A forward genetic screen for genes involved in the arbuscular mycorrhizal symbiosis in the model legume *Medicago truncatula*

Thomas Barrance Irving

Doctor of Philosophy

University of York

Biology

September 2017

Abstract

The arbuscular mycorrhizal symbiosis (AMS) provides a significant part of nutrient uptake for the majority of plant species. Engineering increased symbiotic potential in crops offers great benefits for agriculture, reducing the demand for fertilisers and increasing resiliency to disease and abiotic stress. We attempted to increase understanding of the AMS, by identifying genes involved in the symbiosis, focusing on the poorly understood parts of the symbiotic process outside of the legume common symbiosis pathway.

Prior work had carried out an initial screen of the model legume *Medicago truncatula*, mutagenised with the retrotransposon tnt1, to obtain lines showing a phenotype of impaired arbuscular mycorrhizal colonisation while retaining normal rhizobial colonisation. This project took candidate lines from that screen, and used morphological and genetic phenotyping to confirm four *Medicago* lines with defects in different parts of the AMS.

We developed a computational pipeline to quickly locate the 30-60 tnt1 insertions in each mutant line with Illumina whole genome sequencing (WGS). We backcrossed the mutants to produce populations segregating for the different insertions. This population was genotyped for the insertions located by WGS. Co-segregation analysis was used to show correlation between tnt1 insertions and the impaired arbuscular mycorrhizal colonisation phenotype in these lines. We demonstrate that line NF3438 is a *ram1* mutant (a GRAS transcription factor that is a central regulator of mycorrhizal cell fate), that NF443 is a *kin3* mutant (a membrane kinase of unknown function), and that the delayed colonisation phenotype of NF3209 is likely caused by a somaclonal mutation. We examined the expression of mycorrhizal genes in NF443, and propose a hypothesis for the function of KIN3 in the arbuscular mycorrhizal symbiosis.

Finally, we attempted to replicate reports that plant mycorrhizal colonisation phenotypes are dependent on fungal genotype, and question the assumed universality of signalling across this highly generalist symbiosis.

Table of Contents

Abstract **2**

Table of Contents **3**

List of Figures **8**

List of Tables **11**

Acknowledgements **12**

Author's declaration **13**

Chapter 1 – General Introduction **14**

1.1 – WHAT IS THE ARBUSCULAR MYCORRHIZAL SYMBIOSIS? 14

1.2 – WHY THE AMS MATTERS 17

1.3 – THE COMMON SYMBIOSIS SIGNALLING PATHWAY 18

1.3.1 – How rhizobia have advanced mycorrhizal research 18

1.3.2 – The common symbiosis pathway 20

1.3.3 – Differences in the CSP between the symbioses 22

1.3.4 – After the CSP 24

1.4 – PICKING UP THE PHONE: PRE-CONTACT COMMUNICATION TO ESTABLISH SYMBIOSIS 24

1.4.1 – Plant signals to the fungi 24

1.4.2 – Fungal signals to the plant 28

1.5 – LIFE INSIDE THE ROOT 31

1.5.1 – A highway through the root: how does the plant accommodate hyphal growth 31

1.5.2 – Minerals for bread and butter: symbiotic nutrient exchange 33

1.5.3 – Master switch(es) of the AMS 35

1.6 – HOUSTON, WE HAVE LANDED: THE FUNGI THAT FED THE CONQUEST OF LAND 39

1.7 – AN ANCIENT SYMBIOSIS: WHY HAS THE AMS REMAINED STABLE ACROSS EVOLUTIONARY TIME 41

1.7.1 – The ever-present threat 41

1.7.2 – Control of reciprocal exchange 42

1.7.3 – An updated model? 43

1.7.4 – For every carrot, a stick 45

1.7.5 – Is carbon flux monodirectional? 46

1.7.6 – Conclusions 48

1.8 – LOOKING PAST PHOSPHORUS: WHAT ARE THE BENEFITS OF THE AMS? 49

1.8.1 – A wide array of published benefits 49

1.8.2 – Plant nutrition 50

1.8.3 – Abiotic stress tolerance 51

1.8.4 – Mycorrhizal induced resistance to pests and pathogens 52

1.9 – MOVING BEYOND MONOXENIC CULTURE 53

1.9.1 – Competition 53

1.9.2 – Less to go around: Pathogen and fungivore activity in the field 54

1.9.3 – A trade-off between defence and symbiosis? 54

1.9.4 – Friends as well as foes 56

1.9.5 – A fractal symbiosis: AMF have their own symbionts 57

1.9.5 – What is 'a' fungus? 58
1.10 – A FORWARD GENETIC SCREEN FOR NEW MYCORRHIZAL GENES 59

