MeCN [0.1% formic acid] in H_2O [0.1% formic acid] over 8 min (Acquity UPLC® BEH C18 1.7 µm, 2.1 x 50 mm, 0.6 mL/min). Compounds were visualized with UV absorbance at 265 nm.

5.5.5 Synthesis of Ara*f*$_3$Hyp building block

5.5.5.1 Fmoc-Hyp(Ac$_2$Araf)-OBn

Monoarabinoside **5** was synthesized following the general method for β-arabinosylation using 700 mg (1.58 mmol) of acceptor Fmoc-Hyp-OBn, 1.0 equiv. of donor **3**, and 2.0 equiv. of DDQ. The intermediate mixed acetal **4** was used crude for IAD (see Figure S4 for UPLC traces collected during reaction monitoring). Acidolytic workup and purification by silica gel chromatography (eluent: EtOAc:toluene 2:3 →1:1 v/v), afforded **5** as a white foam (588 mg, 56% over 3 steps). **[α]$_D$** +1.4° (c 0.77, $CHCl_3$). **^{1}H NMR** (500 MHz, $CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 7.78-7.73 (m, 2H, ArH), 7.59-7.51 (m, 2H, ArH), 7.42-7.36 (m, 2H, ArH), 7.34-7.24 (m, 7H, ArH), 5.22, 5.16, 5.13 (3d, 1.5H, J = 12.3 Hz, 1.5PhCH), 5.08-5.01 (m, 2.5H, H1, H3, 0.5PhCH), 4.55 (t_{apt}, 0.5H, $J_{\alpha,\beta}$ = 7.5 Hz, 0.5Hα), 4.52-4.39 (m, 2.5H, 0.5Hα, Hγ, FmocCH_{2a}), 4.36-4.18 (m, 4.5H, H2, H5$_a$ H5$_b$, FmocCH_{2b}, 0.5FmocCH), 4.11-4.08 (m, 1H, H4), 4.00 (t_{apt}, 0.5H, J = 6.8 Hz, 0.5FmocCH), 3.74-3.69 (m, 1.5H, 0.5Hδ$_a$, Hδ$_b$), 3.57-3.55 (m, 0.5H, 0.5Hδ$_a$), 2.60-2.49 (m, 2H, Hβ$_a$, OH), 2.26-2.17 (m, 1H, Hβ$_b$), 2.12, 2.11 (2s, 3H, CH_3CO), 2.04 (s, 3H, CH_3CO) ppm. **^{13}C NMR** (125 MHz, $CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 172.15, 172.12, 170.9, 170.8, 170.7, 154.8, 154.6, 144.3, 144.2, 144.0, 143.7, 141.51, 141.46, 141.4, 135.6, 135.4, 129.2, 128.7, 128.6, 128.5, 128.4, 128.3, 127.9, 127.8, 127.2, 125.3, 125.2, 125.1, 120.12, 120.08, 100.7, 79.4, 79.3, 78.7, 78.5, 76.3, 76.2, 76.1, 75.4, 67.9, 67.7, 67.3, 67.2, 65.3, 65.2, 58.2, 57.9, 52.0, 51.6, 47.31, 47.27, 37.7, 36.6, 21.0, 20.9 ppm. **FTIR**: 3470, 2951, 1740, 1704, 1451, 1420, 1351, 1230, 1189, 1166, 1126, 1077, 1032, 992, 758, 738, 699 cm^{-1}. **HRMS** (ESI+): calcd. for $C_{36}H_{37}NO_{11}Na$ 682.2259, found 682.2260 (M+Na).

5.5.5.2 Fmoc-Hyp(Ac$_4$Araf$_2$)-OBn

Diarabinoside **7** was synthesized following the general method for β-arabinosylation using 588 mg (0.891 mmol) of acceptor **5**, 1.05 equiv. of donor **3**, and 2.0 equiv. of DDQ. The intermediate mixed acetal **6** was used crude for IAD (see Figure S5 for the UPLC trace collected during reaction monitoring). Acidolytic workup and purification by silica gel chromatography (eluent: EtOAc:toluene 1:1 → 3:2 v/v), afforded **7** as a white foam (507 mg, 65% over 3 steps). **[α]$_D$** +34.6° (c 2.15, $CHCl_3$). **^{1}H NMR**(500 MHz, $CDCl_3$) *ca.* 2:1 mixture of rotational isomers: δ 7.78-7.73 (m, 2H, ArH), 7.59-7.56 (m, 2H, ArH), 7.42-7.21 (m, 9H, ArH), 5.23-5.19 (m, 1.33H, H3, 0.33PhCH), 5.15-5.08 (m, 2.33H, 0.67H1, 1.67PhCH), 5.03 (d, 0.33H, $J_{1,2}$ = 4.3 Hz, 0.33H1), 4.99 (t_{apt}, 0.33H, $J_{2',3'/3',4'}$ = 6.0 Hz, 0.33H3'), 4.93-4.90 (m, 1.33H, 0.67H1', 0.67H3'), 4.85 (d, 0.33H, $J_{1',2'}$ = 4.4 Hz, H1'), 4.61,

4.57 (2t_{apt}, 1H, Jα,β = 4.6, Hα), 4.51-4.46 (m, 1.33H, Hγ, 0.33FmocCH_2), 4.45-4.41 (m, 1H, H2), 4.37-4.23 (m, 5H, H5a, H5b, H5a', 1.67 FmocCH_2, 0.33FmocCH), 4.21-4.13 (m, 2H, H2', H5'), 4.11-4.01 (m, 2.67H, H4, H4', 0.67FmocCH), 3.84-3.82 (m, 0.67H, 0.67Hδ_a), 3.67 (dd, 0.33H, $J_{\gamma,\delta}$ = 4.5 Hz, $J_{\delta a,\delta b}$ = 11.5 Hz, 0.33Hδ_b), 3.60-3.56 (m, 1H, 0.33Hδ_a, 0.67Hδ_b), 2.94-2.88 (m, 1H, 2'-OH), 2.55, 2.48 (2m, 1H, Hβ_a), 2.24-2.15 (m, 1H, Hβ_b), 2.12, 2.11, 2.069, 2.065, 2.06, 2.053, 2.049, 1.93 (8s, 12H, 4 x CH_3CO) ppm. **^{13}C NMR** (125 MHz, $CDCl_3$) *ca.* 2:1 mixture of rotational isomers: δ 172.4, 172.2, 170.7, 170.64, 170.57, 170.5, 170.4, 155.0, 154.9, 144.2, 144.11, 144.08, 144.0, 141.5, 141.4, 141.31, 141.28, 135.6, 153.4, 128.6, 128.5, 128.4, 128.2, 127.8, 127.74, 127.71, 127.21, 127.17, 127.15, 125.34, 125.27, 125.2, 125.08, 120.1, 119.9, 101.5, 101.4, 98.7, 98.3, 80.1, 80.0, 79.62, 79.57, 79.55, 79.4, 78.9, 78.7, 76.34, 76.25, 75.4, 74.2, 68.1, 67.8, 67.2, 67.1, 66.0, 65.5, 65.2, 58.0, 57.9, 51.5, 50.8, 47.2, 37.7, 36.5, 20.91, 20.87, 20.8, 20.7 ppm. **FTIR**: 3492, 2956, 2918, 2850, 1742, 1706, 1452, 1422, 1368, 1233, 1194, 1164, 1125, 1034, 905, 760, 741, 699 cm^{-1}. **HRMS** (ESI+): calcd. for $C_{45}H_{49}NO_{17}Na$ 898.2893, found 898.2889 (M+Na).

5.5.5.3 Fmoc-Hyp(Ac$_6$Araf$_3$)-OBn

Triarabinoside **9** was synthesized following the general method for β-arabinosylation using 674 mg (0.770 mmol) of acceptor **7**, 2.0 equiv. of donor **3**, and 2.5 equiv. of DDQ. The intermediate mixed acetal **8** (786 mg, 75%) was obtained after purification by silica gel chromatography (EtOAc:hexane 1:2 →1:1 v/v). Subsequent IAD, acidolytic workup, and purification by silica gel chromatography (eluent: EtOAc:toluene 1:1 → 2:1 v/v), afforded **9** as a white foam (533 mg, 84% [63% over 3 steps]). See Figure S6 for the UPLC trace collected during IAD reaction monitoring. **[α]$_D$** +56.1° (c 1.18, $CHCl_3$). **^{1}H NMR** (500 MHz, $CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 7.78-7.73 (m, 2H, ArH), 7.60-7.53 (m, 2H, ArH), 7.42-7.22 (m, 9H, ArH), 5.22-5.06 (m, 5.5H, H1, H3, 0.5H1', H3', PhCH_2), 5.02-4.97 (m, 2H, 0.5H1', 0.5H", H3"), 4.94 (d, 0.5H, $J_{1",2"}$ = 4.5 Hz, 0.5H1"), 4.59, 4.55 (2dd, 1H, $J_{\alpha,\beta}$ = 6.7, 8.3 Hz, Hα), 4.52 -4.16 (m, 11.5H, Hγ, H2, H2', H2", 2H5, 2H5', H5$_a$", FmocCH_2, 0.5FmocCH), 4.12-4.06 (m, 3H, H4, H4', H5$_b$"), 4.03-3.99 (m, 1H, 0.5H4", 0.5FmocCH), 3.87 (m, 0.5H, 0.5H4"), 3.73-3.65 (m, 1.5H, 0.5Hδ_a, Hδ_b), 3.54 (m, 0.5H, 0.5Hδ), 2.59, 2.51 (2m, 1H, Hβ_a), 2.25, 2.17 (2m, 1H, Hβ_b), 2.12, 2.10, 2.09, 2.08, 2.055, 2.049, 2.01, 1.982, 1.976, 1.94 (10s, 18H, 6 x CH_3CO) ppm. **^{13}C NMR** (125 MHz, $CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 172.3, 172.2, 170.81, 170.76, 170.73, 170.65, 170.6, 170.48, 170.46, 170.38, 170.37, 154.8, 154.7, 144.3, 144.2, 144.0, 143.7, 141.5, 141.41, 141.40, 141.3,

135.6, 135.3, 128.7, 128.6, 128.5, 128.4, 128.2, 127.89, 127.87, 127.85, 127.7, 127.24, 127.21, 125.4, 125.18, 125.15, 125.08, 120.1, 120.0, 120.1, 101.0, 100.9, 99.6, 99.1, 98.5, 98.3, 80.1, 79.82, 79.78, 79.5, 79.3, 78.4, 78.0, 76.6, 76.3, 76.2, 75.3, 74.9, 68.1, 67.8, 67.23, 67.21, 66.3, 66.1, 65.8, 65.7, 65.6, 65.2, 58.1, 57.8, 51.9, 51.1, 47.3, 37.5, 36.5, 20.94, 20.91, 20.89, 20.88, 20.86, 20.79, 20.75, 20.74, 20.71 ppm. **FTIR**: 3514, 2955, 2922, 1739, 1707, 1451, 1421, 1367, 1227, 1164, 1124, 1033, 905, 759, 738, 700 cm^{-1}. **HRMS** (ESI+): calcd. for $C_{54}H_{61}NO_{23}Na$ 1114.3527, found 1114.3524 (M+Na).

5.5.5.4 Fmoc-Hyp(Ac$_7$Araf$_3$)-OBn

To a solution of free alcohol **9** (513 mg, 0.470 mmol) in pyridine (4 mL) was added Ac_2O (2 mL). The solution was stirred at rt for 16 h and then co-evaporated with toluene. Purification by silica gel chromatography (eluent: EtOAc:hex 1:1 → 3:2 v/v) afforded peracetylated trisaccharide **10** as a white foam (478 mg, 90%). **$[\alpha]_D$** +63.8° (c 1.00, $CHCl_3$). **^{1}H NMR** (500 MHz, $CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 7.77-7.72 (m, 2H, ArH), 7.61-7.58 (m, 1.5H, ArH), 7.53 (m, 0.5H, ArH), 7.41-7.20 (m, 9H, ArH), 5.36, 5.28 (2d, 1H, $J_{1",2"}$ = 4.3 Hz, H1"), 5.26-5.09 (m, 5H, H1/H1', H3/H3', H3", $PhCH_2$), 5.05-4.99 (m, 2.5H, H1/H1', H3/H3', 0.5H2"), 4.93 (dd, 0.5H, $J_{2",3"}$ = 7.3 Hz, 0.5H2"), 4.71 (m, 1H, Hα), 4.53-4.22 (m, 9H, H2, H2', $H5_a$, H5a' H5b/H5b', $H5_a$", Hγ, $FmocCH_{2a}$, 0.5FmocCH, 0.5$FmocCH_{2b}$), 4.16-3.94 (m, 5.5H, H4, H4', 0.5H4", $H5_b/H5_b$', $H5_b$", 0.5$FmocCH_{2b}$, 0.5FmocCH), 3.78-3.69 (m, 2H, 0.5H4", $H\delta_a$, 0.5$H\delta_b$), 3.58 (m, 0.5H, 0.5$H\delta_b$), 2.80, 2.63 (2m, 1H, $H\beta_a$), 2.31-2.20 (m, 1H, $H\beta_b$), 2.110, 2.107, 2.09, 2.081, 2.076, 2.063, 2.061, 2.04, 1.95, 1.92, 1.91, 1.82 (12s, 21H, 7 x CH_3CO) ppm. **^{13}C NMR** (125 MHz, $CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 172.6, 172.4, 170.81, 170.77, 170.74, 170.65, 170.6, 170.4, 170.3, 169.9, 154.8, 154.6, 144.6, 144.3, 144.0, 143.7, 141.44, 141.39, 141.3, 135.7, 135.4, 128.7, 128.54, 128.50, 128.4, 128.2, 127.9, 127.7, 127.24, 127.17, 125.6, 125.3, 125.20, 125.16, 120.1, 120.04, 120.00, 98.9, 98.6, 97.82, 97.76, 97.7, 97.6, 80.6, 80.5, 79.8, 79.6, 79.2, 79.0, 77.7, 77.30, 77.25, 77.2, 77.1, 76.8, 76.6, 76.5, 76.4, 75.6, 75.4, 75.2, 74.6, 68.0, 67.8, 67.2, 67.1, 66.4, 66.3, 66.2, 65.5, 65.4, 58.2, 57.9, 51.9, 51.4, 47.3, 37.6, 36.4, 20.93, 20.91, 20.78, 20.77, 20.74, 20.70, 20.6, 20.52, 20.48 ppm. **FTIR**: 2955, 2918, 1741, 1708, 1451, 1421, 1369, 1225, 1164, 1120, 1036, 909, 760, 740, 700 cm^{-1}. **HRMS** (ESI+): calcd. for $C_{56}H_{63}NO_{24}Na$ 1156.3632, found 1156.3633 (M+Na).

5.5.5.5 5 Fmoc-Hyp(Ac_7Araf_3)-OH

Benzyl ester deprotection was accomplished using a chemoselective transfer hydrogenation method reported by Mandal and McMurray (2007)). Specifically, Et_3SiH (508 µL, 3.18 mmol) was added dropwise to a stirred suspension of benzyl ester **10** (361 mg, 318 mmol) and 10% Pd/C (36 mg) in MeOH (5 mL) in a vessel equipped with an argon balloon. After effervescence had ceased, the reaction was allowed to stir for an additional 1 h at rt, before filtering through celite. The filtrate was evaporated to dryness and purified by silica gel chromatography (eluent: MeOH:AcOH:CH_2Cl_2 0:1:99 → 9:1:90 v/v/v, slow gradient) affording glycosylamino acid building block **1** as a white foam (285 mg, 86%). $[\alpha]_D^{20}$ +65.0° (c 1.00, $CHCl_3$), lit: $[\alpha]_D^{26}$ +75.2°. **^{1}H NMR** (500 MHz, $CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 7.75 (d, 1H, *J* = 7.7 Hz, ArH), 7.64 (t_{apt}, 1H, *J* = 8.0 Hz, ArH), 7.60-7.57 (m, 1.5H, ArH), 7.52 (d, 0.5H, *J* = 7.5 Hz, ArH), 7.39 (t_{apt}, 1H, *J* = 7.4 Hz, ArH), 7.34-7.23 (m, 3H, ArH), 5.33 (d, 0.5H, $J_{1",2"}$ = 4.3 Hz, 0.5H1"), 5.26-5.24 (m, 1H, 0.5H1", 0.5H3"), 5.19-5.16 (m, 1.5H, 0.5H1/H1', 0.5H3/H3', 0.5H3"), 5.14-5.11 (m, 1H, 0.5H1/H1', 0.5H3/H3'), 5.04 (dd, 0.5H, *J* = 4.5, 5.9 Hz, 0.5H3/H3'), 5.02-4.99 (m, 2H, H1/H1', 0.5H3/H3' 0.5H2"), 4.93 (dd, 0.5H, $J_{2",3"}$ = 7.1 Hz, 0.5H2"), 4.69, 4.64 (2t_{apt}, 1H, $J_{\alpha,\beta}$ = 7.5 Hz, Hα), 4.53-4.05 (m, 14H, H2, H2', H4, H4', 0.5H4", 2H5, 2H5', H5$_a$", 0.5H5$_b$", Hγ, FmocCH, FmocCH2), 3.97 (dd, 0.5H, $J_{4",5b"}$ = 8.9 Hz, $J_{5a",5b"}$ = 11.4 Hz, 0.5H5$_b$"), 3.75 (m, 0.5H, 0.5Hδ$_a$), 3.71 (m, 0.5H, 0.5H4"), 3.67-3.62 (m, 1H, Hδ$_b$), 3.57 (m, 0.5H, 0.5Hδ$_a$), 2.80, 2.61 (2m, 1H, Hβ$_a$), 2.37-2.27 (m, 1H, Hβ$_b$), 2.102, 2.096, 2.09, 2.07, 2.06, 2.04, 1.96, 1.94, 1.90, 1.80 (10s, 21H, 7 x CH_3CO) ppm. **^{13}C NMR** (125 MHz, $CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 177.0, 175.4, 171.0, 170.9, 170.82, 170.76, 170.7, 170.6, 170.5, 170.4, 170.3, 170.0, 169.9, 155.6, 154.7, 144.4, 144.0, 143.9, 143.7, 141.42, 141.37, 141.32, 141.26, 127.9, 127.8, 127.6, 127.3, 127.21, 127.15, 125.4, 125.2, 125.10, 125.06, 120.1, 120.02, 119.98, 98.7, 98.6, 97.81, 97.75, 97.7, 97.5, 80.5, 80.4, 79.7, 79.6, 79.2, 79.0, 77.7, 77.29, 77.27, 77.2, 77.1, 76.8, 76.6, 76.4, 75.6, 75.4, 75.0, 74.5, 68.1, 68.0, 66.34, 66.30, 66.2, 65.6, 65.4, 58.0, 57.5, 51.8, 51.4, 47.3, 47.2, 37.6, 36.1, 20.93, 20.91, 20.75, 20.65, 20.51, 20.48 ppm. **FTIR**: 2955, 2928, 1423, 1368, 1220, 1164, 1120, 1032, 994, 908, 761, 736, 702, 603, 545, 427 cm^{-1}. **HRMS** (ESI+): calcd. for $C_{49}H_{57}NO_{24}Na$ 1066.3163, found 1066.3164 (M+Na). The data is in agreement with that reported by Kaeothip et al. (2013)).

5.5.6 Synthesis of CLE40a glycopeptide

5.5.6.1 Preloading of 2-Chlorotrityl chloride resin

2-Chlorotrityl chloride resin (1.6 mmol/g resin substitution, 2 equiv.) was swollen in dry CH_2Cl_2 for 30 min then washed with CH_2Cl_2 (5 × 2 mL) and DMF (5 × 2 mL). A solution of Fmoc-Asn(Trt)-OH (1 equiv.) and *i*Pr_2NEt (2 equiv.) in 1:1 v/v DMF:CH_2Cl_2 (10 mL.mmol^{-1} of amino acid) was added and the resin and shaken at rt for 16 h. After filtering, the resin was washed with DMF (5 × 2 mL) and CH_2Cl_2 (5 × 2 mL), and then treated with a capping solution of CH_2Cl_2/MeOH/*i*Pr_2NEt (17:2:1 v/v/v, 10 mL.mmol^{-1} of amino acid) for 3 h. The resin was again washed with DMF (5 × 2 mL), CH_2Cl_2 (5 × 2 mL), and DMF (5 × 2 mL) before submitting to iterative peptide assembly (Fmoc-SPPS).

5.5.6.2 General Fmoc deprotection

The resin was shaken with piperidine:DMF (1:9 v/v, 2 mL, 2 × 3 min) then filtered off and washed with DMF (5 × 2 mL), CH_2Cl_2 (5 × 2 mL) and DMF (5 × 2 mL).

5.5.6.3 General amino acid coupling

A solution of Fmoc-AA(PG)-OH (4 equiv.), PyBOP (4 equiv.) and *N*-methylmorpholine (8 equiv.) in DMF (10 mL.mmol of peptide) was added to the resin-bound peptide (1 equiv). The resin was shaken for 45 min, then filtered off and washed with DMF (5 × 3 mL), CH_2Cl_2 (5 × 3 mL) and DMF (5 × 3 mL). Any amino acid directly following a hydroxyproline or glycosylhydroxyproline was double-coupled using 10 equiv. of Fmoc-AA(PG)-OH, 10 equiv PyBOP and 20 equiv. *N*-methylmorpholine for each coupling.

5.5.6.4 Glycosylamino acid coupling

A solution of glycosylamino acid **1** (1.2 equiv.), HATU (1.2 equiv.), HOAt (1.5 equiv.) and *i*Pr_2NEt (2.4 equiv.) in DMF (10 mL.mmol^{-1} of peptide) was added to the resin-bound peptide (1.0 equiv.) and shaken for 12 h. The resin was filtered off and washed with DMF (5 × 2 mL), CH_2Cl_2 (5 × 3 mL), and DMF (5 × 2 mL).

5.5.6.5 Capping

Following coupling of an Fmoc-AA(PG)-OH or glycosylamino acid **1**, the resin was treated with acetic anhydride/pyridine (1:9 v/v, 2 mL) and shaken for 3 min. The resin was filtered off and washed with DMF (5 × 2 mL), CH_2Cl_2 (5 × 2 mL) and DMF (5 × 2 mL).

5.5.6.6 Cleavage

After washing thoroughly with CH_2Cl_2 (7 × 2 mL), the resin was suspended in a mixture of TFA, triisopropylsilane and water (90:5:5 v/v/v, 40 mL.mmol^{-1} of peptide). The suspension was shaken for 2.5 h and then filtered. The resin was washed with additional TFA, and the combined filtrates were concentrated under a stream of nitrogen gas. The residue was dispersed in toluene with the aid of sonication and then evaporated to dryness on a rotary evaporator. The residue was dried for 16 h under high vacuum before deacetylation.

5.5.6.7 Deacetylation and purification

The cleaved glycopeptide residue was dissolved in anhydrous MeOH (160 mL.mmol of peptide) under argon, and adjusted to ~pH 10 (wet universal indicator paper) with 0.5 M NaOMe in MeOH (~4-8 mL.mmol^{-1} of peptide). The solution was stirred until deacetylation was complete (~30 min as judged by LC-MS) and then neutralized with a drop of formic acid. The reaction mixture was concentrated under a stream of nitrogen, and the resulting solid residue re-suspended in H_2O, filtered and purified by preparative reverse phase HPLC (Waters X-bridge BEH300 C18 5 µm, 19 × 150 mm, 7 mL min^{-1}, 0 to 25% MeCN [0.1%TFA] in H_2O [0.1% TFA] over 45 min, rt ~ 33 min).

Glycopeptide **2a** was prepared on a 12.5 µmol scale according to the general procedures outlined above. After preparative HPLC and lyophilization, glycopeptide **2a** was obtained as a fluffy white solid as the *tetrakis*(trifluoroacetate) salt (6.39 mg, 22%). **Analytical HPLC**: R_t 20.2 min (0 to 30% MeCN [0.1% TFA] in H_2O [0.1% TFA] over 30 min, sample dissolved in H_2O, λ = 214 nm). **LRMS**: m/z 1898 $[M+H]^+$, 949 $[M+2H]^{2+}$, 633 $[M+3H]^{3+}$, 589 $[M+3H\text{-}Araf]^{3+}$, 545 $[m+3H\text{-}2Araf]^{3+}$, 501 $[M+3H\text{-}3Araf]^{3+}$, 475 $[M+4H]^{4+}$, 442 $[m+4H\text{-}Araf]^{4+}$, 409 $[M+4H\text{-}2Araf]^{4+}$, 376 $[M+4H\text{-}3Araf]^{4+}$. **HRMS**: calcd. for $C_{78}H_{121}N_{21}O_{34}$ 1896.8458 $[M+H]^+$, found 1896.8453. See Data S1 for analytical HPLC trace and low resolution ESI-MS spectrum.

5.5.7 Synthesis of unglycosylated CLE40a peptide

5.5.7.1 SPPS

Automated Fmoc-SPPS was carried out on a Biotage Initiator+ Alstra microwave peptide synthesizer equipped with an inert gas manifold. General synthetic protocols for Fmoc-deprotection and capping were carried out in accordance with the manufacturer's specifications. Standardized amino acid couplings were performed for 20 min at 50 °C under microwave irradiation in the presence of amino acid (0.3 M in DMF), Oxyma (0.5 M in DMF) and DIC (0.5 M in DMF).

5.5.7.2 Cleavage and purification

After washing thoroughly with CH_2Cl_2 (7 × 2 mL), the resin was suspended in a mixture of TFA, triisopropylsilane and water (90:5:5 v/v/v, 40 mL.mmol^{-1} of peptide). The suspension was shaken for 2.5 h and then filtered. The resin was washed with additional TFA, and the combined filtrates were concentrated under a stream of nitrogen gas. The residue was suspended in Et_2O and centrifuged. After pouring off the supernatant, the pellet was dissolved in H_2O, filtered and purified by preparative reverse phase HPLC (Waters Sunfire C18 5 µm, 30 × 150 mm, 40 mL.min^{-1}, 0 to 25% MeCN [0.1%TFA] in H_2O [0.1% TFA] over 30 min).