Chapter 2 – Materials and Methods 61
2.1 – MEDIA & MATERIALS 61
2.2 – SEED PREPARATION 61
2.3 – FUNGAL MATERIAL 61
2.3.1 – Production of root organ cultures to propagate AMF 62
2.3.2 – AMF production by hairy root culture 62
2.3.3 – AMF production by stock pot culture 63
2.4 – DNA EXTRACTION, AMPLIFICATION AND SEQUENCING 64
2.4.1 – DNA extraction 64
2.4.2 – Nucleic acid quantification 64
2.4.3 – Triplex PCR 64
2.5 – MICROSCOPY 65
2.5.1 – Ink staining of mycorrhizae 65
2.5.2 – WGA-AF488 staining of mycorrhizae 66

Chapter 3 – Phenotypic analysis of *Medicago truncatula* lines with aberrant responses to AM fungi 67
3.1 – INTRODUCTION 67
3.2 – METHODS 68
3.2.1 – Plant growth 68
3.2.2 – Quantification of gene expression 69
3.3 – RESULTS 69
3.3.1 – Confirming the reduced colonisation phenotype 69
3.3.2 – Arbuscule structure 75
3.3.3 – Reduced colonisation or delayed colonisation? 82
3.3.4 – Does reduced colonisation impair plant growth? 84
3.3.5 – Changes in gene expression in the mutant lines 87
3.4 – DISCUSSION 88
3.4.1 – The mutant phenotypes 88
3.4.2 – Comparison of different methods for measuring colonisation 89

Chapter 4 – Sequencing of *Medicago truncatula* lines exhibiting reduced AM colonisation 91
4.1 – INTRODUCTION 91
4.2 – METHODS 93
4.2.1 – invPCR 93
4.2.2 – WGS 95
4.3 – ATTEMPTS AT FINDING TNT1 INSERTIONS WITH INVPCR 96
4.4 – IDENTIFICATION OF TNT1 INSERTIONS BY WGS 97
4.4.1 – Finding the transposons 97
4.4.2 – Expression of target genes 99
4.4.3 – *In silico* zygosity calling for the tnt1 insertions 100
4.4.4 – Confirming the absence of mutation in known MYC genes 103
4.4.5 – Low confidence tnt1 insertions 104
4.4.6 – Distribution of the tnt1 insertions 106
4.4.7 – Genetic differences between *M. truncatula* accessions 110

4.5 – SELECTING LIKELY CANDIDATE GENES 112
4.5.1 – What makes a gene a good candidate? 112
4.5.2 – NF443: likely candidates 114
4.5.3 – NF1436: likely candidates 116
4.5.4 – NF3209: likely candidates 118
4.5.5 – NF3438: likely candidates 120
4.6 – DISCUSSION 122
4.6.1 – Missing insertions from the invPCR dataset 122
4.6.2 – A17 and R108: why so different? 123
4.6.3 – The role of WGS in future forward genetic screens 125
4.6.4 – The role of *Medicago truncatula* in future work 128

Chapter 5 – Identification of candidate genes 130
5.1 – INTRODUCTION 130
5.2 – METHODS 132
5.2.1 – Backcrossing the mutant lines 132
5.2.2 – Phenotyping and genotyping the F_2 133
5.3 – NF443 CO-SEGREGATION ANALYSIS 134
5.3.1 – F_2 Screen (July 2015) 134
5.3.2 – F_2 Screen (February 2016) 138
5.4 – NF3209 CO-SEGREGATION ANALYSIS 142
5.4.1 – F_2 Screen (July 2014) 142
5.4.2 – F_2 Screen (November 2015) 144
5.4.3 – F_2 Screen (June 2016) 146
5.5 – NF3438 CO-SEGREGATION ANALYSIS 148
5.5.1 – F_2 Screen (August 2014) 148
5.5.2 – F_2 Screen (December 2014) 149
5.6 – DISCUSSION 153
5.6.1 – Patterns of colonisation in the F_2 153
5.6.2 – A pair of insertions correlate with the phenotype in NF443 154
5.6.3 – No loci significantly co-segregate in NF3209 155
5.6.4 – A novel allele of a core MYC gene 159