Peptide **2b** was prepared on a 16 µmol scale according to the general procedures outlined above. After preparative HPLC (rt ~ 16 min) and lyophilization, peptide **2b**was obtained as a fluffy white solid as the *tetrakis*(trifluoroacetate) salt (9.7 mg, 31%). **Analytical HPLC**: R_t 20.5 min (0 to 30% MeCN [0.1% TFA] in H_2O [0.1% TFA] over 30 min, sample dissolved in H_2O, λ = 214 nm). **LRMS**: m/z 1502 $[M+H]^+$, 751 $[M+2H]^{2+}$, 501 $[M+3H]^{3+}$, 376 $[M+4H]^{4+}$. **HRMS**: calcd. for $C_{63}H_{97}N_{21}O_{22}$ 1500.7200 $[M+H]^+$, found 1500.7153. See Data S1 for analytical HPLC trace and low resolution ESI-MS spectrum.

5.5.8 Peptide feeding and root length analysis

Five day-old seedlings (three days after being transplanted to growth pouches) were treated with either hydroxylated peptide **2b**, arabinosylated peptide **2a**, or autoclaved MilliQ® water. For each treatment, 10 µL was applied every 12 hours directly to the tip of

the tap root via small incisions made in the growth pouch. Incisions were subsequently sealed with tape and the tap root length measured. In all treatments n = 9 to 15 plants.

5.5.9 Cloning the GmCLE40a promoter region

A 2.5 kb promoter region located directly upstream of *GmCLE40a* (Glyma.12G054900) was cloned into pGEM®-T easy and subsequently ligated immediately adjacent to the *GUS* coding sequence of modified pCAMBIA1305.1 (pCAMBIA1305.1-Δ35s×2; lacking the duplicated CaMV35s promoter sequence) using T4 DNA ligase (Promega). Positive pro*GmCLE40a*::*GUS* constructs were confirmed by colony PCR and sequencing (Australian Genome Research Facility, Brisbane, Qld, Australia) and transformed from their XL1-Blue *E. coli* strains into electrocompetent *Agrobacterium rhizogenes* K599. Successful transformation was confirmed by PCR prior to use. All primers used in this study are provided in the Key Resources Table.

5.5.10 GUS histochemical assay

GUS activity of transgenic hairy roots was assessed using methods modified from Larkin et al. (Larkin et al. 1996) with X-Gluc staining buffer made using 0.3% (v/v) DMSO in place of 0.1% (v/v) Triton X-100. Harvested hairy roots were treated with fixation buffer (0.5% w/v paraformaldehyde in 100 mM sodium phosphate buffer, pH = 7.2) on ice and under vacuum, rinsed five times with 100 mM sodium phosphate buffer (pH = 7), vacuum infiltrated with X-Gluc staining buffer three times and incubated overnight at 37°C. Stained hairy roots were examined using a clearing solution (Lux et al. 2005) under light microscopes (Nikon models: C-PS/Eclipse E600W).

5.5.11 Bioinformatic analysis

The amino acid sequences of AtCLE40(AT2G27250), and GmCLE40a were used to BLAST for potential orthologues across available genome sequences in Phytozome (https://phytozome.jgi.doe.gov/ (Goodstein et al. 2012). The Phylogenetic tree was created using methods described in Hastwell et al. (Hastwell et al. 2015b) with 1000 bootstrap replications. Geneious Pro v10.0.2 (Kearse et al. 2012) was used to generate the sequence logo of the CLE domain.

5.5.12 NMR conformational analysis

Peptide **2b** was dissolved in 300 µL of 20 mM sodium phosphate (pH 6.5) to a concentration of 3.1 mM. D_2O (15 µL) and 4,4-dimethyl-4-silapentane-1-sulfonic acid (to a final concentration of 10 µM) were added. Peptide **2a** was prepared in the same manner to a final concentration of 2.8 mM.

Spectra were recorded on a Bruker Avance III 600-MHz spectrometer equipped with a TCI cryoprobe. DQF-COSY, 2D TOCSY (mixing time = 70 ms) and NOESY (mixing time = 300 ms) spectra were recorded at both 278 and 298 K. Spectra were analyzed using SPARKY 3.11 (UCSF).

5.5.13 Quantification and statistical analysis

Data are expressed as mean ± SEM. Statistical differences between treatments were determined using Student's t-test as described by Ferguson et al. (Ferguson et al. 2014), with the exception of the growth rate analyses, which were done using the Repeated Measures ANOVA. The statistical details of experiments can be found in the figure legends and Results. The n = 9 to 15 plants refers to biological replicates in each treatment group, where each biological replicate is an individual plant. All statistical differences were calculated in GraphPad Prism 7.01 (La Jolla California, USA).

5.6 Acknowledgements

This work was supported by an Australian Research Council Future Fellowship (FT130100150) to R.J.P. The work was also funded by the Hermon Slade Foundation, and Australian Research Council Discovery Project grants (DP130103084 and DP130102266) to B.J.F. and P.M.G. The Fellowship Fund Inc. is also thanked for provision of a Molly-Budtz Olsen PhD fellowship to A.H.H. We also gratefully acknowledge the funding provided to L.C. by the John A. Lamberton research scholarship and the Agnes Campbell postgraduate prize. We would like to thank Huanan Su, Xitong Chu, and Dongxue Li for their technical assistance.

5.7 References

Broughton W, Dilworth M. 1971. Control of leghaemoglobin synthesis in snake beans. *Biochemical Journal* **125**(4): 1075-1080.

Clark SE, Running MP, Meyerowitz EM. 1993. CLAVATA1, a regulator of meristem and flower development in Arabidopsis. *Development* **119**(2): 397-418.

Clark SE, Running MP, Meyerowitz EM. 1995. CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. *Development* **121**(7): 2057-2067.

Crich D, Pedersen CM, Bowers AA, Wink DJ. **2007.** On the use of 3, 5-O-benzylidene and 3, 5-O-(di-tert-butylsilylene)-2-O-benzylarabinothiofuranosides and their sulfoxides as glycosyl donors for the synthesis of β-arabinofuranosides: importance of the activation method. *The Journal of Organic Chemistry* **72**(5):1553-1565.

Désiré J, Prandi J. 1999. Synthesis of methyl β-d-arabinofuranoside 5-[1d (and l)-myo-inositol 1-phosphate], the capping motif of the lipoarabinomannan of Mycobacteriumsmegmatis. *Carbohydrate Research* **317**(1):110-118.

Ferguson BJ, Li D, Hastwell AH, Reid DE, Li Y, Jackson SA, Gresshoff PM. 2014. The soybean (*Glycine max*) nodulation-suppressive CLE peptide, GmRIC1, functions interspecifically in common white bean (*Phaseolus vulgaris*), but not in a supernodulating line mutated in the receptor PvNARK. *Plant Biotechnology Journal* **12**(8): 1085-1097.

Ferguson BJ, Mathesius U. 2014. Phytohormone regulation of legume-rhizobia interactions. *Journal of Chemical Ecology* **40**(7): 770-790.

Fiers M, Golemiec E, Xu J, van der Geest L, Heidstra R, Stiekema W, Liu CM. 2005. The 14–amino acid CLV3, CLE19, and CLE40 peptides trigger consumption of the root meristem in Arabidopsis through a CLAVATA2-dependent pathway. *The Plant Cell* **17**(9): 2542-2553.

Fletcher JC, Brand U, Running MP, Simon R, Meyerowitz EM. 1999. Signaling of cell fate decisions by CLAVATA3 in Arabidopsis shoot meristems. *Science* **283**(5409): 1911-1914.

Gadikota RR, Callam CS, Wagner T, Del Fraino B, Lowary TL. 2003. 2, 3-Anhydro sugars in glycoside bond synthesis. Highly stereoselective syntheses of oligosaccharides containing α-and β-arabinofuranosyl linkages. *Journal of the American Chemical Society* **125**(14):4155-4165.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**(D1): D1178-D1186.

Greb T, Lohmann JU. 2016. Plant Stem Cells. *Current Biology* **26**(17): R816-R821.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**(5): 696-704.

Hastwell AH, Gresshoff PM, Ferguson BJ. 2015a. Genome-wide annotation and characterization of CLAVATA/ESR (CLE) peptide hormones of soybean (*Glycine max*) and common bean (*Phaseolus vulgaris*), and their orthologues of *Arabidopsis thaliana*. *Journal of Experimental Botany* **66**(17): 5271-5287.

Hastwell AH, Gresshoff PM, Ferguson BJ. 2015b. The structure and activity of nodulation-suppressing CLE peptide hormones of legumes. *Functional Plant Biology* **42**(3): 229-238.

Hastwell AH, de Bang TC, Gresshoff PM, Ferguson BJ. 2017. CLE peptide-encoding gene families in *Medicago truncatula* and *Lotus japonicus*, compared with those of soybean, common bean and Arabidopsis. *Scientific Reports* **7**, 9384.

Hayashi S, Reid DE, Lorenc MT, Stiller J, Edwards D, Gresshoff PM, Ferguson BJ. 2012. Transient Nod factor-dependent gene expression in the nodulation-competent zone of soybean (*Glycine max* [L.] Merr.) roots. *Plant Biotechnology Journal* **10**(8): 995-1010.

Hayashi S, Gresshoff PM, Ferguson BJ. 2014. Mechanistic action of gibberellins in legume nodulation. *Journal of Integrative Plant Biology* **56**(10):971-978.

Hobe M, Müller R, Grünewald M, Brand U, Simon R. 2003. Loss of CLE40, a protein functionally equivalent to the stem cell restricting signal CLV3, enhances root waving in Arabidopsis. *Development Genes and Evolution* **213**(8): 371-381.

Ishiwata A, Akao H, Ito Y. 2006. Stereoselective synthesis of a fragment of mycobacterial arabinan. *Organic Letters* **8**(24):5525-5528.

Ishiwata A, Munemura Y, Ito Y. 2008. NAP ether mediated intramolecular aglycon delivery: a unified strategy for 1, 2-cis-glycosylation. *European Journal of Organic Chemistry* ***2008***(25):4250-4263.

Ishiwata A, Kaeothip S, Takeda Y, Ito Y. 2014. Synthesis of the highly glycosylated hydrophilic motif of extensins. *Angewandte Chemie* **126**(37):9970-9974.

Ito Y, Nakanomyo I, Motose H, Iwamoto K, Sawa S, Dohmae N, Fukuda H. 2006. Dodeca-CLE peptides as suppressors of plant stem cell differentiation. *Science* **313**(5788): 842-845.

Kaeothip S, Boons G-J. 2013. Chemical synthesis of β-arabinofuranosyl containing oligosaccharides derived from plant cell wall extensins. *Organic & Biomolecular Chemistry* **11**(31):5136-5146.

Kaeothip S, Ishiwata A, Ito Y. 2013. Stereoselective synthesis of Arabidopsis CLAVATA3 (CLV3) glycopeptide, unique protein post-translational modifications of secreted peptide hormone in plant. *Organic & Biomolecular Chemistry* **11**(35):5892-5907.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**(12): 1647-1649.

Kereszt A, Li D, Indrasumunar A, Nguyen CD, Nontachaiyapoom S, Kinkema M, Gresshoff PM. 2007. *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. *Nature Protocols* **2**(4): 948-952.

Kinoshita A, Nakamura Y, Sasaki E, Kyozuka J, Fukuda H, Sawa S. 2007. Gain-of-function phenotypes of chemically synthetic CLAVATA3/ESR-related (CLE) peptides in *Arabidopsis thaliana* and *Oryza sativa*. *Plant and Cell Physiology* **48**(12): 1821-1825.

Kondo T, Yokomine K, Nakagawa A, Sakagami Y. 2011. Analogs of the CLV3 peptide: synthesis and structure–activity relationships focused on proline residues. *Plant and Cell Physiology* **52**(1): 30-36.

Kucukoglu M, Nilsson O. 2015. CLE peptide signaling in plants–the power of moving around. *Physiologia Plantarum* **155**(1):74-87.

Larkin P, Gibson J, Mathesius U, Weinman J, Gartner E, Hall E, Tanner G, Rolfe B, Djordjevic M. 1996. Transgenic white clover. Studies with the auxin-responsive promoter, GH3, in root gravitropism and lateral root development. *Transgenic Research* **5**(5):325-335.

Lee YJ, Lee K, Jung EH, Jeon HB, Kim KS. 2005. Acceptor-dependent stereoselective glycosylation: 2 ‘-CB glycoside-mediated direct β-d-arabinofuranosylation and efficient synthesis of the octaarabinofuranoside in mycobacterial cell wall. *Organic Letters* 2005, **7**(15):3263-3266.

Li Y, Singh G. 2001. Synthesis of d-arabinofuranosides using propane-1, 3-diyl phosphate as the anomeric leaving group. *Tetrahedron Letters* **42**(37):6615-6618.

Lin YH, 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**(4): 1074-1086.

Liu Q-W, Bin H-C, Yang J-S. 2013. β-Arabinofuranosylation using 5-O-(2-quinolinecarbonyl) substituted ethyl thioglycoside donors. *Organic Letters* **15**(15):3974-3977.

Lowary TL. 2003. Synthesis and conformational analysis of arabinofuranosides, galactofuranosides and fructofuranosides. *Current Opinion in Chemical Biology* **7**(6):749-756.

Lux A, Morita S, Abe J, Ito K. 2005. An improved method for clearing and staining free-hand sections and whole-mount samples. *Annals of Botany* **96**(6):989-996.

Mandal PK, McMurray JS. 2007. Pd− C-induced catalytic transfer hydrogenation with triethylsilane. *The Journal of Organic Chemistry* **72**(17):6599-6601.

Matsubayashi Y. 2014. Posttranslationally modified small-peptide signals in plants. *Annual Review of Plant Biology* **65**: 385-413.

Meister R, Rajani M, Ruzicka D, Schachtman DP. 2014. Challenges of modifying root traits in crops for agriculture. *Trends in Plant Science* **19**(12):779-788.

Oelkers K, Goffard N, Weiller GF, Gresshoff PM, Mathesius U, Frickey T. 2008. Bioinformatic analysis of the CLE signaling peptide family. *BMC Plant Biology* **8**(1): 1.

Ogawa M, Shinohara H, Sakagami Y, Matsubayashi Y. 2008. Arabidopsis CLV3 peptide directly binds CLV1 ectodomain. *Science* **319**(5861): 294-294.

Ohyama K, Shinohara H, Ogawa-Ohnishi M, Matsubayashi Y. 2009. A glycopeptide regulating stem cell fate in *Arabidopsis thaliana*. *Nature Chemical Biology* **5**(8): 578-580.

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**(1): 67-77.

Okamoto S, Shinohara H, Mori T, Matsubayashi Y, Kawaguchi M. 2013. Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. *Nature Communications* **4**: 2191.

Reid DE, Ferguson BJ, Gresshoff PM. 2011. Inoculation-and nitrate-induced CLE peptides of soybean control NARK-dependent nodule formation. *Molecular Plant-Microbe Interactions* **24**(5): 606-618.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J. 2010. Genome sequence of the palaeopolyploid soybean. N*ature* **463**(7278): 178-183.

Schoof H, Lenhard M, Haecker A, Mayer KF, Jürgens G, Laux T. 2000. The stem cell population of Arabidopsis shoot meristems is maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. *Cell* **100**(6):635-644.

Sharma VK, Ramirez J, Fletcher JC. 2003. The Arabidopsis CLV3-like (CLE) genes are expressed in diverse tissues and encode secreted proteins. *Plant Molecular Biology* **51**(3): 415-425.

Shinohara H, Matsubayashi Y. 2013. Chemical synthesis of Arabidopsis CLV3 glycopeptide reveals the impact of hydroxyproline arabinosylation on peptide conformation and activity. *Plant and Cell Physiology* **54**(3): 369-374.

Shinohara H, Moriyama Y, Ohyama K, Matsubayashi Y. 2012. Biochemical mapping of a ligand-binding domain within Arabidopsis BAM1 reveals diversified ligand recognition mechanisms of plant LRR-RKs. *The Plant Journal* **70**(5): 845-854.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7**(1): 539.

Stahl Y, Wink RH, Ingram GC, Simon R. 2009. A signaling module controlling the stem cell niche in Arabidopsis root meristems. *Current Biology* **19**(11): 909-914.

Strabala TJ, Phillips L, West M, Stanbra L. 2014. Bioinformatic and phylogenetic analysis of the CLAVATA3/EMBRYO-SURROUNDING REGION (CLE) and the CLE-LIKE signal peptide genes in the Pinophyta. *BMC Plant Biology* **14**(1): 47.

Van den Steen P, Rudd PM, Dwek RA, Opdenakker G. 1998. Concepts and principles of O-linked glycosylation. *Critical Reviews in Biochemistry and Molecular Biology* **33**(3):151-208.

Wei Z, Li J. 2016. Brassinosteroids regulate root growth, development, and symbiosis. *Molecular Plant* **9**(1):86-100.

Xie N, Taylor CM. 2010. Synthesis of a dimer of β-(1, 4)-l-arabinosyl-(2 S, 4 R)-4-hydroxyproline inspired by art v 1, the major allergen of mugwort. *Organic Letters* **12**(21):4968-4971.

Xu C, Liberatore KL, MacAlister CA, Huang Z, Chu Y-H, Jiang K, Brooks C, Ogawa-Ohnishi M, Xiong G, Pauly M. 2015. A cascade of arabinosyltransferases controls shoot meristem size in tomato. *Nature Genetics* **47**(7): 784-792.

Yamaguchi YL, Ishida T, Sawa S. 2016. CLE peptides and their signaling pathways in plant development. *Journal of Experimental Botany* **67**(16):4813-4826.

Yin H, Lowary TL. 2001. Synthesis of arabinofuranosides via low-temperature activation of thioglycosides. *Tetrahedron Letters* **42**(34):5829-5832.

Zhang Y, Yang S, Song Y, Wang J. 2014. Genome-wide characterization, expression and functional analysis of CLV3/ESR gene family in tomato. *BMC Genomics* **15**(1):1.

Zhu X, Kawatkar S, Rao Y, Boons G-J. 2006. Practical approach for the stereoselective introduction of β-arabinofuranosides. *Journal of the American Chemical Society* **128**(36):11948-11957.

Chapter 6

Triarabinosylation is required for nodulation-suppressive CLE peptides to systemically inhibit nodulation in *Pisum sativum*

Preface

This chapter uses the triarabinose building block developed in Chapter 5 (Corcilius et al. 2017) to synthesise the nodulation-suppressive CLE peptides of soybean and has been published in Plant, Cell and Environment (https://doi.org/10.1111/pce.13325). Functional characterisation in pea *wild-type* and nodulation mutants was then performed using variations of the peptides to determine if nodule number could be altered. Subsequently, using methods developed in Chapters 3 and 4 (Hastwell et al. 2015a, 2017), the CLE-peptide encoding genes of pea were identified. Additional supplementary tables and figures can also be found online at https://onlinelibrary.wiley.com/doi/abs/10.1111/pce.13325).

6.1 Abstract

Legumes form root nodules to house beneficial nitrogen-fixing rhizobia bacteria. However, nodulation is resource demanding; hence, legumes evolved a systemic signalling mechanism, called Autoregulation of Nodulation (AON), to control nodule numbers. AON begins with the production of CLE peptides in the root, which are predicted to be glycosylated, transported to the shoot, and perceived. We synthesised variants of nodulation-suppressing CLE peptides to test their activity using petiole feeding to introduce CLE peptides into the shoot. Hydroxylated, monoarabinosylated and triarabinosylated variants of soybean GmRIC1a and GmRIC2a were chemically synthesised and fed into recipient *Pisum sativum* (pea) plants, which were used due to the availability of key AON pathway mutants unavailable in soybean. Triarabinosylated GmRIC1a and GmRIC2a suppressed nodulation of wild-type pea, whereas no other peptide variant tested had this ability. Suppression also occurred in the supernodulating hydroxyproline *O*-arabinosyltransferase mutant, *Psnod3*, but not in the supernodulating receptor mutants, *Pssym29*, and to some extent, *Pssym28*. During our study, bioinformatic resources for pea became available and our analyses identified 40 CLE peptide-encoding genes, including orthologues of nodulation-suppressive CLE peptides. Collectively, we demonstrated that soybean nodulation-suppressive CLE peptides can function interspecifically in the AON pathway of pea and require arabinosylation for their activity.

6.2 Introduction

Legumes are important in agriculture systems as a means to alleviate nitrogen fertiliser inputs, thus reducing fossil fuel use, fertiliser run-off and toxic gas emissions (Gresshoff *et al*., 2015; Foyer *et al*., 2016). They also promote soil health by increasing nitrogen levels through a mutualistic symbiotic relationship with bacteria (collectively known as rhizobia) that can convert atmospheric nitrogen gas (N_2) into a form of nitrogen the plant can use (NH_4^+). Agricultural practices take advantage of this, with legumes often used as rotation or cover crops (Jensen *et al*., 2012). Although the symbiosis is beneficial, the host plant regulates the number of nodules it forms as a means of balancing its need for nitrogen with its ability to expend resources forming and maintaining nodule structures. Thus, legumes have complex molecular signalling cascades to control nodulation (Ferguson *et al*., 2010; Reid *et al*., 2011b; Ferguson *et al*., 2018).

A systemic negative feedback signalling pathway that provides legumes with control over their nodule numbers is known as Autoregulation of Nodulation (AON; Kosslak and Bohlool 1984; Delves *et al*., 1986; Reid *et al*., 2011b). The AON pathway begins in response to initial rhizobia infection events, with the production of CLAVATA3/Endosperm Surrounding Region (ESR) related (CLE) peptides. In soybean, these peptides are GmRIC1 and GmRIC2 (Reid *et al*., 2011a), with orthologues in other legumes having also been identified (Okamoto *et al*., 2009; Mortier *et al*., 2010 Reid *et al*., 2011a; Ferguson *et al*., 2014; Nishida *et al*., 2016). While there is no clear distinction between the biological role of GmRIC1 and GmRIC2, there is some temporal separation in their expression patterns (Reid *et al*., 2011a). The AON CLE peptides are produced in the root, post-translationally modified (Okamoto *et al*., 2013; Kassaw et al. 2017), then transported to the shoot where they are perceived by a leucine-rich repeat receptor kinase, called GmNARK in soybean (known orthologues include PvNARK, LjHAR1, MtSUNN, PsSYM29, and GsNARK; Krusell et al. 2002; Nishimura et al. 2002; Searle et al, 2003; Schnabel et al. 2005; Ferguson et al. 2014). CLV2/SYM29 and KLAVIER are proposed to form a heterodimeric complex with NARK (which might also form a homodimer complex) to perceive the CLE peptides, with mutations in either NARK or its dimerisation partners resulting in supernodulation (Miyazawa et al. 2010; Ferguson et al. 2010; Krusell et al. 2011). Interestingly, the homeologous duplicate of GmNARK, called GmCLV1A, has no role in nodulation control, but instead functions in regulating shoot architecture, indicating

that one of the genes has undergone the process of neofunctionalisation (Mirzaei et al. 2017). Following ligand binding by GmNARK, a shoot-derived signal that is proposed to be transported to the root to inhibit further nodulation events is differentially regulated (Lin et al. 2010; Ferguson et al. 2010; Sasaki et al. 2014; Ferguson et al. 2018). This signal might act through the Kelch-Repeat F-box factor Too Much Love (TML), to regulate nodulation, as mutations in its gene also lead to a lack of nodulation control (Magori et al. 2009).

CLE peptides are 12-13 amino acids long, with the few that have been structurally confirmed having a central proline residue that is post-translationally hydroxylated and further modified with a triarabinose moiety containing β1,2 linkages (Shinohara and Matsubayashi 2013, Okamoto et al. 2013; Ferguson and Mathesius 2014; Hastwell et al. 2015b; Okamoto *et al*. 2015; Xu et al. 2015). When synthetic CLE peptides possess this glycan, binding efficiency is increased (AtCLV3; Shinohara and Matsubayashi, 2013) and they exhibit increased biological activity (LjCLE-RS2, Okamoto et al. 2013; GmCLE40a, Corcilius et al. 2017). This modification is likely facilitated by an arabinosyltransferase related to AtHPAT3 (Ogawa-Ohnishi *et al*. 2013; Xu *et al.* 2015), called MtRDN1/PsNOD3 in the case of the AON CLE peptides (Schnabel et al. 2011). Interestingly, only one rhizobia-induced CLE peptide of *M. truncatula*, MtCLE12, appears to require arabinosylation by MtRDN1, whereas MtCLE13 does not (Kassaw et al. 2017).

A similar mechanism to AON, called the nitrate-regulation of nodulation pathway, acts locally and is induced by soil nitrate to enable the plant to inhibit nodulation when ample nitrogen is available (Reid et al. 2011a). This nitrate-regulation of nodulation pathway begins with the production of nitrate-induced CLE peptides (called GmNIC1a and its duplicate GmNIC1b in soybean) which are perceived by the GmNARK receptor located in the root (Reid et al. 2011a; Lim et al. 2014). CLE peptides induced by nitrate to regulate nodulation have not been reported in most other legumes, with the exception of *L. japonicus* where the rhizobia-induced CLE peptides *LjCLE-RS2*, *LjCLE-RS3* and *LjCLE40* are reported to exhibit increased expression with nitrate application (Okamoto et al. 2009; Nishida et al. 2016).

Here, we report that novel triarabinosylated peptides, GmRIC1a and GmRIC2a, of soybean suppress nodulation in pea. This was demonstrated using petiole feeding of peptides that were synthesised by solid-phase peptide synthesis (SPPS) using a synthetic β1,2 triarabinosylated hydroxyproline glycosylamino acid building block (Corcilius et al. 2017) to site selectively incorporate the glycan at position seven of the CLE domain. Using

AON mutant plants, defective in controlling nodule numbers, we showed that the suppressive activity required the PsSYM28 and PsSYM29 receptors, but acted downstream of the PsNOD3 arabinosyltransferase that post-translationally glycosylates the endogenous peptides. Chemically synthesised variants of GmRIC1a and GmRIC2a that were either hydroxylated-only or partially glycosylated were unable to suppress nodulation, demonstrating that triarabinosylation is required for these peptides to function in AON. Subsequently, pea orthologues of the nodulation-suppressive CLE peptides were determined from 40 CLE peptide-encoding gene family members identified in this study. The CLE peptide domains of these pea orthologues were almost identical to those of the soybean peptides fed in this study. Taken together, our findings demonstrate a clear requirement for GmRIC1a and GmRIC2a to be post-translationally modified with a triarabinosylated hydroxyproline moiety to exert their nodulation-suppressive activity.

6.3 Materials and methods

6.3.1 Plant and bacterial growth

Wild-type and mutant *Pisum sativum* (pea) cv Frisson seeds (Postma et al. 1988; Duc and Messager et al. 1989; Sagan and Duc 1996; Li et al. 2009) were sterilised with 70% w/v ethanol before being imbibed with autoclaved Milli-Q® water. Imbibed seeds were germinated in 4 L euro pots with sterile Grade 3 vermiculite topped with approximately 3 cm of autoclaved UQ23 Mix (Central Glasshouse Services, University of Queensland, Australia) to assist germination. All plants were grown in either a E-75L1 or PGC-9/2 growth chamber (Percival Scientific, Perry, IA, USA) under 25°C:23°C, 12 hour day:night conditions. The short-day length condition induced longer internodes to assist with petiole feeding. Plants were watered as required (approximately twice per week) with B & D nutrient solution (Broughton and Dilworth, 1971), supplemented with 1 mM KNO_3, which promotes plant growth but does not inhibit nodulation (Carroll et al. 1985).

Rhizobium leguminosarum RLV248 was grown in liquid yeast mannitol broth (Somerville and Kahn, 1983) at 28°C for 36 hours and diluted to OD=0.1 with either ddH_2O or B & D nutrient solution. Approximately 250 mL of inoculum was applied to each pot 48 hours after petiole feeding commenced (three weeks following germination) and nodule number was counted 14 days after inoculation.

6.3.2 Petiole feeding

Petiole feeding was carried out as per Lin et al. (2010; 2011) with the following modifications. The second petiole of three-week old pea plants was used in the first instance to attach the petiole feeding apparatus. The apparatus consisted of a 3 mL syringe barrel attached to 20 mm of clear silicone tubing having a 2.6 mm internal diameter. This was subsequently connected to 4 cm of silicone tubing having a 1.6 mm internal diameter, which was an appropriate size for attaching to the petiole of the pea plants. The petiole was severed behind the first leaflet, and the basal stipules were left intact and used to help seal the petiole-tubing junction. After one week of feeding, the petioles became chlorotic and the feeding solution (control or peptide) ceased to be taken up by the plant. Thus, a fresh feeding apparatus was attached to a new petiole (usually two higher than the originally-fed petiole). To prevent any loss of peptide solution due to leakage, approximately 500 µL of autoclaved Milli-Q® water was injected into the silicone tubing of the newly attached feeding apparatus and left for 30 minutes prior to adding peptide solutions. Blue food colouring was used in preliminary studies to visualise uptake and ensure solutions were distributed throughout the plant.

6.3.3 Chemical synthesis of GmRIC1a and GmRIC2a (glyco)peptides

GmRIC1a and GmRIC2a peptides were synthesized via solid-phase peptide synthesis (SPPS) according to a previously reported procedure (Corcilius et al. 2017). Six synthetic peptides were prepared in total, each containing hydroxyproline at position 4, and either hydroxyproline, *O*-(β-L-arabinofuranosyl) hydroxyproline (monoarabinosylated hydroxyproline) or *O*-[β-(β1,2-tri-L-arabinofuranosyl)] hydroxyproline (triarabinosylated hydroxyproline) at position 7 of the CLE domain. Synthetic peptides were purified by reversed phase HPLC and characterized by analytical HPLC and both low and high resolution ESI-MS (+ve ion) (see Supporting Information for synthetic peptide characterization data).

6.3.4 Sequence identification and bioinformatic analysis

CLE peptide encoding genes in *Pisum sativum* were identified using BLAST searches of known legume genes identified in Hastwell et al. (2015a and 2017) as well as those from *Arabidopsis thaliana* (Cock and McCormick 2001) with E value = 1 (Altschul et al. 1997

and 2005). The searches were conducted in The Pea RNA-Seq gene atlas (http://bios.dijon.inra.fr/FATAL/cgi/pscam.cgi; Alves-Carvalho et al. 2015). Multiple Sequence Alignments, logo diagrams, signal peptide and phylogenetic analyses were performed as per Hastwell et al. (2015a and 2017).

6.3.5 Statistical analyses

Student's *t*-tests were used to determine statistical differences between treatments and were calculated in GraphPad Prism 7.01 (La Jolla California, USA; *P<0.5, **P<0.01, ***P<0.001). Data are expressed as a mean ±SEM, with n = 6 to 8 plants per treatment, except for untreated plants where n = 14.

6.4 Results

6.4.1 Establishment of petiole feeding as a method to introduce solutions into pea plants

During AON, root-derived CLE peptides travel in the xylem to the shoot, where they are perceived by an LRR receptor kinase (Searle et al. 2003; Reid et al. 2011a; Okamoto et al. 2013). However, feeding CLE peptides to the root can have unwanted false-positive effects, with many inhibiting root growth due to functional redundancy and interacting with other receptors (Whitford et al. 2008; Shinohara and Matsubayashi 2015). Thus, a direct-feeding method to introduce the peptide closer to its correct receptor was desired. Petiole feeding achieves this (Lin et al. 2010, 2011), and pea was selected as the recipient species due to the availability of multiple pea mutants in the AON pathway. When this study commenced, CLE peptide sequences of pea were not available. We therefore focused on GmRIC1a and GmRIC2a of soybean as they have been shown to act interspecifically in other legume species using overexpression studies (Ferguson et al. 2014).

Preliminary experiments feeding water or dye revealed no observable differences in shoot or root weight, shoot height or node number between intact and petiole-fed pea plants (Figure 6.1A, Supplementary Figure 1). This confirmed that petiole feeding could be used to introduce and translocate solutions throughout the plant, and did not induce unwanted

effects, which is consistent with previous reports using other plant species (Lin et al. 2010, 2011).

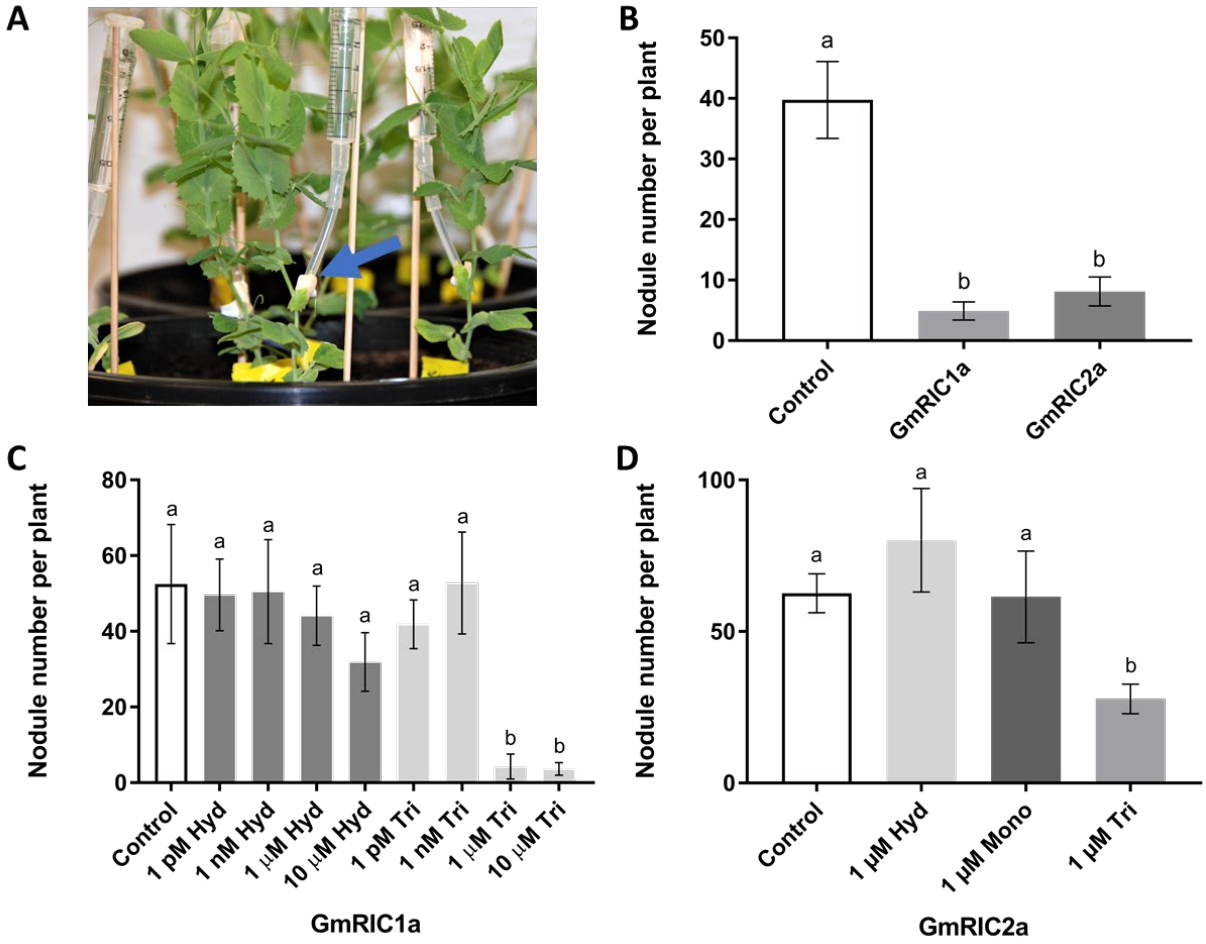


Figure 6.1 Peptide petiole feeding and subsequent nodule number 14 days after inoculation of wild-type pea plants **A** Image of pea plants with petiole feeding apparatus attached (arrow). **B** 1 μM triarabinosylated GmRIC1a, GmRIC2a and water control. **C** 1 pM to 10 μM of hydroxylated (Hyd) or triarabinosylated (Tri) GmRIC1a. **D** 1 μM triarabinosylated (Tri), monoarabinosylated (Mono) or hydroxylated (Hyd) GmRIC2a, and water control. Statistical differences determined using Student's *t*-test. n = 7 to 10 plants per treatment and error bars represent standard error of the mean.

6.4.2 Chemical synthesis of GmRIC1a and GmRIC2a glycopeptide variants

Methods to extract and purify sufficient quantities of endogenous CLE glycopeptides have not been established and therefore chemical synthesis is the only tool available to access CLE glycopeptides for feeding studies. However, this is a considerable undertaking when post-translational modifications are taken into account because of the synthetically-challenging nature of the glycan (Kaeothip and Boons 2013). Despite this challenge, three

successful syntheses of an SPPS-compatible triarabinosylated hydroxyproline 'building block' have been reported (Shinohara and Matsubayashi 2013; Kaeothip et al. 2013; Corcilius et al. 2017) along with examples of its incorporation into native CLE peptides (Shinohara and Matsubayashi 2013, Okamoto et al. 2013; Xu et al. 2015; Corcilius et al. 2017). The most advanced protocol for the synthesis of this triarabinosylated hydroxyproline building block (Figure 6.2, in box) was recently reported (Corcilius et al. 2017), and used in this study to access multi-milligram quantities of homogeneous hydroxyproline-7 triarabinosylated GmRIC1a and GmRIC2a glycopeptides. Briefly, the building block was incorporated into conventional Fmoc-SPPS protocols to obtain the resin-bound and side chain-protected glycopeptides, which were subsequently liberated from the resin and deprotected through treatment with an acidic cleavage cocktail containing trifluoroacetic acid (TFA), triisopropylsilane and water. After deacetylation of the glycan with sodium methoxide in methanol, the residues were purified by preparative reversed phase HPLC affording GmRIC1a and GmRIC2a glycopeptides as their corresponding trifluoroacetate salts in 17% and 28% overall yield, respectively (yield based on initial resin loading of the C-terminal amino acid). The corresponding hydroxyproline-7 monoarabinosylated and unglycosylated variants were also synthesised in order to probe the functional importance of the triarabinosylation modification (Figure 2). All variants were prepared with hydroxyproline at position 4 in analogy with the structures of known CLE peptides.

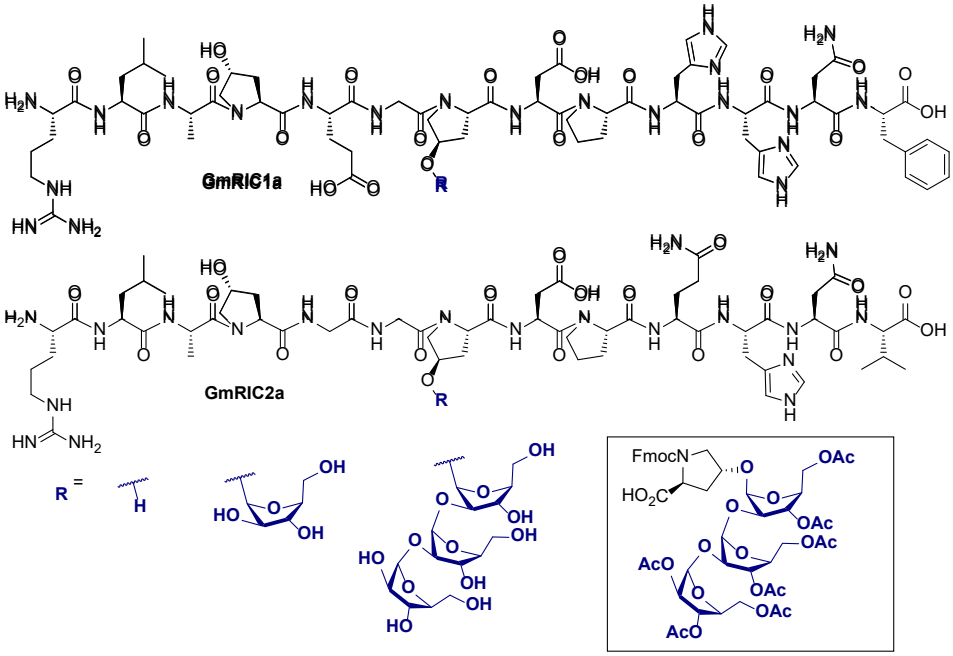


Figure 6.2 Structures of synthetic GmRIC1a and GmRIC2a peptides, and triarabinosylated hydroxyproline building block (in box). Proline 4 is hydroxylated in all variants. Proline 7 is either hydroxylated only (R = H), or further modified by arabinosylation (R = monoarabinose or triarabinose).

6.4.3 GmRIC1a and GmRIC2a glycopeptides suppress nodulation in pea

Petiole feeding was used to determine whether GmRIC1a and GmRIC2a peptide variants could inhibit nodulation in pea. Soybean CLE peptides were used, rather than those of pea, as the transcriptome database enabling identification of pea CLE peptide-encoding gene sequences was not available when this study commenced. The GmRIC1a variants tested had the proline residues at positions four and seven hydroxylated, with or without triarabinosylation at position seven (Figure 6.2), and were fed at concentrations from 1 pM to 10 µM. CLE peptides with no modifications have previously been reported to have no nodulation-suppressive activity and were not used in this study (Okamoto et al. 2009; Mortier et al. 2010).

Nodule inhibition was observed in plants fed with 1 µM or higher of the triarabinosylated variant of GmRIC1a (Figure 6.1, Supplementary Figure 2). In contrast, no significant difference in nodule number was observed with any concentration of the hydroxylated-only variant (Figure 6.1, Supplementary Figures 2 and 3). Triarabinosylated GmRIC1a and GmRIC2a both inhibited nodule number at 1 µM and peptides at this concentration were used in subsequent experiments. Together, this indicates that triarabinosylation is required for the peptides to exert their activity.

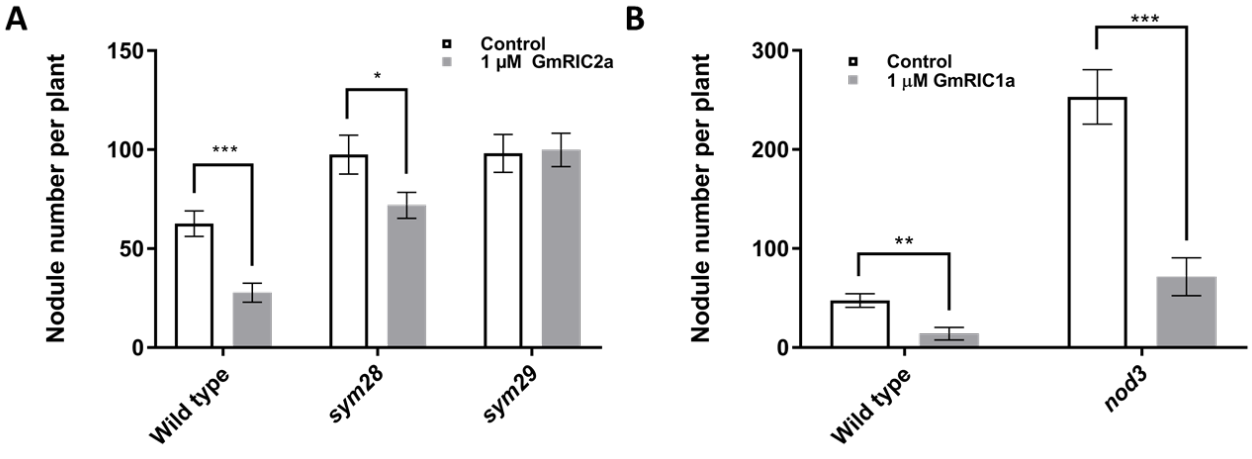


Figure 6.3 Nodule number 14 days after inoculation of wild-type and nodulation-mutant pea plants fed via petiole feeding with either 1 µM triarabinosylated GmRIC1a, triarabinosylated GmRIC2a, or water control. **A** Wild type, *sym28* and *sym29* plants fed with GmRIC2a. **B** Wild type and *nod3* plants fed with GmRIC1a. Statistical differences determined using Student's *t*-test. n = 5-8 plants per treatment and error bars represent standard error of the mean.

6.4.4 The extent of glycosylation can affect the efficacy of CLE peptide activity

All CLE peptides identified to date have been modified with three linked arabinose sugars at their central proline residue. To determine whether these three arabinose sugars are required to suppress nodulation, wild-type pea plants were fed with either the triarabinosylated or monoarabinosylated variant of GmRIC1a. While the triarabinosylated variant significantly suppressed nodulation (Figure 6.1D), the monoarabinosylated variant was unable to do so ($P>0.5$). This further demonstrates that post-translational triarabinosylation is essential for activity.

6.4.5 Nodulation suppressing CLE peptides act downstream of *PsNOD3* but require *PsSYM28* and *PsSYM29* to exert their activity

PsNOD3 encodes a hydroxyproline *O*-arabinosyltransferase (Schnabel et al. 2011) that might be required to post-translationally glycosylate mature, nodulation-suppressing CLE peptides in the root. *PsSYM28* and *PsSYM29* encode for receptors that likely form a complex to perceive nodulation-suppressing CLE peptide ligands in the shoot (Krusell et al. 2002, 2011). Overexpression of rhizobia-induced CLE peptide-encoding genes results in complete suppression of nodulation in wild-type plants of several legumes (Okamoto et al. 2009; Mortier et al. 2010; Reid et al. 2011a), but does not alter nodule numbers in supernodulating receptor mutants (Okamoto et al. 2009; Reid et al. 2011a; Osipova et al. 2012; Ferguson et al. 2014). Interestingly, *MtCLE13* overexpression suppresses nodulation in *Mtrdn1-2* (Kassaw et al. 2017), the orthologue of *PsNOD3*, but not when interspecifically overexpressed in *Psnod3* (Osipova et al. 2012). To establish whether triarabinosylated GmRIC1a or GmRIC2a can suppress nodulation in supernodulating pea mutants, plants were fed via petiole feeding and nodule numbers determined. Soybean was not utilised as there are currently no lines containing mutations in SYM28 and NOD3 orthologues.

Nodule numbers were not affected in *Pssym29* plants fed with GmRIC1a (Figure 6.3A) and only a slight but significant reduction in nodulation was observed in *Pssym28* plants (Figure 6.3A). In contrast, nodulation was significantly reduced when feeding GmRIC2a into *Psnod3* plants (Figure 6.3B). Together, this indicates that SYM29, and to a lower extent SYM28, are required for perception of the nodulation CLE peptides, and that NOD3

is indeed likely responsible for arabinosylation of the peptides, which is required for their function.

6.4.6 Functional redundancy enables other CLE peptide family members to function as nodulation-suppressing CLE peptides

To determine whether other CLE glycopeptides could mimic the activity of the nodulation suppressing CLE peptides, petiole feeding was used to introduce hydroxylated-only or triarabinosylated GmCLE40a variants into wild-type pea plants. GmCLE40a acts to regulate the stem cell population of the root apical meristem (Corcilius et al. 2017) and would not normally be expected to come into contact with receptors of the nodulation suppressing CLE peptides. The CLE domain of GmCLE40a contains six amino acid residues that differ from the GmRIC1a or GmRIC2a CLE domain. Only two of these residues (positions three and twelve) affected GmRIC1a activity when modified via site-directed mutagenesis (Reid et al. 2013). This reduction in activity was only minor at position three of GmRIC1a, and the residue at position 12 of GmCLE40a would only be considered a conservative change from that of GmRIC1a (Asp>His), and thus not likely to have a large impact on activity.

The hydroxylated GmCLE40a variant was not able to suppress nodulation (Figure 6.4), similar to what was observed with hydroxylated GmRIC1a. However, triarabinosylated GmCLE40a did suppress nodulation. In fact, it suppressed nodulation to nearly the same extent as triarabinosylated GmRIC1a (Figure 6.1). These findings demonstrate functional redundancy can occur amongst CLE peptides, and further support the conclusion that triarabinosylation of the nodulation suppressing CLE peptides is required to suppress nodulation in pea.

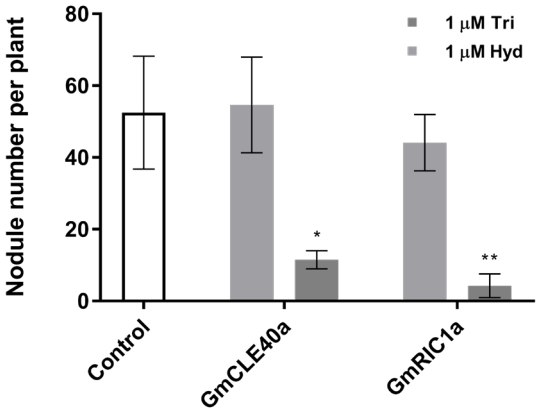


Figure 6.4 Nodule number 14 days after inoculation of wild-type pea plants fed via petiole feeding with 1 μM hydroxylated (Hyd) or triarabinosylated (Tri) GmRIC1a or GmCLE40a, or water control. Statistical differences determined using Student's *t*-test. n = 7 to 8 plants per treatment and error bars represent standard error of the mean.

6.4.7 Identification of CLE peptide-encoding genes of *Pisum sativum*

The complete genome of pea is not yet available and so we used the nodulation suppressing CLE peptides of soybean in this study. However, since commencing our work, several transcriptome analyses have become available that could be used to identify CLE peptide encoding genes of pea (Alves-Carvalho et al. 2015; Tayeh et al. 2015). To identify CLE peptide orthologues of pea, BLAST searches of the UniGene set in The Pea RNA-Seq gene atlas were conducted using CLE peptide-encoding gene sequences of *Medicago truncatula*, *Lotus japonicus*, *Phaseolus vulgaris* and *Arabidopsis thaliana* (Cock and McCormick 2001; Alves-Carvalho et al. 2015; Hastwell et al. 2015a and 2017). The search yielded 40 unique CLE peptide-encoding gene candidates of pea (Figure 6.5, Supplementary Table 1) and a further eight sequences with unclear gene structures and/or analogous CLE peptide domains (Supplementary Table 2). Three of the identified sequences contain multiple CLE domains (Supplementary Table 1, Supplementary Figure 4). It is important to note that without the genome, the entire CLE peptide encoding gene family of pea remains incomplete as only genes that were expressed in the available transcriptome datasets can be identified; hence, there are likely to be more than 40 CLE peptide members in pea.

An initial phylogenetic tree was constructed using the 40 newly identified CLE prepropeptide sequences of pea, along with those previously identified in *M*. *truncatula*, *L*. *japonicus*, *P*. *vulgaris* and *A*. *thaliana* (Supplementary Figure 5) (Cock and McCormick

2001; Hastwell et al. 2015a and 2017). This enabled homologous sequences of pea to be identified. PsCam040153 and PsCam040702 grouped closely with rhizobia-induced CLE peptides, and PsCAM041632 grouped with nitrate-induced CLE peptides (Supplementary Figure 5). An additional phylogenetic tree focusing on nodulation-suppressing CLE peptides was then generated, which included both rhizobia- and nitrate- induced CLE peptides of *G. max* and other legumes (Figure 6.6A) (Reid et al. 2011a; Okamoto et al. 2015). Unsurprisingly, PsCam040153 and PsCam040702 formed a distinct branch with the rhizobia-induced CLE peptide orthologues as in previous phylogenetic analyses (Hastwell et al. 2015a and 2017), whereas no clear branch was observed with the nitrate-induced CLE peptides, despite it grouping in the phylogenetic tree designed using with the complete family of pea CLE prepropeptides (Figure 6.6A).

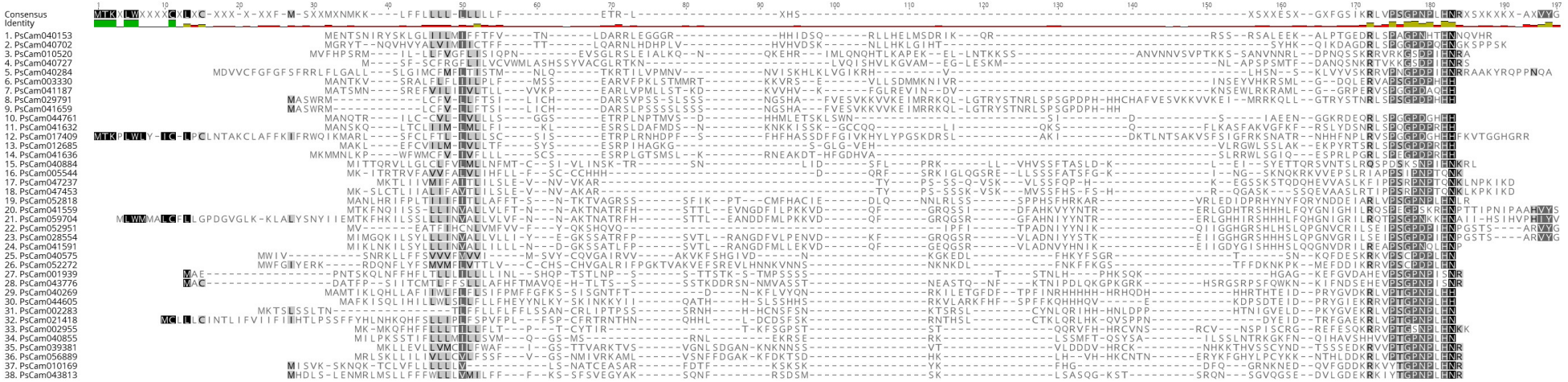


Figure 6.5 Multiple Sequence Alignment of the CLE prepropeptides of *P. sativum*. Shaded nucleotides indicate conservation. Not shown are the multi CLE domain containing prepropeptides.