Chapter 6 – A symbiotic kinase required for the AMS in *Medicago truncatula* 161
6.1 – INTRODUCTION 161
6.1.1 – Methods to prove gene/phenotype association 161
6.1.2 – Obtaining independent mutants 162
6.1.3 – Production of complementation lines 163
6.1.4 – Two pronged approach: independent mutation by CRISPR/Cas9 164
6.2 – METHODS 165
6.2.1 – New Noble Foundation Lines 165
6.2.2 – GoldenGate Cloning 165
6.2.3 – Bacterial transformation 167
6.2.4 – Production and assessment of transgenic composite plants 168
6.2.5 – Confirmation of transformation 168
6.2.6 – Inoculation and colonisation 169
6.2.7 – Quantification of gene expression 169
6.3 – RESULTS 170

6.3.1 – New Noble Foundation lines 170
6.3.2 – Generating independent knock-outs 172
6.3.3 – Complementing the mutants 172
6.3.4 – Changes in gene expression in the mutant lines 173
6.4 – DISCUSSION 175
6.4.1 – Gene expression supports the tyrosine kinase, not the GDSL, as the causal insertion 175
6.4.2 – Mt7g116650 has been previously characterised as a mycorrhizal gene 176
6.4.3 – A question of nodulation 178
6.4.4 – KIN3 is a receptor kinase 178
6.4.5 – Phenotypic differences between NF443 and NF5270 180
6.4.6 – RAM1 expression is significantly reduced in NF443 181
6.4.7 – A proposed model of KIN3 function 182
6.4.8 – Proposals for validation or rejection of the model 184

Chapter 7 – Testing the universality of the CSP 188
7.1 – INTRODUCTION 188
7.1.1 – Implementing improvements in agriculture 188
7.1.2 – How common is the common symbiosis pathway? 189
7.1.3 – A broad-scale experimental exploration of plant mutants on fungal genotypes 191
7.2 – METHODS 192
7.2.1 – Summer 2015 Experiment 192
7.2.2 – Summer 2016 Experiment 194
7.2.3 – PCR to identify AM fungal DNA 195
7.3 – RESULTS 196
7.3.1 – Pilot wild inoculum colonisation experiment 196
7.3.2 – Summer 2015 Experiment 197
7.3.3 – Summer 2016 Experiment 204
7.4 – DISCUSSION 209
7.4.1 – Can changes in plant mass tell us anything about AM colonisation 209
7.4.2 – Outcomes of the experiment 210
7.4.3 – Future directions 211
7.4.4 – Effectors and the evidence for mutualism? 211

Chapter 8 – General Discussion 213
8.1 – GOALS AND ACHIEVEMENTS OF THIS WORK 213
8.1.1 – A general summary 213
8.1.2 – Further understanding the role of KIN3 in the AMS 215
8.2 – MANY GAPS REMAIN IN OUR CURRENT KNOWLEDGE OF THE AMS 215
8.2.1 – Locking the parts of the jigsaw together 215
8.2.2 – Invitation or invasion: how mutualistic is the AMS? 216
8.2.3 – Conservation across Embryophyta: to what extent is the AMS universal? 218
8.3 – LOOKING TOWARDS THE FUTURE 219
8.3.1 – How useful are forward genetics to our understanding of the AMS? 219
8.2.2 – AMF into agriculture: what is the best approach? 221

Appendices 224

APPENDIX 1 – PLASMIDS 224
A1.1 – List of all plasmids used 224
A1.2 – Mt7g116510 complementation plasmids 225
A1.3 – Mt7g116650 complementation plasmids 229
APPENDIX 2 – PRIMERS 233
A2.1 – Genotyping Primers 233
A2.2 – GoldenGate Primers 237
A2.3 – qPCR Primers 239
A2.4 – OTU-typing primers 239
APPENDIX 3 – CO-SEGREGATION GRAPHS 240
APPENDIX 4 – MEDIA AND MATERIALS 244
A4.1 – Media, fertilisers and stock solutions 244
A4.2 – Genetic material 246
A4.3 – Growth Conditions 249
APPENDIX 5 – DATA ANALYSIS 250
A5.1 – Linux scripts for analysis of the Illumina WGS dataset 250
A5.2 – Optimisation of *bwa mem* functional parameters for accurate read mapping 250
APPENDIX 6 – DNA SEQUENCES 252
Abbreviations 253
Bibliography 256

List of Figures

Figure 1.1 – Early AM colonisation of *M. truncatula* roots....15
Figure 1.2 – Wild type and mutant *R. irregularis* arbuscules....16
Figure 1.3 – An overview of the common symbiosis pathway....20
Figure 1.4 – Structure of hypothetical fungal signalling molecules detected by the D14-like pathway....23
Figure 1.5 – Four plant signals promote AMF colonisation....26
Figure 1.6 – Structure determines AMF response to flavonoid group molecules....27
Figure 1.7 – Plant and fungal proteins control intercellular fungal colonisation....32
Figure 1.8 – A fatty acid synthesis pathway feeds the fungus....33
Figure 1.9 – Major nutrient exchanges across the periarbuscular matrix....35
Figure 1.10 – Four master switches control the AM symbiosis....37
Figure 1.11 – Carbon acquisition by orchid protocorms from their fungal symbionts ..47
Figure 1.12 – The interaction of plant hormone networks is critical for symbiont and pathogen interactions....55
Figure 2.1 – Triplex PCR to locate tnt1 insertions....65
Figure 3.1 – Colonisation for five mutant lines....72
Figure 3.2 – Normalised colonisation for five mutant lines....73
Figure 3.3 – NF443 colonisation from experiment 3 is not normally distributed 74
Figure 3.4 – Correction for non-normal data....75
Figure 3.5 – Colonised wild type root....76
Figure 3.6 – Structure of arbuscules in the wild type plant....77
Figure 3.7 – Colonised NF443 root....77
Figure 3.8 – Structure of arbuscules in NF443....78
Figure 3.9 – Colonised NF472 root....79
Figure 3.10 – Structure of arbuscules in NF472....79
Figure 3.11 – Colonisation of NF1436....80
Figure 3.12 – Structure of arbuscules in NF1436....80
Figure 3.13 – Colonised NF3209 root....81
Figure 3.14 – Structure of arbuscules in NF3209....81
Figure 3.15 – Colonised NF3438 root....82
Figure 3.16 – Structure of arbuscules in NF3438....82
Figure 3.17 – NF3209 relative colonisation increases with growth time....83
Figure 3.18 – Plant mutants have no effect on growth at 6 wpi....85
Figure 3.19 – Mycorrhizal inoculation has a significantly effect on plant competition in some lines....86
Figure 3.20 – Expression of genetic markers of nutrient exchange in the four mutant lines....88
Figure 4.1 – Visualisation of the invPCR protocol....94
Figure 4.2 – Example of invPCR reaction....97
Figure 4.3 – Fishing for FST with bwa-mem....98
Figure 4.4 – Visual expectation of homo/heterozygous insertions in Tablet....101
Figure 4.5 – Actual Tablet visualisation of homo/heterozygous insertions....102
Figure 4.6 – Some Illumina reads contain multiple discrete sections of the tnt1 sequence....105
Figure 4.7 – Three FST found by the Noble Foundation are not present in our NF1436