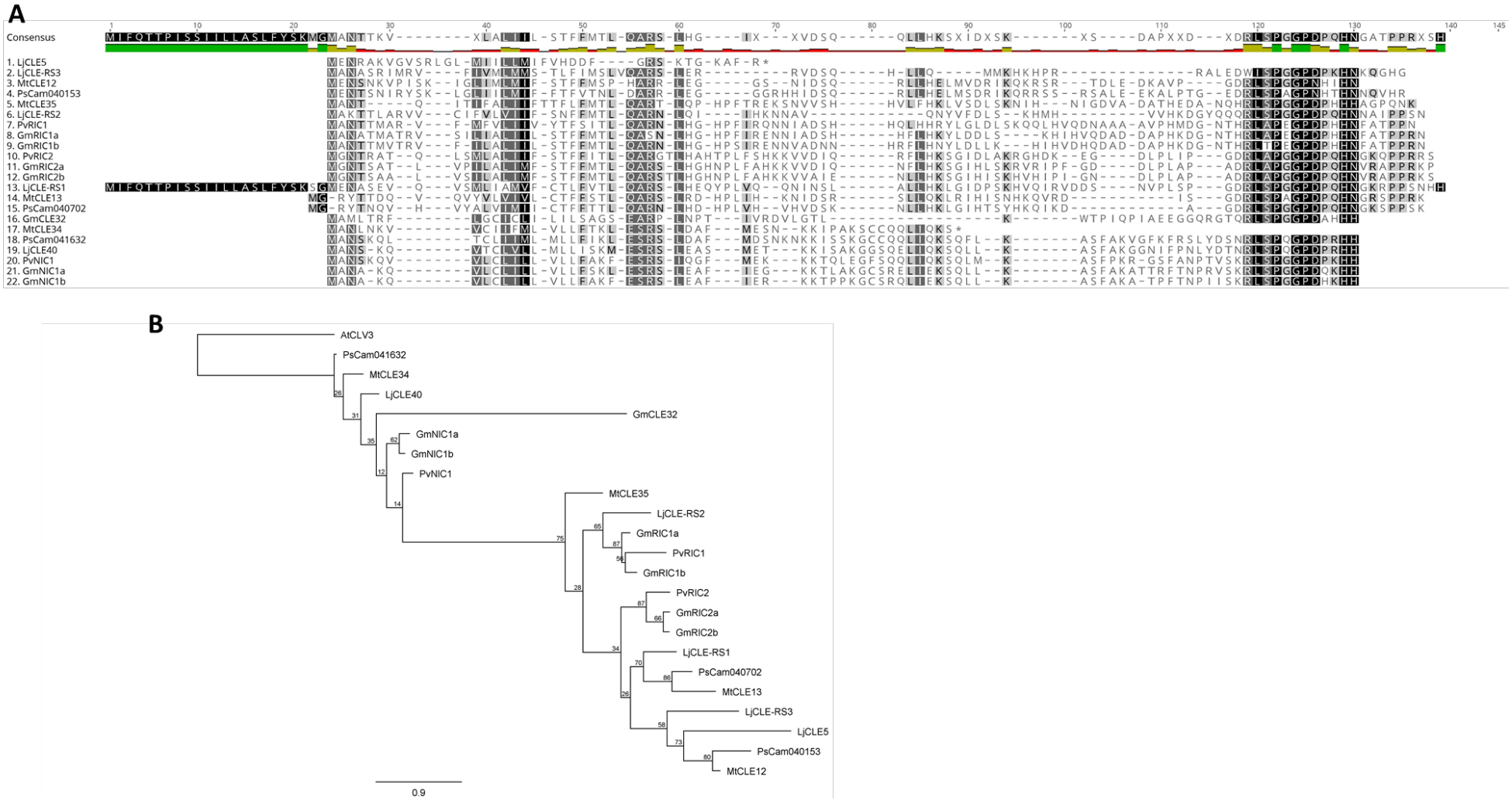


Figure 6.6 Nodulation-suppressive CLE prepropeptides of *P. sativum* and their orthologues in *G. max*, *P. vulgaris*, *L. japonicus*, and *M. truncatula*. **A** Multiple Sequence Alignment where shaded residues indicate conserved residues. **B** Phylogenetic tree, with Bootstrap confidence values expressed as a percentage of 1,000 bootstrap replications, using AtCLV3 as an outgroup.

Based on the sequence and phylogenetic analyses, PsCam040153 and PsCam040702 are the likely orthologues of the rhizobia-induced CLE peptides (GmRIC1, GmRIC2, PvRIC1, PvRIC2, MtCLE12, MtCLE13, LjCLE-RS1, LjCLE-RS2 and LjCLE-RS3). Given that the CLE domain within the prepropeptide represents the functional ligand, the amino acid sequences within that domain were compared to those of previously identified orthologues (Figure 6.6B). PsCam040153 and PsCam040702 have CLE domains that are conserved at six and seven of the eight residues, respectively, that were identified by Reid et al. (2013) as being critical to the activity of nodulation-suppressive CLE peptides (Supplementary Figure 6). However, the two non-conserved residue changes at positions three and eight of the CLE domain are conservative, Ala3>Ser3 in both sequences and Asn8>Asp8 in only PsCam040702 (Supplementary Figure 6). The former is an amino acid found at position three of the majority of CLE domains from the nodulation-suppressive CLE peptides of *M. truncatula* and *L. japonicus*. Hence, these differences seem very unlikely to impact function.

6.5 Discussion

The fundamental mechanisms that provide legumes with control over nodulation requires a better understanding to enable agricultural advances. Using synthetic variants of the nodulation-suppressing CLE peptides, GmRIC1a and GmRIC2a, we show that post-translational modification with a triarabinose moiety is required for activity. These findings are consistent with reports showing that glycosylation is essential for the activity of nodulation-suppressive CLE peptides in *L. japonicus*, CLV3 orthologues in *A. thaliana* and tomato (Shinohara and Matsubayashi 2013, Okamoto et al. 2013; Xu et al. 2015), and CLE40 in soybean (Corcilius et al. 2017).

Activity of the triarabinosylated GmRIC2a peptide is dependent on *PsSYM29*, and to some extent *PsSYM28*, which are the proposed receptors of the nodulation suppressing CLE peptides. This is consistent with over-expression studies, where the peptides acts through these receptors to inhibit nodule numbers (Krusell et al. 2011; Reid et al. 2011a; Osipova et al 2012; Ferguson et al. 2014). These findings agree with the proposed AON pathway, where the CLE peptides are perceived by a receptor complex consisting of NARK/CLV2/KLV, which triggers regulation of a downstream signal that induces nodule number regulation.

In addition to suppressing nodulation in wild-type pea, the glycosylated GmRIC1a peptide also inhibited nodulation in *Psnod3* mutants. NOD3 orthologues are hydroxyproline *O*-arbinosyltransferases proposed to be responsible for catalysing the glycosylation of some CLE peptides in AON (Kassaw et al. 2017). Our study supports the requirement for arabinosylation by NOD3; however, the precise role of the modification remains unknown and may be required for structure, perception and/or stability of the peptide (Shinohara and Matsubayashi 2013). It is also possible that another mechanism for modification is required for other CLE peptides in nodulation (Kassaw et al. 2017). The strong structural redundancy of the synthetic nodulation-suppressive CLE peptides, together with our findings of nodule inhibition using wild-type pea, indicate that GmRIC1a can likely function in *sym28* and *sym29*; and GmRIC2a could function in *nod3*. However, these were not examined due to the complex nature of the chemical synthesis resulting in limited quantities of available peptide.

The CLE peptide-encoding genes identified here considerably enhance our knowledge of the CLE peptide family of pea. The 40 genes identified include orthologues of the well-characterised CLE peptides RIC1, RIC2, NIC1, TDIF and multiple CLE domain containing prepropeptides of other species. Official gene nomenclature was not assigned to the newly identified pea genes as it is highly likely that new CLE peptide-encoding genes will be identified once the pea genome is released. When this occurs, a much more comprehensive study will be required, similar to Hastwell et al. (2015a, 2017), as current gene-identifying resources are limited to tissues and treatments that were used to generate the Pea RNA-Seq gene atlas.

Within a species, different CLE peptide-encoding genes can encode for the same mature peptide sequence, with functional specificity arising from temporal and spatial separation of gene expression in conjunction with divergent receptors. Our findings indicate that synthetic triarabinosylated GmCLE40a, which has high sequence similarity to GmRIC1a and GmRIC2a, can function in AON to suppress nodulation; however, endogenous GmCLE40a is highly unlikely come into contact with AON receptors as it is a component of root apical meristem development (Yamaguchi et al. 2016; Corcilius et al. 2017). This finding highlights the need to plan feeding studies and interpret their results with great care, similar to what has been reported for receptor binding studies that can generate false-positive outcomes (Shinohara and Matsubayashi 2015).

The available germplasm in pea made this study possible as there are multiple mutants available in the AON pathway. In contrast, the duplicated genome of soybean results in functional redundancy that makes selecting mutant lines a challenge. Typically, mutations in soybean are only isolated for duplicate genes that have undergone neofunctionalisation, such as GmNARK and GmCLV1a (Mirzaei et al. 2017). Difficulty in creating stable mutant lines also restricts mutant availability. At the beginning of this study the CLE peptide-encoding genes of pea had not been identified and the pea genome was not available to identify them. However, interspecific studies had shown that the AON mechanism is highly conserved across legumes (Osipova et al. 2012; Ferguson et al. 2014) and so the nodulation suppressing CLE peptides of soybean were used with the AON mutants of pea. Our findings reiterate the conservation of the AON pathway, even between legumes with different nodule development; soybean having determinate nodules and pea having indeterminate nodules. Subsequently, we were able to identify the likely othologues of the nodulation suppressing CLE peptides of pea and established that their mature peptide sequences are highly similar to those of soybean.

CLE peptides are important plant hormones that may provide targets for agricultural advances in nodulation as well as other aspects of plant growth and development. It is important to better understand their patterns of expression, post-translational modifications, and function in molecular signalling pathways. Using recently advanced chemical methods (Corcilius et al. 2017), we demonstrate that this can be achieved using homogeneously modified CLE peptides coupled with precise delivery techniques to reduce off-target effects. Specifically, we chemically synthesised systemically-acting nodulation-suppressive CLE peptides and used a targeted delivery method to demonstrate that they require a specific arabinosyl moiety to function. Further understanding CLE peptides and how they are post-translationally modified by NOD3 is pertinent to expand our knowledge of AON and associated pathways.

6.6 Acknowledgements

This work was funded by the Hermon Slade Foundation, and Australian Research Council Discovery Project grants (DP130103084 and DP130102266) to BJF and PMG. This work was supported by an Australian Research Council Future Fellowship (FT130100150) to RJP. The Fellowship Fund Inc. is thanked for provision of a Molly-Budtz Olsen PhD fellowship to AHH. We also gratefully acknowledge the funding provided to LC by the John

A. Lamberton research scholarship and the Agnes Campbell postgraduate prize. We would also like to thank Andreas Brust, Rob Capon, Dongxue Li, Laura Haaima, Huanan Su, Xitong Chu and Mengbai Zhang for their assistance with this study.

6.7 References

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI- BLAST: a new generation of protein database search programs. *Nucleic Acid Research* **25**(17): 3389-3402.

Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schäffer AA, Yu YK. 2005. Protein database searches using compositionally adjusted substitution matrices. *FEBS Journal* **272**(20): 5101-5109.

Alves-Carvalho S, Aubert G, Carrère S, Cruaud C, Brochot AL, Jacquin F, Klein A, Martin C, Boucherot K, Kreplak J. 2015. Full-length de novo assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *The Plant Journal* **84**(1): 1-19.

Broughton W, Dilworth M. 1971. Control of leghaemoglobin synthesis in snake beans. *Biochemical Journal* **125**(4): 1075-1080.

Carroll BJ, McNeil DL, Gresshoff PM. 1985. 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* **82**: 4162-4166.

Cock JM, McCormick S. 2001. A large family of genes that share homology with CLAVATA3. *Plant Physiology* **126**(3): 939-942.

Corcilius L, Hastwell AH, Zhang M, Williams J, Mackay JP, Gresshoff PM, Ferguson BJ, Payne RJ. 2017. Arabinosylation modulates the growth-regulating activity of the peptide hormone CLE40a from Soybean. *Cell Chemical Biology* **24**: 1347-1355

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**(2): 588-590.

Duc G, Messager A. 1989. Mutagenesis of pea (*Pisum sativum* L.) and the isolation of mutants for nodulation and nitrogen fixation. *Plant Science* **60**: 207–213.

Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE, Gresshoff PM. 2010. Molecular analysis of legume nodule development and autoregulation. *Journal of Integrative Plant Biology* **52**(1): 61-76.

Ferguson BJ, Li D, Hastwell AH, Reid DE, Li Y, Jackson SA, Gresshoff PM. 2014. The soybean (*Glycine max*) nodulation-suppressive CLE peptide, GmRIC1, functions interspecifically in common white bean (*Phaseolus vulgaris*), but not in a supernodulating line mutated in the receptor PvNARK. *Plant Biotechnology Journal* **12**(8): 1085-1097.

Ferguson BJ, Mathesius U. 2014. Phytohormone regulation of legume-rhizobia interactions. *Journal of Chemical Ecology* **40**(7): 770-790.

Ferguson BJ, Mens C, Hastwell AH, Zhang MB, Su H, Jones CH, Chu XT, Gresshoff PM. 2018. Legume nodulation: the host controls the party. *Plant, Cell & Environment*, https://doi.org/10.1111/pce.13348.

Foyer CH, Lam H-M, Nguyen HT, Siddique KH, Varshney RK, Colmer TD, Cowling W, Bramley H, Mori TA, Hodgson JM. 2016. Neglecting legumes has compromised human health and sustainable food production. *Nature Plants* **2**: 16112.

Gresshoff PM, Hayashi S, Biswas B, Mirzaei S, Indrasumunar A, Reid D, Samuel S, Tollenaere A, van Hameren B, Hastwell A. 2015. The value of biodiversity in legume symbiotic nitrogen fixation and nodulation for biofuel and food production. *Journal of Plant Physiology* **172**: 128-136.

Hastwell AH, de Bang TC, Gresshoff PM, Ferguson BJ. 2017. CLE peptide-encoding gene families *in Medicago truncatula* and *Lotus japonicus*, compared with those of soybean, common bean and Arabidopsis. *Scientific Reports* **7**.

Hastwell AH, Gresshoff PM, Ferguson BJ. 2015a. Genome-wide annotation and characterization of CLAVATA/ESR (CLE) peptide hormones of soybean (*Glycine max*) and common bean (*Phaseolus vulgaris*), and their orthologues of *Arabidopsis thaliana*. *Journal of Experimental Botany* **66**(17): 5271-5287.

Hastwell AH, Gresshoff PM, Ferguson BJ. 2015b. The structure and activity of nodulation-suppressing CLE peptide hormones of legumes. *Functional Plant Biology* **42**(3): 229-238.

Jensen ES, Peoples MB, Boddey RM, Gresshoff PM, Hauggaard-Nielsen H, Alves BJ, Morrison MJ. 2012. *C*egumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. A review. *Agronomy for Sustainable Development* **32**(2): 329-364.

Kaeothip S, Boons G-J. 2013. Chemical synthesis of β-arabinofuranosyl containing oligosaccharides derived from plant cell wall extensins. *Organic & Biomolecular Chemistry* **11**(31): 5136-5146.

Kaeothip S, Ishiwata A, Ito Y. 2013. Stereoselective synthesis of Arabidopsis CLAVATA3 (CLV3) glycopeptide, unique protein post-translational modifications of secreted peptide hormone in plant. *Organic & Biomolecular Chemistry* **11**(35): 5892-5907.

Kassaw T, Nowak S, Schnabel E, Frugoli JA. 2017. ROOT DETERMINED NODULATION1 is required for *M. truncatula* CLE12, but not CLE13 peptide signaling through the SUNN receptor kinase. *Plant Physiology* **174**(4): 2445-2456.

Kosslak RM, Bohlool BB. 1984. Suppression of nodule development of one side of a split-root system of soybeans caused by prior inoculation of the other side. *Plant Physiology* **75**(1): 125-130.

Krusell L, Madsen LH, Sato S, Aubert G, Genua A, Szczyglowski K, Duc G, Kaneko T, Tabata S, de Bruijn F. 2002. Shoot control of root development and nodulation is mediated by a receptor-like kinase. *Nature* **420**(6914): 422-426.

Krusell L, Sato N, Fukuhara I, Koch BE, Grossmann C, Okamoto S, Oka-Kira E, Otsubo Y, Aubert G, Nakagawa T. 2011. The Clavata2 genes of pea and *Lotus japonicus* affect Autoregulation of Nodulation. *The Plant Journal* **65**(6): 861-871.

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**: 955-67.

Lim CW, Lee YW, Lee SC, Hwang CH. 2014. Nitrate inhibits soybean nodulation by regulating expression of CLE genes. *Plant Science* **229**: 1-9.

Lin YH, 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**(4): 1074-1086.

Lin Y-H, Lin M-H, Gresshoff PM, Ferguson BJ. 2011. An efficient petiole-feeding bioassay for introducing aqueous solutions into dicotyledonous plants. *Nature Protocols* **6**(1): 36-45.

Magori S, Oka-Kira E, Shibata S, Umehara Y, Kouchi H, Hase Y, Tanaka A, Sato S, Tabata S, Kawaguchi M. 2009. TOO MUCH LOVE, a root regulator associated with the long-distance control of nodulation in *Lotus japonicus*. *Molecular Plant-Microbe Interactions* **22**(3): 259-268.

Mirzaei S, Batley J, El-Mellouki T, Liu S, Meksem K, Ferguson BJ, Gresshoff PM. 2017. Neodiversification of homeologous CLAVATA1-like receptor kinase genes in soybean leads to distinct developmental outcomes. *Scientific Reports* **7**(1): 8878.

Miyazawa H, Oka-Kira E, Sato N, Takahashi H, Wu G-J, Sato S, Hayashi M, Betsuyaku S, Nakazono M, Tabata S. 2010. The receptor-like kinase KLAVIER

mediates systemic regulation of nodulation and non-symbiotic shoot development in *Lotus japonicus*. *Development* **137**(24): 4317-4325.

Mortier V, Den Herder G, Whitford R, Van de Velde W, Rombauts S, D'haeseleer K, Holsters M, Goormachtig S. 2010. CLE peptides control *Medicago truncatula* nodulation locally and systemically. *Plant Physiology* **153**(1): 222-237.

Nishida H, Handa Y, Tanaka S, Suzaki T, Kawaguchi M. 2016. Expression of the *CLE-RS3* gene suppresses root nodulation in *Lotus japonicus*. *Journal of Plant Research* **129**(5): 909-919.

Nishimura R, Hayashi M, Wu G-J, Kouchi H, Imaizumi-Anraku H, Murakami Y, Kawasaki S, Akao S, Ohmori M, Nagasawa M. 2002. HAR1 mediates systemic regulation of symbiotic organ development. *Nature* **420**(6914): 426-429.

Ogawa-Ohnishi M, Matsushita W, Matsubayashi Y. 2013. Identification of three hydroxyproline O-arabinosyltransferases in *Arabidopsis thaliana*. *Nature Chemical Biology* **9**(11): 726-730.

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**(1): 67-77.

Okamoto S, Shinohara H, Mori T, Matsubayashi Y, Kawaguchi M. 2013. Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. *Nature Communications* **4**.

Okamoto S, Suzuki T, Kawaguchi M, Higashiyama T, Matsubayashi Y. 2015. A comprehensive strategy for identifying long-distance mobile peptides in xylem sap. *The Plant Journal* **84**(3): 611-620.

Osipova MA, Mortier V, Demchenko KN, Tsyganov VE, Tikhonovich IA, Lutova LA, Dolgikh EA, Goormachtig S. 2012. *WUSCHEL-RELATED HOMEOBOX5* gene expression and interaction of CLE peptides with components of the systemic control add two pieces to the puzzle of Autoregulation of Nodulation. *Plant Physiology* **158**(3): 1329-1341.

Postma JG, Jacobsen E, Feenstra WJ. 1988. Three pea mutants with an altered nodulation studied by genetic analysis and grafting. *Journal of Plant Physiology* **132**: 424–430.

Reid DE, Ferguson BJ, Gresshoff PM. 2011a. Inoculation-and nitrate-induced CLE peptides of soybean control NARK-dependent nodule formation. *Molecular Plant-Microbe Interactions* **24**(5): 606-618.

Reid DE, Ferguson BJ, Hayashi S, Lin Y-H, Gresshoff PM. 2011b. Molecular mechanisms controlling legume Autoregulation of Nodulation. *Annals of Botany* **108**(5): 789-795.

Reid DE, Li D, Ferguson BJ, Gresshoff PM. 2013. Structure–function analysis of the GmRIC1 signal peptide and CLE domain required for nodulation control in soybean. *Journal of Experimental Botany*: ert008.

Sagan M, Duc G. 1996. *Sym28* and *Sym29*, two new genes involved in regulation of nodulation in pea (*Pisum sativum* L.). *Symbiosis* **20**: 229–245.

Sasaki T, Suzaki T, Soyano T, Kojima M, Sakakibara H, Kawaguchi M. 2014. Shoot-derived cytokinins systemically regulate root nodulation. *Nature Communications* **5**.

Schnabel E, Journet E, 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.

Schnabel E, Kassaw T, Smith L, Marsh J, Oldroyd GE, Long SR, Frugoli J. 2011. ROOT DETERMINED NODULATION 1 regulates nodule number in *M. truncatula* and defines a highly conserved, uncharacterized plant gene family. *Plant Physiology* **157**(1): 328-340.

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**(5603): 109-112.

Shinohara H, Matsubayashi Y. 2013. Chemical synthesis of Arabidopsis CLV3 glycopeptide reveals the impact of hydroxyproline arabinosylation on peptide conformation and activity. *Plant and Cell Physiology* **54**(3): 369-374.

Shinohara H, Matsubayashi Y. 2015. Reevaluation of the CLV3-receptor interaction in the shoot apical meristem: dissection of the CLV3 signaling pathway from a direct ligand-binding point of view. *The Plant Journal* **82**(2): 328-336.

Somerville JE, Kahn ML. 1983. Cloning of the glutamine synthetase I gene from *Rhizobium meliloti*. *Journal of Bacteriology* **156**(1): 168-176.

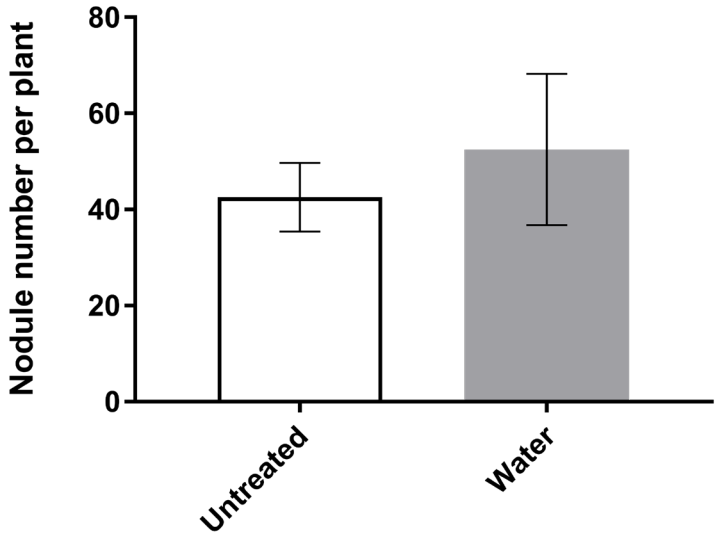
Tayeh N, Aluome C, Falque M, Jacquin F, Klein A, Chauveau A, Bérard A, Houtin H, Rond C, Kreplak J. 2015. Development of two major resources for pea genomics: the GenoPea 13.2 K SNP Array and a high-density, high-resolution consensus genetic map. *The Plant Journal* **84**(6): 1257-1273.

Whitford R, Fernandez A, De Groodt R, Ortega E, Hilson P. 2008. Plant CLE peptides from two distinct functional classes synergistically induce division of vascular cells. *Proceedings of the National Academy of Sciences* **105**(47): 18625-18630.

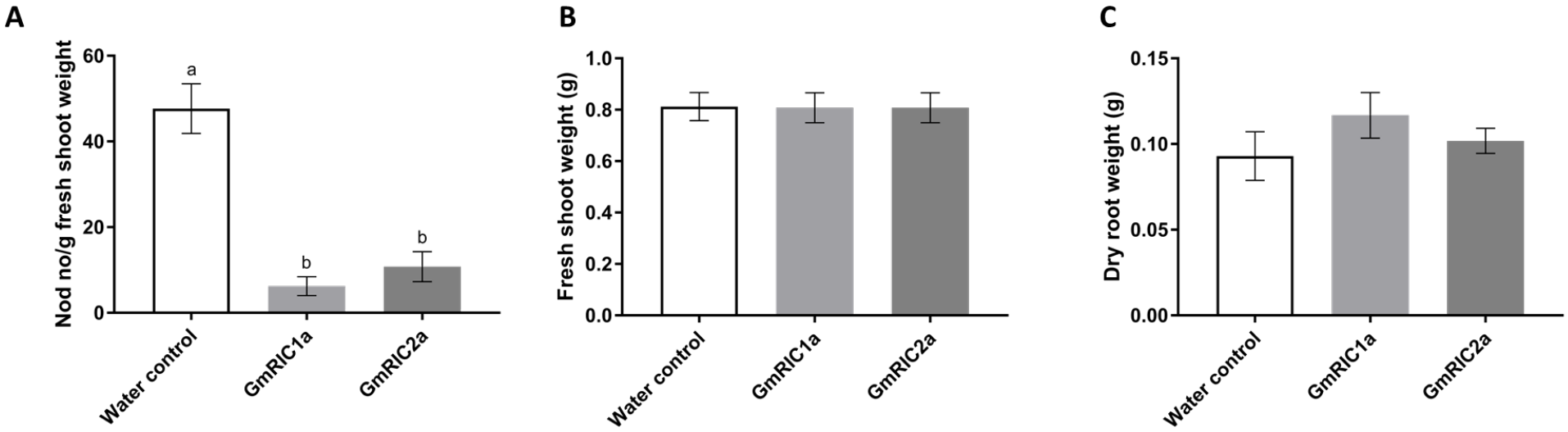
Xu C, Liberatore KL, MacAlister CA, Huang Z, Chu Y-H, Jiang K, Brooks C, Ogawa-Ohnishi M, Xiong G, Pauly M. 2015. A cascade of arabinosyltransferases controls shoot meristem size in tomato. *Nature Genetics* **47**(7): 784-792.