seed stock....107
Figure 4.8 – Did WGS discover all tnt1 insertions present in the mutant lines?....110
Figure 4.9 – R108 and A17 genomes show evidence of chromosome rearrangement111
Figure 4.10 – R108 and A17 genomes show evidence of inversion....112
Figure 4.11 – Chromosome map of NF443 showing location of all tnt1 insertions....115
Figure 4.12 – Expression of genes affect by candidate insertions in NF443....116
Figure 4.13 – Chromosome map of NF1436 showing location of all tnt1 insertions...117
Figure 4.14 – Expression of genes affect by candidate insertions in NF1436....118
Figure 4.15 – Chromosome map of NF13209 showing location of all tnt1 insertions.119
Figure 4.16 – Expression of genes affect by candidate insertions in NF3209....120
Figure 4.17 – Chromosome map of NF3438 showing location of all tnt1 insertions...121
Figure 4.18 – Expression of genes affect by candidate insertions in NF3438....122
Figure 4.19 – 40% of insertions in NF3209 missing from the invPCR dataset could not have been found with the restriction enzymes used....123
Figure 5.1 – Backcrossing strategy....131
Figure 5.2 – NF443 high/low colonisation screen....137
Figure 5.3 – Colonisation of two NF443 F_2 populations....139
Figure 5.4 – Co-segregation analysis for NF443 at 5 wpi....140
Figure 5.5 – Co-segregation analysis for NF443 at 7 wpi....141
Figure 5.6 – Colonisation of two NF3209 F_2 populations....143
Figure 5.7 – NF3209 high/low colonisation screen....144
Figure 5.8 – Colonisation of the November 2015 NF3209 screen control plants....145
Figure 5.9 – Colonisation of the June 2016 NF3209 F_2 population....146
Figure 5.10 – Colonisation of the August 2014 NF3438 F_2 population....148
Figure 5.11 – Colonisation of the December 2014 NF3438 F_2 population....150
Figure 5.12 – NF3438 high/low colonisation screen (part 1)....151
Figure 5.13 – NF3438 high/low colonisation screen (part 2)....152
Figure 5.14 – Co-segregation analysis for NF3438....152
Figure 5.15 – NF3209 chromosome map, showing locations of insertions and future regions of interest....158
Figure 5.16 – NF3438 is a novel allele of *ram1*....160
Figure 6.1 – Split plant roots expressing DsRed....169
Figure 6.2 – NF15212 plants did not contain a tnt1 insertion near Mt7g116650....170
Figure 6.3 – NF19713 plants did not contain a tnt1 insertion near Mt7g116510....171
Figure 6.4 – tnt1-related genotype of NF2901 at locus Mt7g446190 is unclear....172
Figure 6.5 – Gene expression in NF443 and NF3438....174
Figure 6.6 – Gene expression in NF443 and NF3438....174
Figure 6.7 – Relative gene expression of NF443 candidates....176
Figure 6.8 – Structure of KIN3....179
Figure 6.9 – Genomic structure of Mt7g116650 in NF443 and NF5270....181
Figure 6.10 – Proposed role of KIN3 in maintain RAM1 expression during the AMS...184
Figure 7.1 – Diverse arrays of signalling molecules and/or effectors maintain generalism....190
Figure 7.2 – Location of wild inoculum sampling sites (2015) on Walmgate Stray, York 193
Figure 7.3 – Location of wild inoculum sampling sites (2016) on Walmgate Stray, York 194
Figure 7.4 – Colonisation of WT test plants 6 wpi with wild inoculum....197
Figure 7.5 – Fresh weight of WT test plants 6 wpi with wild inoculum....197

Figure 7.6 – Whole plant fresh weight by treatment; wild inoculum 2015 199
Figure 7.7 – Plant fresh weight by genotype; single pots; wild inoculum 2015 201
Figure 7.8 – Shoot dry weight by genotype; single pots; wild inoculum 2015 201
Figure 7.9 – Plant fresh weight by genotype; mixed pots; wild inoculum 2015 202
Figure 7.10 – Pot effect; wild inoculum 2015 203
Figure 7.11 – Shoot dry weight by genotype; mixed pots; wild inoculum 2015 204
Figure 7.12 – Plant fresh weight by genotype; wild inoculum 2016 207
Figure 7.13 – Plant dry weight by genotype; wild inoculum 2016 207
Figure 7.14 – Detection of AMF rDNA10 in wild type plants 209
Figure 7.15 – Detection of AMF rDNA10 in mutant plants 209
Figure 8.1 – Knowledge of the AMS tends to collect into discrete modules 216
Figure 9.0.1 – LMP1_51 227
Figure 9.0.2 – LMP1_51Y 228
Figure 9.0.3 – LMP1_51YM 229
Figure 9.0.4 – LMP1_65 230
Figure 9.0.5 – LMP1_65Y 231
Figure 9.0.6 – LMP1_65YM 232
Figure 9.0.7 – NF3209 insertions that did not co-segregate with MYC phenotype (discovered by invPCR) 240
Figure 9.0.8 – NF3209 insertions that did not co-segregate with MYC phenotype (discovered by WGS) 243