Yamaguchi YL, Ishida T, Sawa S. 2016. CLE peptides and their signaling pathways in plant development. *Journal of Experimental Botany* **67**, 4813-4826.

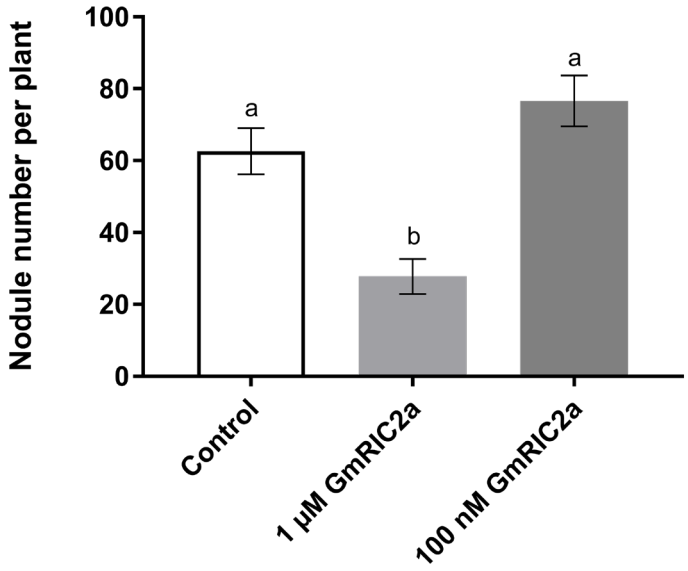
6.8 Supplementary material



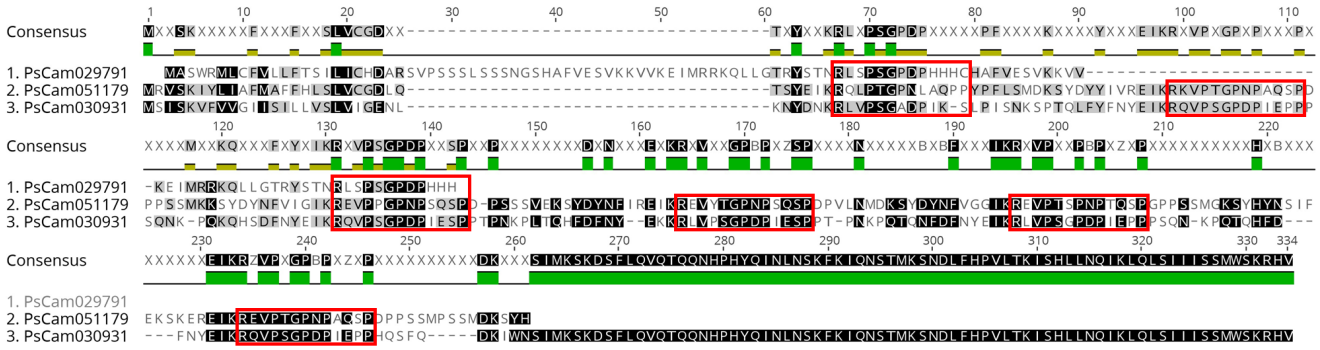
Supplementary Figure 1. Nodule number of plants 14 days after rhizobial inoculation with (water) and without (untreated) an attached petiole-feeding apparatus. No statistical differences were observed using the Student's *t*-test (P=0.52) n = 8 to 14 plants per treatment and error bars represent standard error of the mean.



Supplementary Figure 2. Phenotype 14 days after inoculation of wild-type pea plants fed via petiole feeding with 1 µM triarabinosylated GmRIC1a, GmRIC2a or water control. **A** Nodule number per g of fresh shoot weight. **B** Fresh shoot weight **C** Dried root weight. Statistical differences determined using Student's *t*-test. n = 8 to 10 plants per treatment and error bars represent standard error of the mean.



Supplementary Figure 3. Nodule number 14 days after inoculation of wild-type pea plants fed via petiole feeding with 1 µM or 100 nM triarabinosylated GmRIC2a compared with water control. Statistical differences determined using Student's *t*-test. n = 8 to 14 plants per treatment and error bars represent standard error of the mean.



Supplementary Figure 4. Multiple Sequence Alignment of the multi CLE domain containing CLE prepropeptides identified in *P. sativum*. The CLE domains of the prepropeptides are highlighted by a red box.

Supplementary Figure 5. Phylogenetic tree of the CLE prepropeptides of *P. sativum*, *M. truncatula*, *L. japonicus*, *A. thaliana* and *P. vulgaris,* including AtRGF1 as an outgroup. Bootstrap confidence values are expressed as a percentage from 100 bootstrap replications.

Supplementary Figure 6. Sequence Logo Diagrams of orthologous CLE domains from *G. max*, *P. vulgaris*, *L. japonicus*, *M. truncatula* and *P. sativum*. **A** Rhizobia-induced CLE peptides. **B** Nitrate-induced CLE peptide. LjCLE5, MtCLE34 and GmCLE32 are not included.

Supplementary Table 1. Features of the pea CLE prepropeptides

Identifier	Prepropeptide Length (AA)	CLE Domain(s)	Predicted Signal Peptide End
PsCam001939	89	HEVPSGPNPISNR	35
PsCam002283	95	RLVPSGPNPLHN	26
PsCam002955	80	RVPTGSNPLHNKK	24
PsCam003330	84	RVAPSGPDPHHH	25
PsCam005544	87	RIAPPSIPNPTQN	26
PsCam010169	77	RKVYTGPNPLHNR	26
PsCam010520	108	RVRKGSDPIHNRA	27
PsCam012685	67	RLSPSGPDPRHH	22
PsCam017409	144	RVSPGGPDGHHHF	21
PsCam021418	117	RLVPTGPNPLHN	20
PsCam028554	115	RILSEIPSGPDPI	30
PsCam029791	117	RLSPSGPDPHHHC	22
		RLSPSGPDPHHH	
PsCam030931	282	RLVPSGADPIKSL	22
		RQVPSGPDPIEPP	
		RQVPSGPDPIESP	
		RLVPSGPDPIESP	
		RLVPSGPDPIEPP	
		RQVPSGPDPIEPP	
PsCam039381	90	RVVPTGPNPLHNR	24
PsCam040153	91	RLSPAGPNHTHNN	29
PsCam040269	104	RLVPTGPNPLHH	29
PsCam040284	103	RVPNGPDPIHNRR	35
PsCam040575	82	RKVPSCPDPLH	26
PsCam040702	83	RLSPGGPDPQHNG	27
PsCam040727	84	RTVKKGSDPIHNR	21
PsCam040855	78	HVVPTGPNPLHN	25
PsCam040884	91	RQSPDSKSNPIH	24
PsCam041187	82	RVSPGGPDAQHH	26
PsCam041559	118	RQSPEGPSKRHNP	29
PsCam041591	106	REAPSGPNQLHNP	30
PsCam041632	84	RLSPQGPDPRHH	23
PsCam041636	79	RLPGRLSPEGPDP	26
PsCam041659	77	RLSPSGPDPHHH	24
PsCam043776	110	HEVPSGPNPISNR	29
PsCam043813	91	RKIYTGPNPLHNR	34
PsCam044605	101	RRVPTGPNPLHH	27
PsCam044761	73	RLSPGGPDGHHH	23
PsCam047237	83	KFIPPSRPNPTQN	25
PsCam047453	86	RTIPPSRPNPTQN	20
PsCam051179	225	RQLPTGPNLAQPP	22
		REIKRKVPTGPNP	
		REVPPGPNPSQSP	
		REVYTGPNPSQSP	
		REVPTSPNPTQSP	
		REVPTGPNPAQSP	
PsCam052272	106	RRVPSCPDPLHN	30
PsCam052818	106	RLVPSGPNPLHNL	31
PsCam052951	85	RILSEIPSGPDPI	np
PsCam056889	94	RLVPTGPNPLHNR	23
PsCam059704	148	RQTPSGPNKKHNA	18

np - No predicted signal peptide

Supplementary Table 2. Sequences with a CLE domain from PsCAmour Low Copy Number that are not included in analyses reported here.

Identifier	CLE Domain(s)	Notes
PsCam059992	RILSEIPSGPDPIHN	Four highly similar sequences to PsCam052951 and includes sequences with no clear start or stop codons
PsCam053544	RILSEIPSGPDPIH	
PsCam052913	RILSEIPSGPDPIHN	Multiple CLE domains
	RILSEIPSGPDPIHN	
PsCam046649	LHDVPGGPNPLHN	Possible intron containing sequence with no clear start codon
PsCam052264	RLIHTGPNPLHN	No predicted intron boundaries. Possible signal peptide in different frame
PsCam059958	RILKPSPSGPNKK	88.5% similar to PsCam059704 and no clear start or stop codons
PsCam059934	RILSEIPSGPDPI	Multiple CLE domains and no start codon
	RILSEIPSGPDPI	74.2% Simliar to PsCam028554 (which doesn't have multiple CLE domains)
PsCam052264	RLIHTGPNPLHN	Repeated sequence

Chapter 7

The soybean (*Glycine max*) nodulation-suppressive CLE peptide, GmRIC1, functions interspecifically in common white bean (*Phaseolus vulgaris*), but not in a supernodulating line mutated in the receptor PvNARK

Preface

This journal article has been published in Plant Biotechnology Journal (2014, doi:10.1111/pbi.12216). This article demonstrates that the nodulation-suppressive CLE peptides of soybean can function interspecifically in common bean. The endogenous nodulation-suppressive CLE peptides of common bean were also identified and genetically characterised using methods described by Hastwell et al. (2015b) in Chapter 3. Additional supplementary tables and figures may be found online at http://onlinelibrary.wiley.com/doi/10.1111/pbi.12216/abstract.

7.1 Summary

Legume plants regulate the number of nitrogen-fixing root nodules they form via a process called the Autoregulation of Nodulation (AON). Despite being one of the most economically important and abundantly consumed legumes, little is known about the AON pathway of common bean (*Phaseolus vulgaris*). We used comparative- and functional-genomic approaches to identify central components in the AON pathway of common bean. This includes identifying *PvNARK*, which encodes a LRR receptor kinase that acts to regulate root nodule numbers. A novel, truncated version of the gene was identified directly upstream of *PvNARK*, similar to *Medicago truncatula*, but not seen in *Lotus japonicus* or soybean. Two mutant alleles of *PvNARK* were identified that cause a classic shoot-controlled and nitrate-tolerant supernodulation phenotype. Homeologous over-expression of the nodulation-suppressive CLE peptide-encoding soybean gene, *GmRIC1*, abolished nodulation in wild-type bean, but had no discernible effect on *PvNARK*-mutant plants. This demonstrates that soybean GmRIC1 can function interspecifically in bean, acting in a PvNARK-dependent manner. Identification of bean *PvRIC1*, *PvRIC2* and *PvNIC1*, orthologues of the soybean nodulation-suppressive CLE peptides, revealed a high degree of conservation, particularly in the CLE domain. Overall, our work identified four new components of bean nodulation control and a truncated copy of *PvNARK*, discovered the mutation responsible for two supernodulating bean mutants and demonstrated that soybean GmRIC1 can function in the AON pathway of bean.

7.2 Introduction

Legumes account for a large source of biologically available nitrogen through a highly specialized symbiotic relationship with nitrogen-fixing soil bacteria termed rhizobia. The relationship is signified by the formation of novel organs on the legume root, called nodules (reviewed in Ferguson et al. 2010). Nodules act to house the rhizobia, providing them with the conditions necessary to generate useable forms of reduced nitrogen. This process is frequently exploited in agriculture, where legumes are used in rotation or alongside other crops to increase yields and improve the nitrogen content and structure of soils (Jensen et al. 2012).

Forming and maintaining nodules require a complex interaction of various plant hormones and signals (e.g. Ding et al. 2008; Ferguson and Mathesius 2003; Ferguson et al. 2005, 2011; Hirsch and Fang 1994; Ryu et al. 2012). This is costly to the host plant in terms of resources. As a result, legumes have developed mechanisms that enable them to optimize nodule formation (and hence nitrogen acquisition) under an array of growing conditions. This includes regulating their nodule numbers in response to environmental factors, such as the nitrogenous content of the soil (e.g. Carroll et al. 1985a,b; Reid et al. 2011a), soil acidity (Ferguson et al. 2013; Lin et al. 2012) and stress (e.g. ethylene) (Gresshoff et al. 2009; Guinel and Geil 2002).

Legumes also control their nodule numbers via an inbuilt signalling mechanism known as the Autoregulation of Nodulation (AON) (Delves et al. 1986; Kosslak and Bohlool 1984; Reid et al. 2011b). AON is triggered following the first nodulation events, and is predicted to begin with the production of a root-derived CLAVATA/ESR-related (CLE) peptide(s) (Lim et al. 2011; Mortier et al. 2010, 2012; Okamoto et al. 2009, 2013; Reid et al. 2011a,b, 2013). In soybean, two AON CLE peptide-encoding genes have been identified, called *Rhizobia-Induced CLE1* (*GmRIC1*) and *Rhizobia-Induced CLE2* (*GmRIC2*) (Reid et al. 2011a). *GmRIC1* expression is induced early following nodule initiation (within 12 h), whereas *GmRIC2* expression is triggered later (approximately 72 h) (Hayashi et al. 2012; Reid et al. 2011a, 2013).

GmRIC1 and GmRIC2 are predicted to be transported via the xylem to the shoot, where they are thought to be perceived by a specialized LRR receptor kinase (Krusell et al. 2002; Nishimura et al. 2002; Schnabel et al. 2005; Searle et al. 2003). In soybean, this receptor is called Nodulation Autoregulation Receptor Kinase (GmNARK; Searle et al. 2003). NARK

may function in a complex with other receptors, such as CLAVATA2 and KLAVIER (Krusell et al. 2011; Miyazawa et al. 2010). Additional work using soybean has identified other components that may also interact with GmNARK, or function downstream of it, including GmKAPP1 and GmKAPP2 (Miyahara et al. 2008) and GmUFD1a (Reid et al. 2012). Soybean plants having mutations in *GmNARK* lack AON control (Olsson et al. 1989) and exhibit a nitrate-tolerant supernodulation phenotype (Carroll et al. 1985a,b; Day et al. 1987).

Recent elegant work using *Lotus japonicus* has shown that the nodulation-inhibiting LjCLE-RS2 peptide is transported in the xylem as a post-translationally arabinosylated glycopeptide that directly binds to LjHAR1 (the orthologue of GmNARK) to suppress nodulation (Okamoto et al. 2013). Constitutive over-expression of either *GmRIC1* or *GmRIC2* in soybean hairy roots can completely abolish nodule formation, acting systemically through the shoot in a GmNARK-dependent manner (Reid et al. 2011a, 2013). The perception of GmRIC1 or GmRIC2 by GmNARK leads to the production of a novel Shoot-Derived Inhibitor (SDI) signal (Ferguson et al. 2010; Reid et al. 2011b). This signal is subsequently transported to the roots where it acts to inhibit further nodulation events (Lin et al. 2010a,b, 2011a).

In addition to exhibiting a supernodulation phenotype, *GmNARK* mutants of soybean exhibit nitrate-tolerant symbiosis (*nts*), where normally inhibitory levels of nitrate fail to reduce nodule numbers (Carroll et al. 1985a,b; Day et al. 1987). This provided the first clue that GmNARK could be involved in both AON and the nitrate regulation of nodulation. Recent work identified a CLE peptide-encoding gene that acts in nitrate regulation of nodulation and is highly similar to *GmRIC1* and *GmRIC2*, called *Nitrogen-Induced CLE1* (*GmNIC1*) (Reid et al. 2011a). Like *GmRIC1* and *GmRIC2*, *GmNIC1* is expressed in the root and acts in a GmNARK-dependent manner. However, unlike *GmRIC1* and *GmRIC2, GmNIC1*is induced specifically by available nitrogenous compounds, not rhizobia. It is also perceived locally by GmNARK in the root, not systemically in the shoot (Reid et al. 2011a). Over-expression of *GmNIC1* in soybean causes a significant reduction in nodule formation, but not to the extent of *GmRIC1* and *GmRIC2* (Reid et al. 2011a). Collectively, this functional characterization of GmNIC1 confirmed that GmNARK does have a role in both AON and nitrate regulation of nodulation.

All legumes appear to regulate their nodule numbers, suggesting AON is a well-conserved process. To investigate this hypothesis further, we tested the interspecific function of the

soybean (*Glycine max*) nodulation-inhibiting CLE peptide, *GmRIC1*, in the AON pathway of common white bean (*Phaseolus vulgaris*). Both species are similar in growth and stature and are amenable to *Agrobacterium rhizogenes*-mediated transformation (e.g. Estrada-Navarrete et al. 2007; Kereszt et al. 2007; Lin et al. 2011b). Moreover, they represent two of the world's most widely grown legume crops. However, unlike soybean where nodulation control has been well studied (Ferguson 2013), and despite its immense agricultural, nutritional and economic importance, very little is understood about how bean plants regulate their nodule numbers.

In the 1980s, a mutagenesis experiment using the common white bean variety, Ontario Agricultural College (OAC) Rico, was conducted at the Harrow Research Centre in Canada, resulting in the isolation of a supernodulating bean line called R32 (Park and Buttery 1988). Similar to previously described supernodulating mutants of soybean, the R32 mutant (Figure 7.1) is characterized by profuse and nitrate-tolerant nodulation (Park and Buttery 1988, 1989). The phenotype of the mutant is inherited as a single, recessive gene (Park and Buttery 1989) that predominately acts in the shoot to control nodule numbers of the root (Buttery and Park 1990, 1993; Hamaguchi et al. 1993). In the absence of compatible rhizobia, mutant R32 plants grow similarly to their OAC Rico parent, but they exhibit significantly reduced root and shoot systems when induced to form nodules (Becher et al. 1997; Buttery and Park 1990). Although the mutant forms significantly more nodules, both the average nodule size (Buttery and Park 1990; Hansen et al. 1993a) and the number of nodules that appear red (Hansen et al. 1992) are greatly reduced compared with the parent cultivar. Unsurprisingly, the symbiotic nitrogen fixation efficiency per nodule, as determined by both acetylene reduction and ^{15}N-labelled studies, is also significantly reduced in the mutant (Becher et al. 1997; Buttery and Park 1990; Buttery et al. 1990; Hansen et al. 1992, 1993a). Despite having nitrate-tolerant nodulation, other characteristics of R32 exhibit a normal response to nitrate (Buttery and Park 1990; Buttery et al. 1990; Hansen et al. 1992, 1993a). This includes symbiotic nitrogen fixation, which is detrimentally affected by nitrate in both R32 and its OAC Rico parent (Becher et al. 1997; Hansen et al. 1992, 1993a,b). Collectively, these R32 mutant traits strongly resemble those typified by classical supernodulating mutant lines isolated from other legume species that were later found to have a mutation in their orthologue of GmNARK (reviewed in Ferguson et al. 2010; Reid et al. 2011b).

A second supernodulating mutant of common white bean generated at the Harrow Research Centre in Canada was derived via mutagenesis of the Swan Valley variety and

named SV145 (Park and Buttery 1988). Though far less studied, SV145 is also characterized by profuse and nitrate-tolerant nodulation (Park and Buttery 1988, 1989). Complementation tests using genetic crosses demonstrated that the phenotypes of the R32 and SV145 mutants are under the control of the same gene (Park and Buttery 1989).

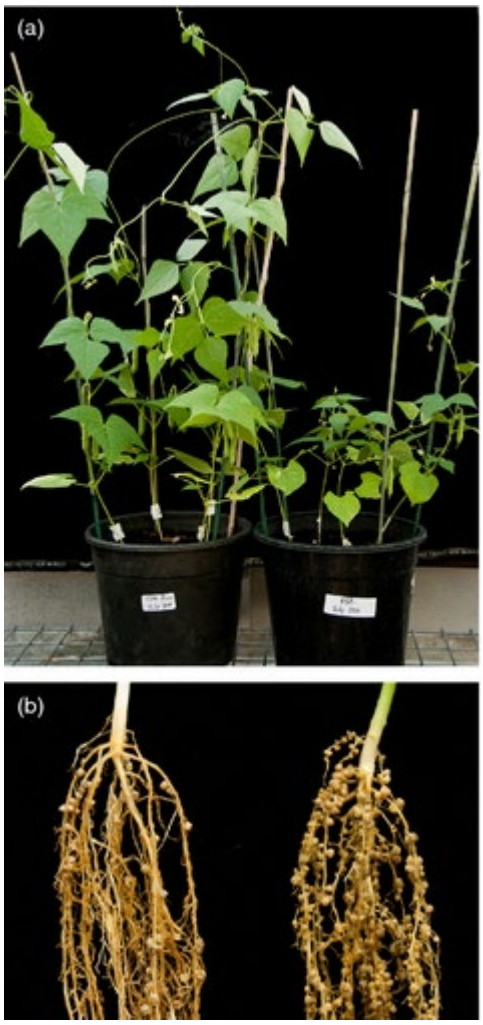


Figure 7.1 *Phaseolus vulgaris* wild type, OAC Rico and its supernodulation *PvNARK* mutant, *R32*. **(a)** Shoot and **(b)** root phenotypes of 2-month-old OAC Rico (left) and R32 (right) plants, inoculated with *Rhizobium phaseoli* strain CC511.

Here, we report on our use of comparative genomics with the *GmNARK* sequence of soybean to identify *PvNARK* in bean. A truncated copy of *PvNARK* was also found directly upstream of *PvNARK*. We subsequently identified distinct mutations in the *PvNARK* gene of R32 and SV145 that are responsible for the supernodulation phenotypes of these mutants. Heterologous over-expression of *GmRIC1*revealed that the soybean CLE peptide can negatively regulate nodule formation in wild-type bean, demonstrating that soybean *GmRIC1* is compatible with the AON pathway of bean. The inhibition was found to be PvNARK-dependent, as nodule numbers were not reduced in R32 plants over-expressing

GmRIC1. Comparative genomic approaches were also used to identify candidate orthologues of *GmRIC1, GmRIC2* and *GmNIC1* in bean. Collectively, our work identified four new components acting in bean nodulation control plus a truncated copy of *PvNARK*, discovered mutations that cause supernodulating in two separate bean lines, and established that soybean GmRIC1 can function in the AON pathway of bean.

7.3 Results

7.3.1 Identification of the *NARK* orthologue in common white bean

The sequence of *PvNARK* was determined via comparative genomic approaches using bean and soybean, followed by sequencing the full-length genomic and CDS DNA sequences of bean. The gene is located on chromosome 11 of bean. It is 3367 bp in length, including a single intron of 400 bp beginning at base pair 2600 (Figure 7.2a). These structural features are highly similar to that of orthologues genes reported in *Lotus japonicus*, *Glycine max* and *Medicago truncatula* (Krusell et al. 2002; Nishimura et al. 2002; Schnabel et al. 2005; Searle et al. 2003). The *PvNARK* sequences of the wild-type cultivars OAC Rico and Swan Valley were found to be 100% identical, including that of the intron.

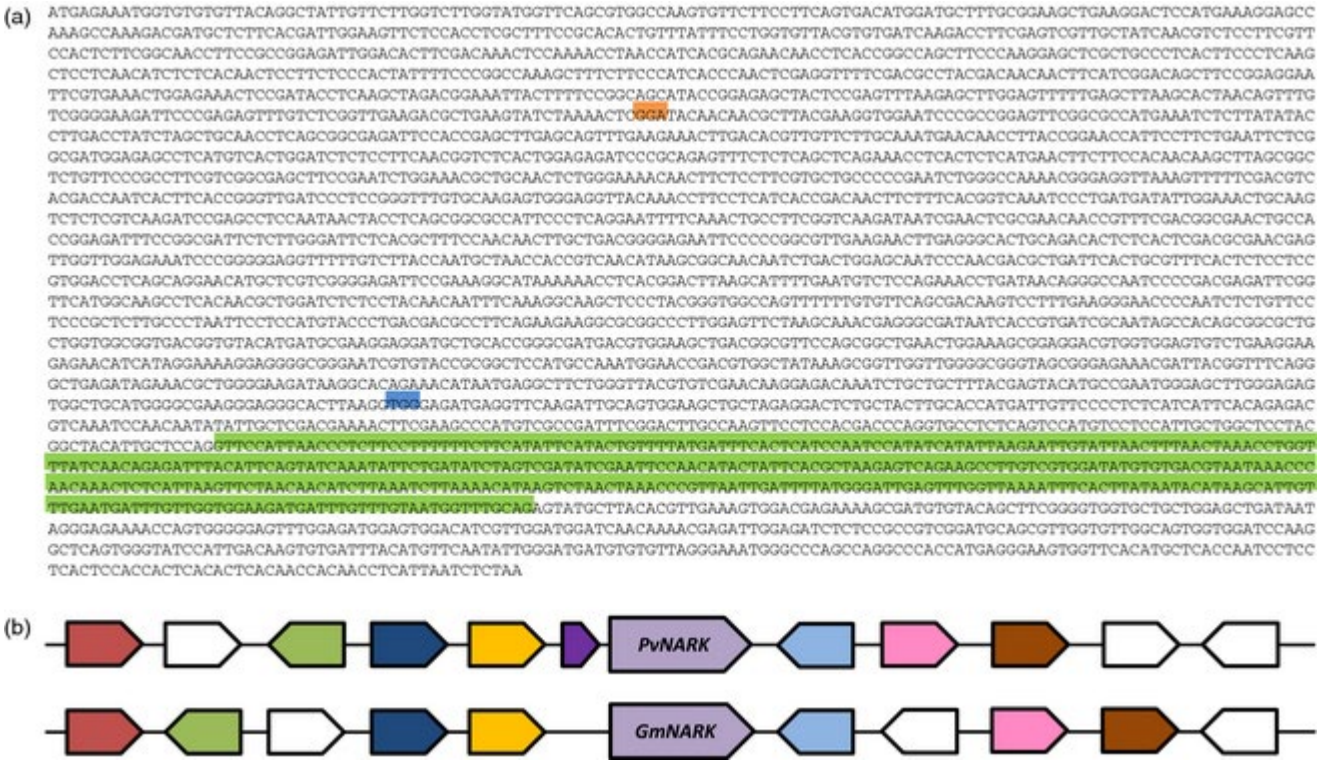


Figure 7.2 The *Nodulation Autoregulation Receptor Kinase* gene of *Phaseolus vulgaris*. **(a)** Genomic DNA sequence of *PvNARK*. An intron of 400 bp is highlighted in green. The GGA codon highlighted in

orange is mutated to GAA in the supernodulating bean mutant, R32, resulting in a glycine residue being changed to glutamic acid. The TGG codon highlighted in blue is mutated to TGA in the supernodulating bean mutant, SV145, resulting in a pre-mature stop codon. **(b)** Phytozome cluster analysis of the genomic environments of *PvNARK* and *GmNARK* reveals a well-conserved region of synteny existing between the two species. A truncated version of *NARK*(dark purple) is located directly upstream of *PvNARK*, but not *GmNARK*.