List of Tables

Table 1.1 – Symbiotic gene homologs across four important plant models. 19
Table 4.1 – Estimated coverage of the *Medicago* genomes sequenced 95
Table 4.2 – Insertions found in 7 mutant lines with invPCR 96
Table 4.3 – Summary of tnt1 insertions found by WGS in the four mutant lines 99
Table 4.4 – Expression of genes affected by tnt1 insertions 100
Table 4.5 – List of known mycorrhizal genes 104
Table 4.6 – Locating insertions from the R108 scaffold assembly 106
Table 4.7 – Did WGS discover all tnt1 insertions present in the mutant lines? 109
Table 4.8 – Promising candidate insertions for causing the mycorrhizal phenotype in NF443 114
Table 4.9 – Promising candidate insertions for causing the mycorrhizal phenotype in NF1436 116
Table 4.10 – Promising candidate insertions for causing the mycorrhizal phenotype in NF3209 118
Table 4.11 – Promising candidate insertions for causing the mycorrhizal phenotype in N3438 120
Table 5.1 – Description of the seven F_2 screen performed 134
Table 5.2 – NF3209 insertions located by invPCR 143
Table 5.3 – NF3438 insertions located by invPCR 149
Table 7.1 – Lines used in wild inoculum experiment (Summer 2015) 193
Table 7.2 – Lines used in wild inoculum experiment (Summer 2016) 195
Table 7.3 – Fertilisation regime for the wild inoculum experiment (Summer 2016) 195
Table 9.1 – Addgene plasmids for complementation 224
Table 9.2 – PCR amplicons for complementation 224
Table 9.3 – Plasmids produced for complementation 225
Table 9.4 – NF443 genotyping primers 233
Table 9.5 – NF3209 genotyping primers 234
Table 9.6 – NF3438 genotyping primers 236
Table 9.7 – New NF line genotyping primers 237
Table 9.8 – Primers for LightRun sequencing 238
Table 9.9 – Primers for producing the GoldenGate cassettes 238
Table 9.10 – Primers for *Medicago* MYC gene qPCR 239
Table 9.11 – Primers for OTU typing primers Glomeromycotina 239
Table 9.12 – Solid growth media 244
Table 9.13 – Fertilisers, liquid media and stock solutions 245
Table 9.14 – Mixed solutions for experiments 246
Table 9.15 – *M. truncatula* ecotypes and mutants used in this project 247
Table 9.16 – Glomeromycotina lines used in this project 248
Table 9.17 – Bacterial strains used in this project 248
Table 9.18 – Concentrations of antibiotics used in this project 248
Table 9.19 – Growth conditions used in this project 249
Table 9.20 – DNA sequences used in this project 252

Acknowledgements

I would to thank my supervisors, Michael Schultze and Peter Young, for giving me the opportunity to undertake a doctorate, and for all their help and teaching during the past four years. Further thanks to Michael for all his help proof-reading this thesis, and putting up with my terrible spelling. I offer my thanks to the members of my thesis advisory panel, Thorunn Helgason, Mike Haydon and Daniela Barilla, for all their help and advice on the writing and practical work during this project.

The investigators as a whole are indebted to the BBSRC for funding, and to our collaborators for material without which the project could not have happened; to Samuel Roberts Noble Foundation for producing and curating the tnt1-mutagenised *Medicago* lines, to Jeremy Murray for providing us with *Rhizophagus* hairy root cultures and pre-public access to the *Medicago* R108 genome, and to PlantWorks Ltd for providing mycorrhizal inocula.

I would also like to thank Jeremy, along with Frans Maathuis and Uta Paszkowski, for their helpful advice on a variety of topics, as well as the other members of the L2 and D0 lab groups, for their help and camaraderie, particularly thanking Thomas Hartley for his most essential randomisation script

Finally, I would like to thank the numerous undergraduate students whose project work contributed to this thesis, in particular Tasmin Spurgeon for her work producing CRISPR/Cas9 knock out plasmids for *Medicago* transformation (see **Chapter 6**), and also to Leigh Betts, Laura Bouvet, Sigurd Harborough, Louisa Keith and Shiraz Ziya for their prior work producing invPCR amplicons to help locate tnt1 insertions in the various mutant lines (see **Section 4.3**).

Author's declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been published or presented for an award at this, or any other, University. Some parts of this work (primarily **Chapters 3** and **6**) have been presented in posters or oral presentations at conferences.

All parts of the work are my own, excepting that specifically labelled as the work of six undergraduate students undertaking final year projects in the Schultze lab. These collaborators were Leigh Betts, Laura Bouvet, Sigurd Harborough, Louisa Keith and Shiraz Ziya (for work presented in **Section 4.3**, then used for the authors own analysis in **Section 5.4.1** and **5.5**) and Tasmin Spurgeon (for her part in an ongoing project described in **Section 6.1.4** and **6.3.2**)).

All sources are acknowledged as references.