To unequivocally verify whether *PvNARK* and *GmNARK* are orthologous genes, their genomic environments were compared. A number of neighbouring genes were found to be well conserved within the genomic environments of the two species, both in terms of their predicted sequence and their orientation within the genome (Figure 7.2b). This level of genetic synteny validates that the *PvNARK*sequence identified here is indeed orthologous to that of the previously reported *GmNARK*. In addition, the ~2-kb promoter region located directly upstream of *PvNARK* (Figure S1) revealed a number of shared motifs previously identified in the *GmNARK* promoter (Nontachaiyapoom et al. 2007).

7.3.2 The amino acid sequence of PvNARK is highly conserved with its orthologues in other legume species

To further characterize NARK in bean, the protein sequence of the receptor kinase, including its predicted domains, was determined. The complete PvNARK sequence is 989 amino acids in length (Figure 7.3a). Four main domains were identified, including a leucine-rich repeat-containing N-terminal signal peptide, a leucine-rich repeat domain, a transmembrane domain and a serine/threonine kinase domain (Figure 7.3a). These domains, and their arrangement within the PvNARK protein, are highly conserved with NARK orthologues reported in other legume species (Krusell et al. 2002; Nishimura et al. 2002; Schnabel et al. 2005; Searle et al. 2003). Indeed, *PvNARK* and *GmNARK*, which are 85.5% similar in nucleic acid sequence, encode proteins that are 85.1% similar in amino acid sequence.

A phylogenetic tree was constructed to further compare the orthologous sequences of NARK in bean, soybean, *Lotus*, *Medicago*, and pea (Figure 7.3b). Predictably, all five proteins are very highly conserved in sequence. PvNARK is most similar to GmNARK, consistent with the two species being closely related and bean having only diverged from soybean ~20 million years ago (Lavin et al. 2005; Lin et al. 2010a,b). Also of interest is that the NARK sequences of determinate nodule-forming species are more closely related

than those of the indeterminate nodule-forming species, which again is most likely a direct reflection of evolutionary divergence.

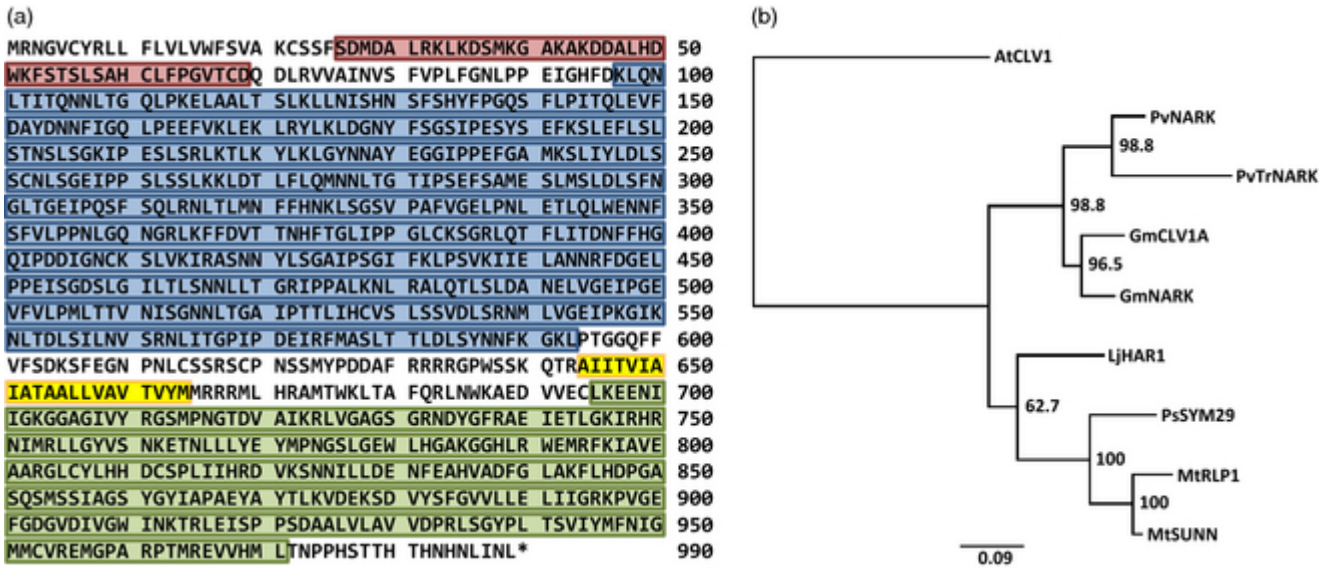


Figure 7.3 Nodulation Autoregulation Receptor Kinase of *Phaseolus vulgaris*. **(a)** PvNARK is comprised of four domains: leucine-rich repeat-containing N-terminal signal peptide (red), leucine-rich repeats (blue), transmembrane (yellow) and serine/threonine kinase (green). **(b)** Phylogenetic tree of PvNARK and its orthologues in soybean (GmNARK), *Lotus japonicus* (LjHAR1), *Medicago truncatula* (MtSUNN) and pea (PsSYM29), in addition to its orthologue in *Arabidopsis thaliana* (AtCLV1) as the outgroup. Also included are the truncated copies of PvNARK in bean (PvTrNARK) and MtSUNN in *Medicago truncatula* (MtRLP1) and the homeologous copy of GmNARK (formerly GmCLV1B) in soybean (GmCLV1A). The tree is shown with bootstrap confidence values expressed as a percentage from 1000 bootstrap replications.

An alignment of the amino acid sequences of the NARK orthologues of bean, soybean, *Lotus*,*Medicago*, and pea further demonstrated their high level of conservation (Figure S2). This is particularly true in regions of the protein that are predicted to represent a distinct domain (Figure S2), which comprises most of the total NARK sequence (Figure 7.3a). This degree of conservation indicates that most of the NARK protein is required for proper function and that only minimal sequence changes can be tolerated without disrupting optimum functionality.

7.3.3 Identification of a truncated copy of *PvNARK*

Interestingly, a truncated version of *PvNARK* was identified just 4.9 kb upstream of the parent *PvNARK* gene (Figure 7.2b). This truncated gene is 1962 bp in length and is designated here as *PvTruncatedNARK* (*PvTrNARK*). The protein sequence and predicted

domains of PvTrNARK were also determined. It is 653 amino acids in length (Figure S3), with two complete domains predicted. This includes the leucine-rich repeat-containing N-terminal signal peptide and the leucine-rich repeat domain. The protein terminates in what would be the middle of the transmembrane domain of PvNARK. Alignment of PvTrNARK with PvNARK shows that the two genes are highly conserved (Figure S3), particularly in regions predicted to encode functional domains, with 80% similarity in their amino acid sequences.

A similar truncated version of *MtSUNN*, the orthologue of *PvNARK* in *M. truncatula*, has also been reported. It is positioned 6.2 kb upstream of *MtSUNN* and is designated as *MtRLP1* (Schnabel et al. 2005). No such truncated gene has been reported upstream of the *PvNARK* orthologues in soybean (*GmNARK*) or *L. japonicus* (*LjHAR1*). However, a truncated version of *GmCLV1A*, the homeologous copy (i.e. duplicate) of *GmNARK*, has been identified 4.7 kb upstream of *GmCLV1A* (S. Mirzaei, J. Batley, K. Meksem, B.J. Ferguson and P.M. Gresshoff, unpublished). This may be indicative of either an evolutionary deletion in some species or an independent origin of the truncation. In all cases, the function of the truncated copy of the gene, if any, remains unknown.

7.3.4 The supernodulating mutant lines of common white bean, R32 and SV145, have a mutation in *PvNARK*

The R32 and SV145 mutants of common white bean exhibit supernodulation phenotypes that are highly reminiscent to those exemplified by NARK mutants of other legume species (Figure 1). Thus, the sequence of *PvNARK* was determined in R32 and SV145 to establish whether a mutation in this gene was responsible for the phenotype of the mutants.

R32 was found to have a point mutation at base pair 674 of the *PvNARK* coding sequence, represented by a transition from G to A (highlighted in orange in Figure 7.2a). This alters the codon from GGA to GAA, which causes a change in amino acid 225 from glycine to glutamic acid. Such a change can be quite influential to protein function as glycine is small (MW = 57.05) and uncharged, whereas glutamic acid is much larger (MW = 129.12) and acidic.

SV145 was found to have a point mutation located at base pair 2373 of the *PvNARK* coding sequence. This mutation is also represented by a transition from G to A (highlighted in blue in Figure 7.2a), which alters a TGG codon for tryptophan at amino acid

791 into a pre-mature stop codon (TGA). The mutation is located in the predicted kinase domain, and results in a truncated PvNARK protein. Such a mutation would be expected to disrupt the protein's function, similar to a mutation in the *nts382*supernodulating mutant of soybean, which has a pre-mature stop codon in its kinase domain (Searle et al. 2003) located downstream of the one identified here in SV145.

Like soybean, stable transformation of bean is not readily achievable and hence complementation of NARK in R32 and SV145 mutant shoots is not an option. However, the identification of distinct mutations in their *PvNARK* genes coupled with the fact that their phenotypes closely resemble that of verified NARK mutant lines isolated from other legume species (excessive nodulation that is nitrate tolerant, shoot controlled, inherited as a single, recessive gene, etc.), and the fact that their phenotype can be rescued when crossed into wild-type bean but cannot be rescued in complementation tests with one another, conclusively demonstrates that a mutation in *PvNARK* must be responsible for the supernodulation phenotypes of the R32 and SV145 mutants.

7.3.5 The *PvNARK* mis-sense mutation of R32 is located in the putative ligand binding site of the LRR domain

Previous reports have suggested that the *NARK* gene is a 'hotspot' for mutation, as two mutants have been identified in bean (Park and Buttery 1989), in addition to fifteen in the much more thoroughly investigated soybean (Delves et al. 1988), with a relatively high mutation frequency reported for both species. This provides the nodulation and LRR-RLK communities with a large collection of mutant lines to investigate structure-function relationships. To date, all reported mutations in bean and soybean are located in various regions of the leucine-rich repeat (LRR) or kinase domains, and/or cause a pre-mature stop codon resulting in a truncation of the protein (Figure 7.4a).

The R32 mutation affects amino acid 225, which is located within the LRR domain of PvNARK. To better delineate the precise location in the 3D structure of the LRR domain, and to establish how this location might affect the function of PvNARK, molecular modelling of the LRR domain was performed. The LRR domain was predicted to form a horseshoe shape (Figure 7.4b), typical of many other reported LRRs, including that of the orthologous sequence in soybean, GmNARK (Reid et al. 2011a). Amino acid 225 of PvNARK is predicted to be located on the inside of the LRR, directly in the central binding cleft of the domain (Figure 7.4b). This location represents the proposed CLE-ligand

binding site of the receptor (Reid et al. 2011a). A mutation in this region could have a significant impact on the function of the LRR domain's ability to recognize and bind to its ligand partner, potentially even rendering it completely ineffective. When coupled with the fact that the mutation results in a major amino acid shift, from glycine to glutamic acid, it is not surprising why the R32 mutant exhibits such a strong supernodulating phenotype.

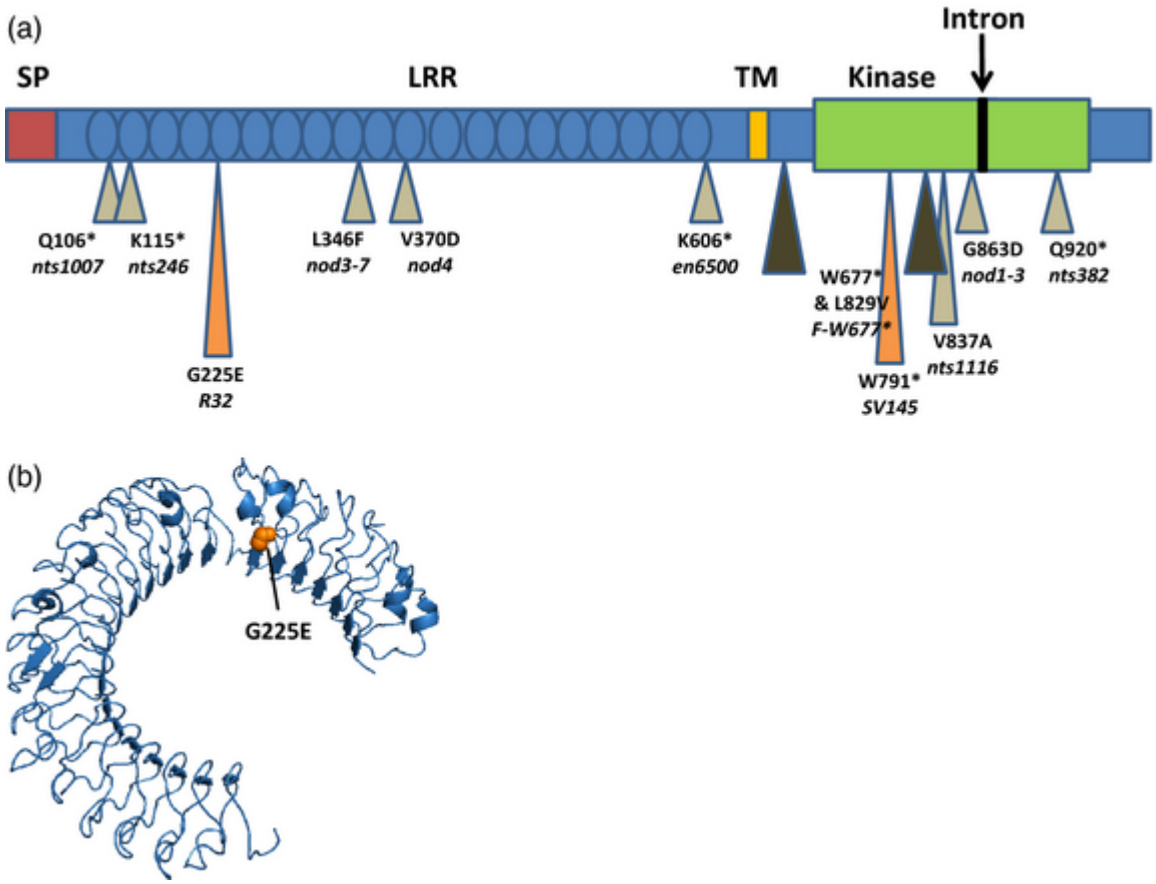


Figure 7.4 Known NARK mutations in soybean (*Glycine max*) and bean (*Phaseolus vulgaris*). **(a)** Illustration of the NARK protein structure depicting known mutations, including the location, amino acid change and mutant name. Mutations are shown for bean (orange arrows) and various lines of soybean (grey arrows), including the *F-W677** mutant (olive green arrows), which was recently isolated from a TILLING population (J. Batley, S. Liu, S. Mirzaei, T. El-Mellouki, B.J. Ferguson, P.M. Gresshoff, P.M. and K. Meksem, unpublished) and contains two separate mutations. Not shown is the *FN37* mutant of *Glycine soja* (wild soybean), which was generated using fast neutron technology and is completely devoid of the *NARK* gene (Men et al. 2002; Searle et al. 2003). **(b)** Phyre modelling (Kelley and Sternberg 2009) (based on PDB: 1ZIW) of the NARK leucine-rich repeat domain of bean. The site of the mutated amino acid residue (G225E; highlighted in orange) of the supernodulation mutant, R32, is depicted flanking the central cleft of the LRR domain that likely defines the ligand binding site.

7.3.6 The soybean CLE peptide-encoding gene, *GmRIC1*, functions interspecifically in the AON pathway of bean in a PvNARK-dependent manner

The soybean gene, *GmRIC1*, encodes a hormone-like CLE peptide signal that has previously been shown to completely inhibit nodulation when transgenically over-expressed in soybean plants (Reid et al. 2011a). To identify whether this soybean gene can function interspecifically in the AON pathway of bean, *GmRIC1* was heterologously over-expressed in *Agrobacterium rhizogenes*–mediated hairy roots of common white bean plants. Over-expression of *GmRIC1* significantly reduced the nodule number of wild-type bean plants, compared with empty-vector control plants, on both a per-plant (P = 6.13 × 10^{-7}) (Figure 7.5) and a per-gram root dry weight (P = 2.09 × 10^{-5}) bases. This near abolishment of nodulation is similar to what was found using soybean (Reid et al. 2011a) and demonstrates that *GmRIC1* can function in the AON pathway of common white bean. This indicates that this part of the AON pathway of the two species is highly conserved.

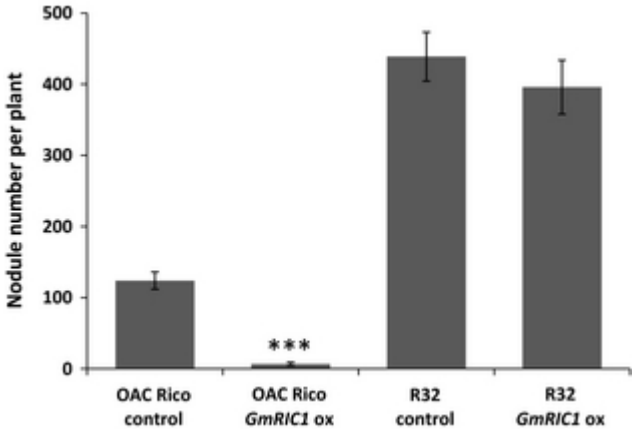


Figure 7.5 Hairy-root over-expression of the soybean AON CLE peptide, *GmRIC1*, in bean and its effect on nodule formation. Nodulation of OAC Rico (wild type) hairy roots over-expressing *GmRIC1* was significantly inhibited compared with those induced with the vector-only control (***P < 0.001). In contrast, nodule numbers were not reduced in *R32* mutant plants. Error bars indicate mean ± standard error (SE) (n = 10–12).

To determine whether *GmRIC1* suppression of nodule development in bean functions via PvNARK, *GmRIC1* was heterologously over-expressed in *Agrobacterium rhizogenes*–mediated hairy roots of mutant R32 plants. No significant difference in nodule number was detected between empty-vector control and *GmRIC1* over-expressing R32 plants (P = 0.41) (Figure 7.5), including when root weight was taken into account (P = 0.81). This clearly indicates that *GmRIC1* expression functions via PvNARK in the AON pathway of bean to control nodule numbers. When coupled with previous studies using grafting to

demonstrate that the R32 mutation functions in the shoot (Buttery and Park 1990, 1993; Hamaguchi et al. 1993), our findings demonstrate that GmRIC1, or a product of its action, must be transported to the shoot to interact with PvNARK. This fits exceedingly well with the current model of the AON process (Ferguson 2013; Ferguson et al. 2010; Hayashi et al. 2013; Reid et al. 2011b) and is consistent with the recent finding that the nodulation-suppressive CLE peptide, LjCLE-RS2 of *L. japonicus*, is transported to the shoot and directly binds to LjHAR1 (Okamoto et al. 2013).

7.3.7 Nodulation-suppressive CLE peptide orthologues of bean and soybean are highly conserved

Following on from our establishment that soybean *GmRIC1* functions in the AON pathway of bean, we set out to identify the bean orthologues of the three known soybean nodulation-suppressive CLE peptide-encoding genes: *GmRIC1, GmRIC2* and *GmNIC1* (Reid et al. 2011a). Comparative genomic approaches were used to BLAST the sequences of the soybean genes against the available scaffolds of the bean genome. As was carried out for *PvNARK*, candidates were further verified by assessing their genomic environments for genetic synteny. This was achieved via cluster analyses using *PvRIC1, PvRIC2*and *PvNIC1*, and both homeologous copies of *GmRIC1, GmRIC2* and *GmNIC1*. Approximately 75% of soybean genes have a homeologous partner due to duplication events occurring ~59 and 13 million years ago (Schmutz et al. 2010). Because bean diverged from soybean ~20 million years ago, it only has the older duplication event in common with soybean, and hence most bean genes exist as a single copy (Lavin et al. 2005; Lin et al. 2010a,b). In all cases, the ‘b’ copy of the soybean gene represents the presumed non-functional duplicate, which was previously found to exhibit little-to-no expression (Reid et al. 2011a).

PvNIC1 and *PvRIC1* are located adjacent to one another, 18.4 Kb apart, on bean chromosome 5. This is similar to *GmNIC1b* and *GmRIC1*, which are located 11.6 Kb apart on soybean chromosome 12, and also to *GmNIC1* and *GmRIC1b*, which are located 9.3 Kb apart on soybean chromosome 13 (Figure 7.6a; Reid et al. 2011a). An additional gene is conserved upstream of all three gene pairs, illustrating synteny within their genomic environments (Figure 7.6a). *PvRIC2* is located on bean chromosome 11 and exhibits genetic synteny with the genomic environment of *GmRIC2*, which is located on chromosome 6 of the soybean genome (Figure 7.6a; Reid et al. 2011a).

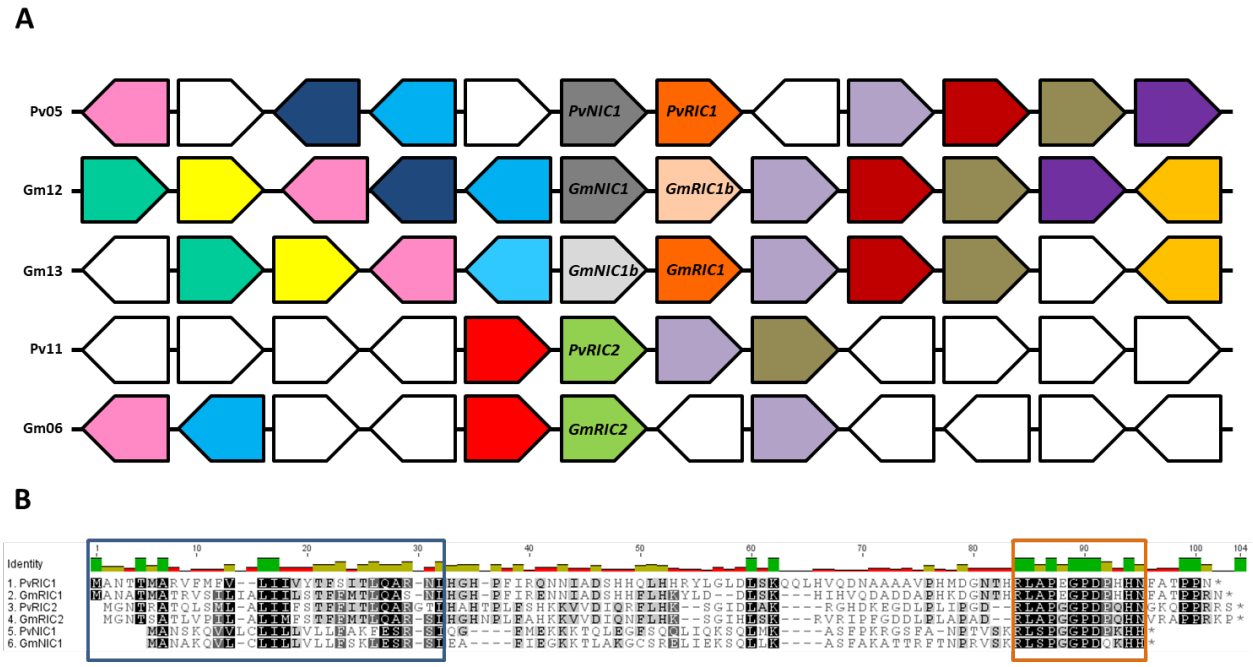


Figure 7.6 Microsynteny of inoculation- (*RIC1* and *2*) and nitrate- (*NIC1*) responsive CLE peptide-encoding genes of bean and soybean and a multiple sequence alignment of their predicted products. **(a)** Phytozome cluster analyses reveal well-conserved regions of synteny existing between the genomic environments of both species. *PvNIC1* is located directly upstream of *PvRIC1*, consistent with the duplicated, orthologous region of soybean. Both the functional and homeologous (*b*, duplicate, predicted non-functional) copy of the soybean genes are shown. **(b)** Predicted amino acid sequences illustrating the highly conserved signal peptide domains (outlined in a blue box) and the 12 amino acid CLE domains (enclosed by an orange box).

The nucleic acid sequences of *PvRIC1, PvRIC2* and *PvNIC1* are 79.3%, 82.1% and 80.5% similar to that of *GmRIC1, GmRIC2* and *GmNIC1*, respectively, and the products they encode are 69.3%, 70.5% and 69.5% similar in amino acid sequence, respectively. The predicted amino acid sequences of the three bean genes and the three functional soybean genes show that the signal peptide domain is fairly well conserved (Figure 7.6b). Even more conserved are the CLE peptide domains and two amino acid residues located in the variable domain, which may have an important role in the post-translational processing of the peptide (Figure 7.6b). Also consistent is the presence of a conserved C-terminal domain containing two proline residues in the RIC1/2 peptides, which is noticeably absent in the NIC1 peptides of the two species (Figure 7.6b). This domain may have a role in protecting against proteolytic degradation during long-distance transport, something that would not affect the locally acting NIC1 peptide. Of interest is the proline located at position 9 in the CLE peptide domain. It is found in all of the sequences examined except

for GmNIC1, which instead has a glutamine residue (Figure 7.6b). The exact function of this proline remains unknown, but it is possible that the weaker activity exhibited when over-expressing *GmNIC1*, compared with that of *GmRIC1* or *GmRIC2* (Reid et al. 2011a), may be associated with the residue change reducing GmNIC1 activity.

7.3.8 Nodulation-suppressive CLE peptides of bean are induced following the induction of nodule organogenesis

Wild-type OAC Rico plants were treated with either water or the compatible *Rhizobium phaseoli* strain CC511 and the transcript abundance of *PvRIC1* and *PvRIC2* was determined in their root systems 1 and 5 days later. Relatively little expression was observed in the water-treated root systems. In *R. phaseoli* inoculated plants, *PvRIC1* was found to be significantly up-regulated in expression after 1 day, followed by a decline in expression by day 5 (Figure 7.7). On the contrary, *PvRIC2* was not induced at day 1, but was significantly up-regulated in expression 5 days after rhizobia inoculation (Figure 7.7). These findings demonstrate that both genes are induced by inoculation with compatible *R. phaseoli*, and hence the triggering of nodule organogenesis, but that their pattern of expression differs from one another.

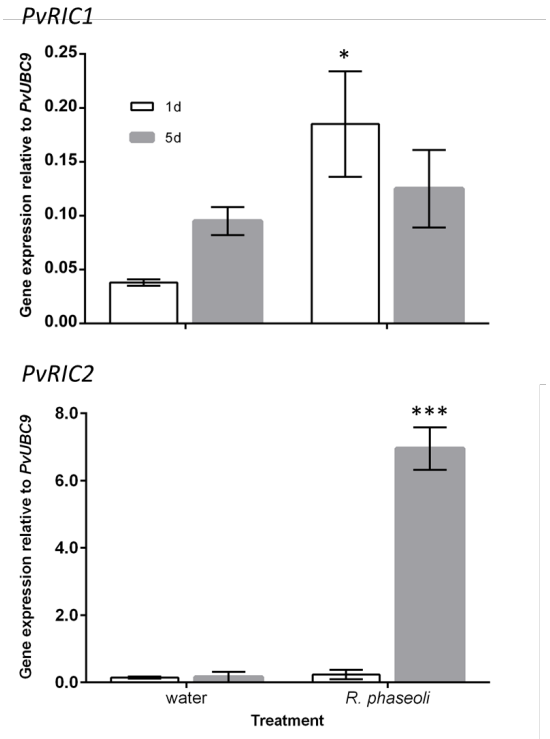


Figure 7.7 Expression of the nodulation CLE peptide-encoding genes, *PvRIC1* and *PvRIC2*, following the induction of nodule organogenesis. Values show the relative transcript abundance of *PvRIC1* and *PvRIC2* at 1 and 5 days following the exogenous application of either water (control) or a compatible

strain of *Rhizobium phaseoli*. Expression levels were determined using qRT-PCR. *n* = 9 per biological replicate. Error bars indicate SE. Asterisks above the bars represent statistically significant differences between treatments (Student's *t* test, where *$P \leq 0.05$ and ***$P \leq 0.001$).

7.4 Discussion

Although bean is one of the most widely grown legume crops, and of considerable agricultural, nutritional and economic importance, molecular components required to regulate nodulation in bean have not been reported to date. Here, we capitalized on the recent sequencing of the bean and soybean genomes, and used comparative and functional genomics to identify four key components acting in the regulation of bean nodulation, *PvNARK, PvRIC1, PvRIC2* and *PvNIC1*, as well as a truncated copy of *PvNARK*, called *PvTrNARK*. In addition to enhancing the molecular knowledge base of bean nodulation, the identification of these components now enables functional characterization studies that could aid future attempts to optimize bean nodulation in agriculture, and could benefit the targeting and selection of superior bean varieties. Moreover, comparisons with orthologous components in other legume species will provide new insight into evolutionary legume genomics.

The genomic DNA sequence and surrounding environment of *PvNARK* are highly similar to that of its orthologues in other legume species, including the position of a single intron (Krusell et al. 2002; Nishimura et al. 2002; Schnabel et al. 2005; Searle et al. 2003). Accordingly, the corresponding PvNARK protein sequence is also highly similar to its orthologues, particularly in regard to the positioning, length and amino acid residues of its four identified domains. This extent of conservation indicates that much of the sequence is essential to maintain function. Moreover, the high frequency of NARK mutants isolated from various mutagenesis populations (Delves et al. 1988; Park and Buttery 1989) further supports the notion that a high level of conservation is essential to preserve function.

The identification of a truncated copy of *PvNARK* located directly upstream of the parent gene is similar to that reported for the *NARK* orthologue in *M. truncatula* (Schnabel et al. 2005), but not in *L. japonicus* or soybean. However, a truncated version of soybean *GmCLAVATA1A*, the homeologue of *GmNARK*, has been identified (S. Mirzaei, J. Batley, K. Meksem, B.J. Ferguson and P.M. Gresshoff, unpublished). The fact that the truncated copies have two evolutionary conserved domains, and that both *MtRLP1* (Schnabel et al. 2005) and *GmTrCLV1A* (S. Mirzaei, J. Batley, K. Meksem, B.J. Ferguson and P.M.

Gresshoff, unpublished) are expressed, suggests that they could serve a biological function. This may occur as part of a heterodimer complex with a protein partner having a functional kinase domain capable of relaying a signal response, similar to the role of *AtCLV2* in the CLAVATA signalling pathway of Arabidopsis (Jeong et al. 1999). Potential interactions between this truncated component and other factors, such as a G-protein subunit, or factors proposed to bind with orthologues of NARK, such as Klavier (Miyazawa et al. 2010) and CLAVATA2 (Krusell et al. 2011), are also of great interest to resolve.

Our identification of distinct point mutations in *PvNARK* of R32 and SV145 demonstrates that a mutation in this gene is indeed responsible for the supernodulation phenotype of these mutants. The R32 mutation results in a transition from glycine to glutamic acid, and is located in the predicted central binding cleft of the leucine-rich repeat domain. This location, and change in amino acid residue, could have a considerable impact on the receptor's ability to effectively bind putative CLE peptide ligands, potentially even rendering it completely ineffective. Two supernodulation mutants of soybean, *nod3-7*and *nod4*, have also been found to have a mutation in this region (Reid et al. 2011a). The SV145 mutation results in a pre-mature stop codon located in the kinase domain of PvNARK. This mutation results in a truncated PvNARK protein that appears to be non-functional. A similar truncation located further downstream in the kinase domain has been reported in the supernodulating mutant of soybean, *nts382* (Searle et al. 2003).

Consistent with the phenotype of NARK mutants isolated from other legume species, both AON and nitrate-tolerant nodulation are disrupted in the R32 and SV145 mutants, demonstrating that NARK is an essential component of these processes. Likewise, the supernodulation phenotype of R32 is shoot-controlled (Buttery and Park 1990, 1993; Hamaguchi et al. 1993) and inherited as a single, recessive gene (Park and Buttery 1989) that is epistatic to other genes acting in nodule development (Park and Buttery 1994, 1997). Moreover, R32 exhibits significantly reduced root and shoot systems when induced to form nodules (Becher et al. 1997; Buttery and Park 1990), and it forms smaller nodules (Buttery and Park, 1990; Hansen et al. 1993a) with reduced symbiotic nitrogen fixation efficiency (Becher et al. 1997; Buttery and Park 1990; Buttery et al. 1990; Hansen et al. 1992, 1993a). These phenotypes are all conserved amongst NARK mutants of various legume species (Krusell et al. 2002; Nishimura et al. 2002; Schnabel et al. 2005; Searle et al. 2003).

Identifying the *PvNARK* mutation in R32 enabled us to test the interspecific function of the soybean nodulation-suppressive CLE peptide-encoding gene, *GmRIC1*, in the AON pathway of bean. Heterologous over-expression of *GmRIC1* almost completely suppressed nodulation in wild-type bean, but not in R32 *PvNARK*-mutant plants. This demonstrates that soybean *GmRIC1* can function in the AON pathway of bean to control nodule numbers, and that this occurs systemically through PvNARK. Similar findings were reported using soybean (Reid et al. 2011a), signifying that the AON pathways of soybean and bean, and perhaps all legumes, are analogous. Indeed, previous grafting studies using various legume species suggest this conservation could cross an even further evolutionary distance (e.g. Harper et al. 1997; Lohar and VandenBosch 2005; Sheng and Harper, 1997) and it was recently shown that *MtCLE13* of *M. truncatula* can function in pea (Osipova et al. 2012).

A previous study investigating the AON pathways of different legumes used various grafting techniques with supernodulating *NARK* mutants of soybean (En6500) and bean (R32) (Hamaguchi et al. 1993). However, reciprocally grafted plants failed to nodulate and instead the authors focused on Y and X grafts harbouring two shoots, one of which belonged to the same species as the rootstock. Failure of the reciprocally grafted plants to nodulate may be indicative of issues surrounding the re-establishment of vascular connections, which would have influenced outcomes, resulting in some contradictory interpretations with the work reported here.

The CLE peptide-encoding genes of bean, *PvRIC1* and *PvRIC2*, were shown here to be induced by a compatible strain *R. phaseoli*, indicating that they are triggered during nodule organogenesis. Interestingly, the two genes exhibit a different pattern of expression. *PvRIC1* was induced within 1 day of inoculation, and then declined in expression thereafter, whereas the induction of *PvRIC2* was not evident until 5 days after inoculation. These expression patterns are consistent with that reported for their respective orthologues in soybean (Reid et al. 2011a) and may indicate that RIC1 is an early signal of AON, representing the degree of rhizobia infection, whereas RIC2 is a later signal of AON, used to signify the extent of successful nodule maturation. The combination of these two signals may aid the plant in optimizing nodulation control.

CLE peptides generally have three domains: an N-terminal signal transit peptide, a variable region and a 12–14 amino acid CLE domain that, after processing and modification, is thought to represent the final signal ligand (Ni and Clark 2006). Many CLE

peptides also possess a C-extension domain present after the CLE domain at the C-terminal region (Oelkers et al. 2008). Future studies aimed at better understanding the role of these domains will be highly informative, building on recent domain-swap investigations using *GmRIC1* and *GmNIC1* of soybean (Reid et al. 2013).

The signal peptide domain is likely critical for exporting the peptide from its originating cell and/or out of the endoplasmic reticulum. While there is less conservation outside the CLE domain, some residues within the signal peptide are highly conserved and may represent key regions for transport, post-translational modifications and/or cleavage. The function of the C-extension domain also remains unclear, although it is proposed to provide the peptide with protection from proteolytic cleavage by a predicted carboxy-peptidase (Ni et al. 2011; Reid et al. 2011a, 2013). In soybean and bean, the long-distance transported RIC1 and RIC2 feature the domain, while the locally acting NIC1 does not. It is possible that such protection is only required while in long-distance transport; hence, the reason for its absence in the NIC1. Some individual amino acids within the variable region of the nodulation CLE peptides are also strongly conserved and may represent key residues for proteolytic cleavage of the prepropeptide. Also of interest is the proline residue located at position 9 within the CLE domain. This residue is conserved amongst all of the nodulation CLE peptides of bean and soybean, except for GmNIC1, which has a glutamine residue at that location. Establishing whether this amino acid change is responsible for the weaker ability of GmNIC1 to suppress nodule development compared with that of GmRIC1 and GmRIC2 (Reid et al. 2011a, 2013) is of interest to determine.

Some plant pathways can be influenced by more than one CLE peptide; that is, redundancy of function exists. This is clearly evident in the AON pathway, which can be driven by either RIC1 or RIC2. Moreover, due to their high level of amino acid conservation, it is predicted that the CLE domain of many CLE peptides can function in more than one signalling pathway. For example, the RIC1, RIC2 and NIC1 CLE peptides of bean and soybean all have very well-conserved residues in their CLE domains, and are all predicted to be perceived by the NARK receptor; however, RIC1 and RIC2 function systemically in AON, whereas NIC1 functions locally in the nitrate regulation of nodulation pathway. As such, the promoter, N-terminal signal peptide, tissue, cell type and compartmental localization within the cell may all have a role in determining the precise activity of a particular CLE peptide. Moreover, post-translational modifications of the peptide ligand, such as glycosylation, hydroxylation and arabinosylation, are likely to play a major role in its specificity (Matsubayashi 2011; Meng et al. 2010; Ohyama et al. 2009;

Okamoto et al. 2013). However, there may also be some level of receptor specificity, with each receptor having an optimal CLE peptide ligand (Meng et al. 2010).

7.5 Conclusion and future prospective

Here, we identified four components acting in the regulation of bean nodulation: *PvNARK, PvRIC1, PvRIC2* and *PvNIC1*, in addition to a truncated copy of *PvNARK*, called *PvTrNARK*, and showed that *GmRIC1* of soybean can function in the AON pathway of bean to regulate nodulation in a PvNARK-dependent manner. In addition, mutations in *PvNARK* were identified in the R32 and SV145 supernodulation mutants of common white bean. Future studies aimed at functionally characterizing the CLE peptides and establishing their precise mode of transport and perception, in addition to delineating the importance of their domains and key amino acid residues, will further enhance the understanding of CLE peptide signalling, potentially leading to advances in improving bean nodulation in agriculture.

7.6 Experimental procedures

7.6.1 Plant growth conditions

Seeds of the bean (*Phaseolus vulgaris*) wild-type cultivar, OAC Rico, and its supernodulation mutant, R32 (Park and Buttery 1988, 1989), were surface-sterilized in 70% ethanol for 20 s, followed by rinsing five times with sterile water. Unless otherwise stated, the seeds were sown in sterile grade 3 vermiculite in 4-L pots. Plants were watered daily and supplemented with a modified nutrient solution lacking nitrogen (Herridge, 1982) twice per week. Glasshouse growth conditions were controlled using a 16-h day/8-h night cycle at 28 °C/23 °C, respectively.

For qRT-PCR studies, sterilized seeds of OAC Rico were planted in 2-L pots containing sterile grade 2 vermiculite, and grown in a controlled growth chamber (PGC-9/2; Percival Scientific, Perry, IA) at 28:24 °C over a 16:8 h day:night regime. Plants were watered every second day with distilled water and inoculated with *Rhizobium phaseoli* (detailed below) 5 days following germination. Whole root systems were harvested 6 and 10 days following germination (i.e. 1 and 5 days post-inoculation, respectively) and snap-frozen in liquid nitrogen and stored at −80 °C (n = 9 plants per biological replicate).

7.6.2 Bacteria growth conditions

Rhizobium phaseoli strain CC511 was grown for 48 h at 28 °C in Yeast Mannitol Broth (YMB). Cultures were diluted with water to a final concentration of OD_{600} = 0.01 prior to inoculating plants. Approximately 150 mL of this final concentration was applied per pot.

Agrobacterium rhizogenes strain K599 (Savka et al. 1990) was grown for 48 h at 28 °C on minimal medium agar plates (Broughton and Dilworth 1971) containing 30 µg/mL rifampicin and 50 µg/mL kanamycin.

7.6.3 Identifying the genomic and CDS sequences of PvNARK and PvTrNARK

Candidates of *PvNARK* (Phvul.011G042000) and *PvTrNARK* (Phvul.011G041000) were initially identified via comparative genomics by BLASTp searches of *P. vulgaris* genomic scaffolds (available in the research group of S. Jackson) with the genomic DNA sequence of soybean *GmNARK* (Searle et al. 2003). Once identified, the *PvNARK* candidate was used to design gene-specific primers for its genomic DNA sequence using Primer3 (available at http://biotools.umassmed.edu/bioapps/primer3_www.cgi) software (Table S1). To unequivocally determine the sequence, and to identify a potential mutation in the R32 supernodulation mutant, leaf tissue was collected from mature plants of OAC Rico and R32. The tissue was immediately snap-frozen and homogenized in liquid nitrogen using a mortar and pestle. Seeds of the Swan Valley wild type and its supernodulation mutant SV145 were also homogenized in liquid nitrogen using a mortar and pestle. DNA was subsequently extracted from the samples according to the methods outlined in Winnepenninckx et al. (1993).

To determine the full-length CDS DNA sequence of *PvNARK*, additional leaf tissue was collected from mature plants of OAC Rico and R32. Total RNA was extracted and cDNA synthesis performed according to Hayashi et al. (2012). All genomic and CDS DNA fragments were amplified via PCR using i-Taq (Scientifix, http://www.scientifix.com.au/) DNA polymerases in a PTC-200TM Programmable Thermal Controller (MJ Research, Inc., http://www.mj-research.com/) using specific primers shown in Table S1 and 1.0 µL (100 ng) of genomic DNA, or 1.0 µL (25 ng) of CDS DNA, as template. All gene-specific primers were designed using Primer3 (available at http://biotools.umassmed.edu/bioapps/primer3_www.cgi) software. Samples were heated to 94 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 45 s, annealing at 58

or 60 °C (depending on the primer pair) for 45 s, elongation at 72 °C for 2 min and a final extension at 72 °C for 10 min.

Amplified DNA products were cleaned up using Gel/PCR DNA Fragments Extraction kit (IBI Scientific, cat. No. IB47020), prior to being sent for pyrosequencing at the Australian Genome Research Facility (AGRF, Brisbane, Australia) using the AB 3730*xl* capillary separation DNA Analyzer platform (Applied Biosystems). All reactions were performed for at least three biological replicates and a target amplicon of approximately 1000 bp.

7.6.4 Bioinformatic analyses

Candidates of *PvRIC1* (Phvul.005G096900), *PvRIC2* (Phvul.011G135900) and *PvNIC1*(Phvul.005G097000) were identified via BLASTp searches of *P. vulgaris* genomic scaffolds available at Phytozome (http://www.phytozome.net/) using orthologues of the soybean CLE peptide-encoding genes *GmRIC1, GmRIC2* and *GmNIC1*. Complete coding sequences and multiple sequence alignments, using either the predicted CLE peptide sequences or various NARK orthologues, were subsequently performed using Geneious Pro 5.5.2. The phylogenetic tree of various NARK orthologues was generated using PhyML 3.0 (Guindon and Gascuel, 2003) in Geneious Pro 6.0. The tree was constructed using the maximum likelihood approach. A branch was supported in 1000 bootstrap replications. Genetic synteny within the genomic environments of *GmRIC1, GmRIC2, GmNIC1*, *PvRIC1, PvRIC2* and *PvNIC1*, and also between *GmNARK* and *PvNARK*, was obtained using Phytozome cluster analyses (http://www.phytozome.net/). Domain predictions (Conserved Domain Database, NCBI) of *PvNARK*, and signal peptide predictions (SignalP 3.0) (Bendtsen et al. 2004) of *GmRIC1, GmRIC2, GmNIC1*, *PvRIC1, PvRIC2* and *PvNIC1*, were carried out using the predicted peptide sequences. Molecular modelling of the LRR domains of GmNARK and PvNARK was carried out using amino acid residues 1–601 of the NARK protein (GenBank: AAN74865.1) at the Phyre web server (Kelley and Sternberg 2009) before visualization in the PYMOL molecular graphics system (DeLano, 2002). The best-ranked hit for Phyre threading was to the human toll-like receptor 3 ectodomain (PDB id: 1ziw).

7.6.5 Hairy-root transformation

Agrobacterium rhizogenes–mediated hairy-root transformation was used to induce transgenic roots on bean plants using methods previously described for soybean (Kereszt et al. 2007; Lin et al. 2011b; Reid et al. 2011a, 2013). Bean plants were inoculated with *A. rhizogenes* strain K599 (Savka et al. 1990) harbouring a p15SRK2 integrative vector modified to over-express the full-length coding sequence of the soybean gene, *GmRIC1* (Reid et al. 2011a). Plants with established hairy-root systems were inoculated with *R. phaseoli* strain CC511. Nodule numbers were recorded 18 days post-inoculation, and root system dry weights were subsequently obtained. In all cases, n = 10–12 plants per treatment.

7.6.6 RNA extraction and cDNA synthesis for qRT-PCR studies

RNA was extracted from root tissue and contaminating DNA removed according to Hayashi et al. (2012). cDNA was subsequently synthesized in 20 mL reactions using the Takara Blue Print cDNA synthesis kit according to the manufacturer's instructions. Resulting cDNA was verified using PCR with *PvUBC9* primers (Phvul.006G110100; Hernández et al. 2007; Table S2).

7.6.7 qRT-PCR

Quantitative real-time PCR (qRT-PCR) was performed according to the methods outlined in Hayashi et al. (2012). Primers were designed using Primer3 (Untergrasser et al. 2012) and are listed in Table S2. LinRegPCR 7.5 software (Ramakers et al. 2003) was used to determine PCR efficiency for each reaction. *PvUBC9* (Hernández et al. 2007) was used as a housekeeping gene to determine transcript abundance.

7.6.8 Statistical analysis

Statistical differences amongst treatment groups were determined using Student's t tests to generate P values. A threshold of $P \leq 0.05$ was used to indicate a significant statistical difference between values. For qRT-PCR studies, all values and SEs were generated using group averages.

7.7 Acknowledgements

This work was funded by the Australian Research Council for the provision of a Centre of Excellence grant (CEO348212) and Discovery Project grants (DP130103084 and DP130102266) as well as the University of Queensland for strategic funds and Research Scholarship support. We would like to thank Dr Alireza Navabi and Terry Rupert, of Agriculture and Agri-Food Canada, and the Department of Plant Agriculture, the University of Guelph, for the provision of OAC Rico, Swan Valley, R32 and SV145 seed. Alina Tollenaere, Satomi Hayashi, Stacey Cook, Saeid Mirzaei and Bethany van Hameren are thanked for technical assistance.

7.8 References

Becher M, Schepl U, Schubert S. 1997. N 2 Fixation during different physiological stages of *Phaseolus vulgaris* OAC Rico and its supernodulating mutant R32BS15: The role of assimilate supply to and export from nodules. *Journal of Plant Physiology* **150**(1): 31-36.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* **340**(4): 783-795.

Broughton W, Dilworth M. 1971. Control of leghaemoglobin synthesis in snake beans. *Biochemical Journal* **125**(4): 1075-1080.

Buttery B, Park S. 1990. Effects of nitrogen, inoculation and grafting on expression of supernodulation in a mutant of *Phaseolus vulgaris* L. *Canadian Journal of Plant Science* **70**(2): 375-381.

Buttery B, Park S-J. 1993. Characterization of some non-fixing mutants of common bean (*Phaseolus vulgaris* L.). *Canadian Journal of Plant Science* **73**(4): 977-983.

Buttery B, Park S, Bernard S, Streit W, Werner D. 1990. Effects of *Rhizobium* inoculum concentration strain and combined nitrogen on growth and nodulation of a supernodulating common bean and its parent line. *Canadian Journal of Plant Science* **70**(4): 987-996.

Carroll BJ, McNeil DL, Gresshoff PM. 1985a. A supernodulation and nitrate-tolerant symbiotic (*nts*) soybean mutant. *Plant Physiology* **78**(1): 34-40.

Carroll BJ, McNeil DL, Gresshoff PM. 1985b. 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* **82**(12): 4162-4166.

Day A, Price G, Schuller K, Gresshoff P. 1987. Nodule physiology of a supernodulating soybean (*Glycine max*) mutant. *Functional Plant Biology* **14**(5): 527-538.

DeLano WL. 2002. The PyMOL molecular graphics system. San Carlos, CA: DeLano Scientific.

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**(2): 588-590.

Delves AC, Carroll BJ, Gresshoff PM. 1988. Genetic analysis and complementation studies on a number of mutant supernodulating soybean lines. *Journal of Genetics* **67**(1): 1-8.

Ding Y, Kalo P, Yendrek C, Sun J, Liang Y, Marsh JF, Harris JM, Oldroyd GE. 2008. Abscisic acid coordinates nod factor and cytokinin signaling during the regulation of nodulation in *Medicago truncatula*. *The Plant Cell* **20**(10): 2681-2695.

Estrada-Navarrete G, Alvarado-Affantranger X, Olivares J-E, Guillén G, Díaz-Camino C, Campos F, Quinto C, Gresshoff PM, Sanchez F. 2007. Fast, efficient and reproducible genetic transformation of *Phaseolus* spp. by *Agrobacterium rhizogenes*. *Nature Protocols* **2**(7): 1819-1824.

Ferguson BJ. 2013. The development and regulation of soybean nodules: *In A Comprehensive Survey of International Soybean Research - Genetics, Physiology, Agronomy, and Nitrogen Relationships* (Board, J.E., ed.), pp. 31–47. Rijeka, Croatia: InTech – Open Access.

Ferguson BJ, Mathesius U. 2003. Signaling interactions during nodule development. *Journal of Plant Growth Regulation* **22**(1): 47-72.

Ferguson BJ, Ross JJ, Reid JB. 2005. Nodulation phenotypes of gibberellin and brassinosteroid mutants of *Pisum sativum*. *Plant Physiology* **138**(4): 2396-2405.

Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE, Gresshoff PM. 2010. Molecular analysis of legume nodule development and autoregulation. *Journal of Integrative Plant Biology* **52**(1): 61-76.

Ferguson BJ, Foo E, Ross JJ, Reid JB. 2011. Relationship between gibberellin, ethylene and nodulation in *Pisum sativum*. *New Phytologist* **189**(3): 829-842.

Ferguson B, Lin M-H, Gresshoff PM. 2013. Regulation of legume nodulation by acidic growth conditions. *Plant Signaling & Behavior* **8**(3): e23426.

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**(9): 818-823.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**(5): 696-704.

Guinel FC, Geil R. 2002. A model for the development of the rhizobial and arbuscular mycorrhizal symbioses in legumes and its use to understand the roles of ethylene in the establishment of these two symbioses. *Canadian Journal of Botany* **80**(7): 695-720.

Hamaguchi H, Kokubun M, Yoneyama T, Hansen AP, Akao S. 1993. Control of supernodulation in intergeneric grafts of soybean and common bean. *Crop Science* **33**(4): 794-797.

Hansen AP, Martin P, Buttery BR, Park SJ. 1992. Nitrate inhibition of N2 fixation in *Phaseolus vulgaris* L. cv. OAC Rico and a supernodulating mutant. *New Phytologist* **122**(4): 611-615.

Hansen AP, Yoneyama T, Kouchi H, Hiraoka K. 1993a. Respiration and nitrogen fixation of hydroponically cultured *Phaseolus vulgaris* L. cv. OAC Rico and a supernodulating mutant. *Planta* **189**(4): 546-556.

Hansen AP, Yoneyama T, Kouchi H, Hiraoka K. 1993b. Respiration and nitrogen fixation of hydroponically cultured Phaseolus vulgaris L. cv. OAC Rico and a supernodulating mutant: II. Distribution and respiration of recently fixed 13 C-labelled photosynthate and the effect of sink removal on carbon partitioning. *Planta* **189**(4): 546-556.

Harper J, Corrigan K, Barbera A, Abd-Alla M. 1997. Hypernodulation of soybean, mung bean, and hyacinth bean is controlled by a common shoot signal. *Crop Science* **37**(4): 1242-1246.

Hayashi S, Reid DE, Lorenc MT, Stiller J, Edwards D, Gresshoff PM, Ferguson BJ. 2012. Transient Nod factor-dependent gene expression in the nodulation-competent zone of soybean (*Glycine max* [L.] Merr.) roots. *Plant Biotechnology Journal* **10**(8): 995-1010.

Hayashi S, Gresshoff PM, Ferguson BJ 2013. Systemic signalling in legume nodulation: nodule formation and its regulation. *Long-Distance Systemic Signaling and Communication in Plants*: (Baluska, F., ed.), pp. 219–229. Heidelberg, Germany: Springer-Verlag.

Hernández G, Ramírez M, Valdés-López O, Tesfaye M, Graham MA, Czechowski T, Schlereth A, Wandrey M, Erban A, Cheung F. 2007. Phosphorus stress in common bean: root transcript and metabolic responses. *Plant Physiology* **144**(2): 752-767.

Herridge DF. 1982. Relative abundance of ureides and nitrate in plant tissues of soybean as a quantitative assay of nitrogen fixation. *Plant Physiology* **70**(1): 1-6.

Hirsch AM, Fang Y. 1994. Plant hormones and nodulation: what's the connection? *Plant Molecular Biology* **26**(1): 5-9.

Jensen ES, Peoples MB, Boddey RM, Gresshoff PM, Hauggaard-Nielsen H, Alves BJ, Morrison MJ. 2012. Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. A review. *Agronomy for Sustainable Development* **32**(2): 329-364.

Jeong S, Trotochaud AE, Clark SE. 1999. The Arabidopsis *CLAVATA2* gene encodes a receptor-like protein required for the stability of the CLAVATA1 receptor-like kinase. *The Plant Cell* **11**(10): 1925-1933.

Kelley LA, Sternberg MJ. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**(3): 363-371.

Kereszt A, Li D, Indrasumunar A, Nguyen CD, Nontachaiyapoom S, Kinkema M, Gresshoff PM. 2007. *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. *Nature Protocols* **2**(4): 948-952.

Kosslak RM, Bohlool BB. 1984. Suppression of nodule development of one side of a split-root system of soybeans caused by prior inoculation of the other side. *Plant Physiology* **75**(1): 125-130.

Krusell L, Madsen LH, Sato S, Aubert G, Genua A, Szczyglowski K, Duc G, Kaneko T, Tabata S, de Bruijn F. 2002. Shoot control of root development and nodulation is mediated by a receptor-like kinase. *Nature* **420**(6914): 422-426.

Krusell L, Sato N, Fukuhara I, Koch BE, Grossmann C, Okamoto S, Oka-Kira E, Otsubo Y, Aubert G, Nakagawa T. 2011. The *Clavata2* genes of pea and *Lotus japonicus* affect autoregulation of nodulation. *The Plant Journal* **65**(6): 861-871.

Lavin M, Herendeen PS, Wojciechowski MF. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Systematic Biology* **54**(4): 575-594.

Lim CW, Lee YW, Hwang CH. 2011. Soybean nodule-enhanced CLE peptides in roots act as signals in GmNARK-mediated nodulation suppression. *Plant and Cell Physiology* **52**(9): 1613-1627.

Lin J-Y, Stupar RM, Hans C, Hyten DL, Jackson SA. 2010a. Structural and functional divergence of a 1-Mb duplicated region in the soybean (*Glycine max*) genome and comparison to an orthologous region from *Phaseolus vulgaris*. *The Plant Cell* **22**(8): 2545-2561.

Lin YH, Ferguson BJ, Kereszt A, Gresshoff PM. 2010b. Suppression of hypernodulation in soybean by a leaf-extracted, NARK-and Nod factor-dependent, low molecular mass fraction. *New Phytologist* **185**(4): 1074-1086.

Lin Y-H, Lin M-H, Gresshoff PM, Ferguson BJ. 2011a. An efficient petiole-feeding bioassay for introducing aqueous solutions into dicotyledonous plants. *Nature Protocols* **6**(1): 36-45.

Lin M-H, Gresshoff PM, Indrasumunar A, Ferguson BJ. 2011b. *pHairyRed*: A novel binary vector containing the *DsRed2* reporter gene for visual selection of transgenic hairy roots. *Molecular Plant* **4**(3): 537-545.

Lin M-H, Gresshoff PM, Ferguson BJ. 2012. Systemic regulation of soybean nodulation by acidic growth conditions. *Plant Physiology* **160**(4): 2028-2039.

Lohar DP, VandenBosch KA. 2005. Grafting between model legumes demonstrates roles for roots and shoots in determining nodule type and host/rhizobia specificity. *Journal of Experimental Botany* **56**(416): 1643-1650.

Matsubayashi Y. 2014. Posttranslationally modified small-peptide signals in plants. *Annual Review of Plant Biology* **65**: 385-413.

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* **1**(3): 147-155.

Meng L, Ruth KC, Fletcher JC, Feldman L. 2010. The roles of different CLE domains in Arabidopsis CLE polypeptide activity and functional specificity. *Molecular Plant* **3**(4): 760-772.

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**(37): 25381-25391.

Miyazawa H, Oka-Kira E, Sato N, Takahashi H, Wu G-J, Sato S, Hayashi M, Betsuyaku S, Nakazono M, Tabata S. 2010. The receptor-like kinase KLAVIER mediates systemic regulation of nodulation and non-symbiotic shoot development in *Lotus japonicus*. *Development* **137**(24): 4317-4325.

Mortier V, Den Herder G, Whitford R, Van de Velde W, Rombauts S, D'haeseleer K, Holsters M, Goormachtig S. 2010. CLE peptides control *Medicago truncatula* nodulation locally and systemically. *Plant Physiology* **153**(1): 222-237.

Mortier V, De Wever E, Vuylsteke M, Holsters M, Goormachtig S. 2012. Nodule numbers are governed by interaction between CLE peptides and cytokinin signaling. *The Plant Journal* **70**(3): 367-376.

Ni J, Clark SE. 2006. Evidence for functional conservation, sufficiency, and proteolytic processing of the CLAVATA3 CLE domain. *Plant Physiology* **140**(2): 726-733.

Ni J, Guo Y, Jin H, Hartsell J, Clark SE. 2011. Characterization of a CLE processing activity. *Plant Molecular Biology* **75**(1-2): 67-75.

Ni J, Guo Y, Jin H, Hartsell J, Clark SE. 2011. Characterization of a CLE processing activity. *Plant Molecular Biology* **75**(1-2): 67-75.

Nishimura R, Hayashi M, Wu G-J, Kouchi H, Imaizumi-Anraku H, Murakami Y, Kawasaki S, Akao S, Ohmori M, Nagasawa M. 2002. HAR1 mediates systemic regulation of symbiotic organ development. *Nature* **420**(6914): 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**(7): 769-780.

Oelkers K, Goffard N, Weiller GF, Gresshoff PM, Mathesius U, Frickey T. 2008. Bioinformatic analysis of the CLE signaling peptide family. *BMC Plant Biology* **8**(1): 1.

Ohyama K, Shinohara H, Ogawa-Ohnishi M, Matsubayashi Y. 2009. A glycopeptide regulating stem cell fate in *Arabidopsis thaliana*. *Nature Chemical Biology* **5**(8): 578-580.

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**(1): 67-77.

Okamoto S, Shinohara H, Mori T, Matsubayashi Y, Kawaguchi M. 2013. Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. *Nature Communications* **4**: 2191.

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**(4): 1347-1352.

Osipova MA, Mortier V, Demchenko KN, Tsyganov VE, Tikhonovich IA, Lutova LA, Dolgikh EA, Goormachtig S. 2012. *WUSCHEL-RELATED HOMEOBOX5* gene

expression and interaction of CLE peptides with components of the systemic control add two pieces to the puzzle of autoregulation of nodulation. *Plant Physiology* **158**(3): 1329-1341.

Park SJ, Buttery B. 1988. Nodulation mutants of white bean (*Phaseolus vulgaris* L.) induced by ethyl-methane sulphonate. *Canadian Journal of Plant Science* **68**(1): 199-202.

Park SJ, Buttery B. 1989. Inheritance of nitrate-tolerant supernodulation in EMS-induced mutants of common bean (*Phaseolus vulgaris* L.). *Journal of Heredity* **80**(6): 486-488.

Park SJ, Buttery B. 1994. Inheritance of non-nodulation and ineffective nodulation mutants in common bean (*Phaseolus vulgaris* L.). *Journal of Heredity* **85**(1): 1-3.

Park SJ, Buttery B. 1997. Complementation of nodulation genes of various mutants in common bean (*Phaseolus vulgaris* L.). *Journal of Heredity* **88**(6): 543-545.

Ramakers C, Ruijter JM, Deprez RHL, Moorman AF. 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters* **339**(1): 62-66.

Reid DE, Ferguson BJ, Gresshoff PM. 2011a. Inoculation-and nitrate-induced CLE peptides of soybean control NARK-dependent nodule formation. *Molecular Plant-Microbe Interactions* **24**(5): 606-618.

Reid DE, Ferguson BJ, Hayashi S, Lin Y-H, Gresshoff PM. 2011b. Molecular mechanisms controlling legume autoregulation of nodulation. *Annals of Botany* **108**(5): 789-795.

Reid DE, Hayashi S, Lorenc M, Stiller J, Edwards D, Gresshoff PM, Ferguson BJ. 2012. Identification of systemic responses in soybean nodulation by xylem sap feeding and complete transcriptome sequencing reveal a novel component of the autoregulation pathway. *Plant Biotechnology Journal* **10**(6): 680-689.

Reid DE, Li D, Ferguson BJ, Gresshoff PM. 2013. Structure–function analysis of the Gm RIC1 signal peptide and CLE domain required for nodulation control in soybean. *Journal of Experimental Botany* **64**(6): 1575-1585.

Ryu H, Cho H, Choi D, Hwang I. 2012. Plant hormonal regulation of nitrogen-fixing nodule organogenesis. *Molecules and Cells* **34**(2): 117-126.

Savka MA, Ravillion B, Noel G, Farrand S. 1990. Induction of hairy roots on cultivated soybean genotypes and their use to propagate the soybean cyst nematode. *Phytopathology* **80**(5): 503-508.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J. 2010. Genome sequence of the palaeopolyploid soybean. N*ature* **463**(7278): 178-183.

Schnabel E, Journet E, 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**(6): 809-822.

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**(5603): 109-112.

Sheng C, Harper JE. 1997. Shoot versus root signal involvement in nodulation and vegetative growth in wild-type and hypernodulating soybean genotypes. *Plant Physiology* **113**(3): 825-831.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**(15): e115.

Winnepenninckx B, Backeljau T, De Wachter R. 1993. Extraction of high molecular weight DNA from molluscs. *Trends in genetics:* **9**(12): 407-407.

Chapter 8

8.1 General discussion and conclusion

Legume nodulation requires several complex signalling mechanisms to establish, control and maintain the symbiosis between the host plant and the beneficial rhizobia bacteria. Of particular interest are the mechanisms and signalling molecules which prevent rhizobia from over colonizing the roots of the plant, thus maintaining homeostasis. Without these, rhizobia have the ability to excessively draw on photoassimilates and significantly reduce the above ground biomass of the host plant. One such signalling molecule is CLAVATA/ESR-related (CLE) peptides, which are not only involved in nodulation control, but a myriad of other plant growth and developmental processes (Hastwell et al. 2015b). The research in this thesis identifies, genetically characterises gene families and further functionally characterises some of these CLE peptide-encoding genes from a range of legume species.

The CLE peptide-encoding gene family of five legumes (soybean, common bean, *M. truncatula*, *L. japonicus* and pea), were identified in Chapters 3, 4 and 6 (Hastwell et al. 2015a; 2017 and 2018). To date, very few of those identified have been functionally characterised, of which, most are related to nodulation processes. One such approach to identifying the potential role of these genes is to first bioinformatically compare them to genes with a known function within the model species, *A. thaliana*. However, this is also limiting as many of the CLE peptide-encoding genes are not yet, or poorly characterised; *A. thaliana* lacks the ability to undergo two common symbiosis: mycorrhization and nodulation; and many techniques available in *A. thaliana* are not yet available or optimised in legumes. More recently, a thorough genome-wide analysis of CLE peptide-encoding genes in response to macronutrient application and starvation as well as nodulation was completed in *M. truncatula* (de Bang et al. 2017). When these findings are used in conjunction with the bioinformatic resources presented in this thesis together with the evidence presented that indicates a high level of functional conservation across species (Ferguson et al. 2014; Hastwell et al. 2015a; 2017 and 2018), there is potential to significantly advance the characterisation of CLE peptides in other legumes.

Outside of the Fabaceae family and *A. thaliana*, advances in the identification and functional characterisation of CLE peptide-encoding genes and their associated pathways

have been made. This includes an array of CLE peptides from 57 genomes (Goad et al. 2017); and more comprehensive studies in poplar (Han et al. 2016); tomato (Xu et al. 2015; Zhang et al. 2014); rice (Kinoshita et al. 2007) ; maize (Je et al. 2016); radish (Dodueva et al, 2013; Gancheva et al. 2016); and the only species to contain CLE peptides outside of the plant kingdom, nematodes (Guo et al. 2011; Lu et al. 2009; Wang et al. 2010; Wang et al. 2011; Wubben et al. 2015). These resources used together will facilitate the extrapolation of research findings and any agronomic advancements across different species.

To further characterise CLE peptide-encoding genes and possibly improve plant growth and development, it would be beneficial to use the bioinformatic resources in Chapters 3, 4 and 6, in conjunction with methods such as targeted transcriptomic analyses (de Bang et al. 2017), peptide feeding studies (Corcilius et al. 2017; Hastwell et al. 2018) and other molecular biology techniques, such as those used in Chapter 8 (Ferguson et al. 2014), to gain a clearer understanding of the role CLE peptide-encoding genes within plant development.

One recently developed technique, CRISPR/Cas9-mediated gene editing, will prove to be an invaluable tool and has already been utilised to gain mutants of the 32 CLE peptide-encoding genes of *A. thaliana* (Yamaguchi et al. 2017). It is cumbersome to generate stable transformants in model legumes using such techniques (Cai et al. 2018), however, it will be more valuable analyse the CLE peptides and associated pathways of crop legumes using stable mutations rather than using the well-established chimeric hairy-root transformation technique. Using CRISPR, fruit size has shown to be increased in tomato (Xu et al. 2015) and it is likely that similar advances can be made in legumes.

Furthermore, this type of reverse genetic characterisation workflow could be applied to other peptide and protein families such as CEP, RALF and arabinosyltransferases, where little analysis has been performed outside the model species *A. thaliana*, which as has developmental pathway limitations and in some cases findings cannot be extrapolated to other species.

A significant hinderance to further elucidating the function of CLE peptides is the inability to isolate and characterise endogenous functionally mature CLE peptides. A further limitation is that those identified via overexpression and subsequent mass spectrometry contain an uncommon post-translational glycosylation which cannot be synthesised using

traditional peptide synthesis methods. The few previous methods for synthesising the central triarabinosylated hydroxyproline have been improved upon in Chapter 5 (Corcilius et al. 2017). Despite previous reports of synthetic CLE peptides exhibiting activity without glycosylation, the modification considerably increased the activity of GmCLE40a (identified in Hastwell et al. 2015a; Chapter 3), which suppresses root growth when applied exogenously (Corcilius et al. 2017; Chapter 5). With newly available technology, the function of two other CLE peptides in nodulation with and without arabinosylation were also examined (Hastwell et al. 2018; Chapter 6). Developing an arabinosylated CLE peptide library would significantly advance our understanding of their associated molecular signalling mechanisms. However, caution also needs to be taken to ensure that the peptide is applied correctly spatially and temporally as studies show significant redundancy in the structure and function of CLE peptides (Hastwell et al. 2015b; 2018).

The functional reason behind arabinosylation also needs to be explored. There is some evidence to suggest that it promotes conformational changes that allow the ligand to be perceived (Shinohara and Matsubayashi 2013), however this was not supported by findings in Corcilius et al. (2017). It is possible that the modification is instead related to transport, degradation prevention, cleavage or is indeed involved in pathway and/or peptide specific ligand-receptor binding (Gomord and Faye 2014). Further studies are required for each peptide and pathway to determine the role of CLE peptide post-translational modifications with arabinose and other glycosyl groups.

Many signalling components of the Autoregulation of Nodulation and Nitrate regulation of nodulation pathways are yet to be identified and characterised such as the shoot-to-root signal that occurs following CLE peptide perception. Although some findings suggest a shoot derived cytokinin is the systemic negative regulator of nodulation (Sasaki et al. 2014), there is recent evidence presented that a microRNA promoting nodulation is subsequently downregulated following CLE perception and this in turn suppresses nodulation (reviewed in Ferguson et al. 2018). Other unknown mechanisms include the binding conformation of the CLE ligand and receptor kinase and the subsequent signalling cascade before the miRNA suppression; the method of CLE peptide root-to shoot transport; and the root-derived inhibitor (or promoter) of nodulation following perception of the nitrate-induced CLE peptide. One approach which will prove useful in identifying and characterising new factors and mechanisms in nodulation regulation is to utilise newly developed omics approaches in combination that examine more than a single target. This

could include looking at the transcriptome, peptidome, degradome and the exome in addition to traditional genomic studies.

Although the molecular signalling mechanisms of nodulation are of great interest to determine, the biological significance of these pathways should be simultaneously explored. One important consideration is the relationship between nodulation; mycorrhization; lateral root development; and the acquisition and allocation of nutrient resources (Yokota and Hayashi 2011; Ferguson et al. 2018). Many of these developmental processes share similar CLAVATA-like pathway mechanisms as described in Chapter 1. How these interact to provide plants with evolutionary advantages will better our understanding of nutrient use and may help to alleviate problems arising from nutrient use efficiency.

Finally, together the studies presented in this thesis identified and bioinformatically characterised CLE peptide-encoding gene families which can now be used as a resource to further characterise the role of individual genes within plant development; characterised a number of CLE peptides in nodulation and increased our understanding of nodulation control pathways, including establishing conservation between legume species; and developed an improved method for synthesising arabinosylated CLE peptides. The findings presented have significantly increased our understanding of CLE peptides, plant signalling and development.

8.2 References

Cai Y, Chen L, Liu X, Guo C, Sun S, Wu C, Jiang B, Han T, Hou W. (2018). CRISPR/Cas9-mediated targeted mutagenesis of *GmFT2a* delays flowering time in soya bean. *Plant Biotechnology Journal* **16**(1): 176-185.

Corcilius L, Hastwell AH, Zhang M, Williams J, Mackay JP, Gresshoff PM, Ferguson BJ, Payne RJ. 2017. Arabinosylation modulates the growth-regulating activity of the peptide hormone CLE40a from Soybean. *Cell Chemical Biology* **24**: 1347-1355

de Bang TC, Lundquist PK, Dai X, Boschiero C, Zhuang Z, Pant P, Torres-Jerez I, Roy S, Nogales J, Veerappan V, and Dickstein R. 2017. Genome-wide identification of *Medicago* peptides involved in macronutrient responses and nodulation. *Plant Physiology* **175**(4):1669-1689.

Dodueva IE, Kiryushkin AS, Yurlova EV, Osipova MA, Buzovkina IS, Lutova LA. 2013. Effect of cytokinins on expression of radish CLE genes. *Russian Journal of Plant Physiology* **60**(3):388-395.

Ferguson BJ, Li D, Hastwell AH, Reid DE, Li Y, Jackson SA, Gresshoff PM. 2014. The soybean (*Glycine max*) nodulation-suppressive CLE peptide, GmRIC1, functions interspecifically in common white bean (*Phaseolus vulgaris*), but not in a supernodulating line mutated in the receptor PvNARK. *Plant Biotechnology Journal* **12**(8): 1085-1097.

Ferguson BJ, Mens C, Hastwell AH, Zhang MB, Su H, Jones CH, Chu XT, Gresshoff PM. 2018. Legume nodulation: the host controls the party. *Plant, Cell & Environment*, https://doi.org/10.1111/pce.13348.

Gancheva MS, Dodueva IE, Lebedeva MA, Tvorogova VE, Tkachenko AA, Lutova LA. 2016. Identification, expression, and functional analysis of CLE genes in radish (*Raphanus sativus* L.) storage root. *BMC Plant Biology* **16**(1): 7.

Goad DM, Zhu C, Kellogg EA. 2017. Comprehensive identification and clustering of CLV3/ESR-related (CLE) genes in plants finds groups with potentially shared function. *New Phytologist* **216**(2): 605-616.

Gomord V, Faye L. (2004). Posttranslational modification of therapeutic proteins in plants. *Current Opinion in Plant Biology* **7**(2):171-181.

Guo Y, Ni J, Denver R, Wang X, Clark SE. 2011. Mechanisms of molecular mimicry of plant CLE peptide ligands by the parasitic nematode *Globodera rostochiensis*. *Plant Physiology* **157**:476–484.

Han H, Zhang G, Wu M, Wang G. 2016. Identification and characterization of the *Populus trichocarpa* CLE family. *BMC genomics* **17**(1):174.

Hastwell AH, Corcilius L, Williams J, Gresshoff PM, Payne RJ Ferguson BJ. 2018. Triarabinosylation is required for nodulation-suppressive CLE peptides to systemically inhibit nodulation in *Pisum sativum*. *Plant, Cell & Environment*, https://doi.org/10.1111/pce.13325.

Hastwell AH, de Bang TC, Gresshoff PM, Ferguson BJ. 2017. CLE peptide-encoding gene families *in Medicago truncatula* and *Lotus japonicus*, compared with those of soybean, common bean and Arabidopsis. *Scientific Reports* **7**.

Hastwell AH, Gresshoff PM, Ferguson BJ. 2015a. Genome-wide annotation and characterization of CLAVATA/ESR (CLE) peptide hormones of soybean (*Glycine max*) and common bean (*Phaseolus vulgaris*), and their orthologues of *Arabidopsis thaliana*. *Journal of Experimental Botany* **66**(17): 5271-5287.

Hastwell AH, Gresshoff PM, Ferguson BJ. 2015b. The structure and activity of nodulation-suppressing CLE peptide hormones of legumes. *Functional Plant Biology* **42**(3): 229-238.

Je BI, Gruel J, Lee YK, Bommert P, Arevalo ED, Eveland AL, Wu Q, Goldshmidt A, Meeley R, Bartlett M, Komatsu M, Sakai H, Jönsson H, Jackson D. 2016. Signaling from maize organ primordia via FASCIATED EAR3 regulates stem cell proliferation and yield traits. *Nature Genetics* **48**: 785-791.

Kinoshita A, Nakamura Y, Sasaki E, Kyozuka J, Fukuda H, Sawa S. 2007. Gain-of-function phenotypes of chemically synthetic CLAVATA3/ESR-related (CLE) peptides in *Arabidopsis thaliana* and *Oryza sativa*. *Plant and Cell Physiology* **48**(12): 1821-1825.

Lu S-W, Chen S, Wang J, Yu H, Chronis D, Mitchum MG, Wang X. 2009. Structural and functional diversity of *CLAVATA3/ESR (CLE)*-like genes from the potato cyst nematode *Globodera rostochiensis*. *Molecular Plant-Microbe Interactions*. **22**:1128–1142.

Sasaki T, Suzaki T, Soyano T, Kojima M, Sakakibara H, Kawaguchi M. (2014). Shoot-derived cytokinins systemically regulate root nodulation. *Nature Communications* **5**: 4983.

Shinohara H, Matsubayashi Y. 2013. Chemical synthesis of Arabidopsis CLV3 glycopeptide reveals the impact of hydroxyproline arabinosylation on peptide conformation and activity. *Plant and Cell Physiology* **54**: 369-374.

Wang J, Joshi S, Korkin D, Mitchum MG. 2010. Variable domain I of nematode CLEs directs post-translational targeting of CLE peptides to the extracellular space. *Plant Signalling and Behaviour* **5**(12): 1633-1635.

Wang J, Replogle AM, Hussey R, Baum T, Wang X, Davis EL, Mitchum MG. 2011. Identification of potential host plant mimics of CLAVATA3/ESR (CLE)-like peptides from the plant-parasitic nematode *Heterodera schachtii*. *Molecular Plant Pathology* **12**(2): 177-186.

Wubben MJ, Gavilano L, Baum TJ, Davis EL. 2015. Sequence and spatiotemporal expression analysis of CLE-motif containing genes from the reniform nematode (*Rotylenchulus reniformis* Linford & Oliveira). *Journal of Nematology* **47**(2):159-165.

Xu C, Liberatore KL, MacAlister CA, Huang Z, Chu YH, Jiang K, Brooks C, Ogawa-Ohnishi M, Xiong G, Pauly M, Van Eck J 2015. A cascade of arabinosyltransferases controls shoot meristem size in tomato. *Nature Genetics* 47(7): 784-796.

Yamaguchi YL, Ishida T, Yoshimura M, Imamura Y, Shimaoka C, Sawa S. 2017. A collection of mutants for CLE-peptide-encoding genes in *Arabidopsis* generated by CRISPR/Cas9-mediated gene targeting. *Plant and Cell Physiology* **58**(11): 1848-1856.

Yokota K, Hayashi M. (2011). Function and evolution of nodulation genes in legumes. *Cellular and Molecular Life Sciences*. ;68(8): 1341-1351.

Zhang Y, Yang S, Song Y, Wang J. 2014. Genome-wide characterization, expression and functional analysis of CLV3/ESR gene family in tomato. *BMC Genomics*. **15**(1):827.