Chapter 1 – General Introduction

1.1 – What is the arbuscular mycorrhizal symbiosis?

The arbuscular mycorrhizal symbiosis (AMS) is a mutualistic relationship between plant roots and filamentous fungi of the subphylum Glomeromycotina[1] (**Figure 1.1**), which occurs in 80% of extant land plants[2]. The symbiosis is a major source of phosphate and other nutrients for the plant host in most natural ecosystems, and also provides the host with resistance to a wide spectrum of biotic and abiotic stresses. In return, the obligate fungal symbiont receives photosynthate from the host, its only source of fixed carbon. The AMS is characterised by the arbuscule, an elaborate branched fungal structure that forms inside the cells of the root cortex, and which is the main site of nutrient exchange (**Figure 1.2**). This relationship predates the colonisation of land, with genetic data suggesting an association as early as 1 billion years ago[3,4]. The AMS was likely responsible for the development of modern plants, forming primitive 'roots' that allowed the first terrestrial colonists to exploit their new substrate.

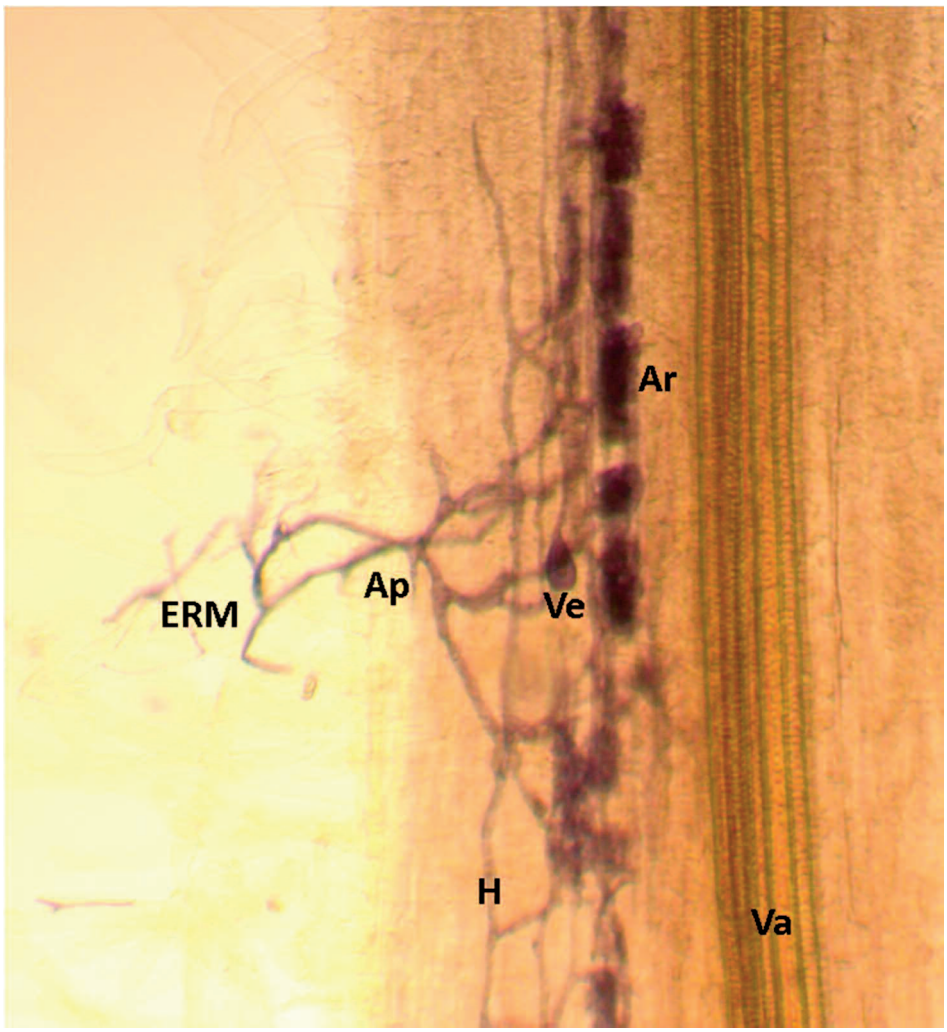


Figure 1.1 – Early AM colonisation of *M. truncatula* roots
Light field microscope image of early root colonisation of *Medicago truncatula* roots (yellow) by *Rhizophagus irregularis* (blue). External hyphae (ERM) emerging from a fungal spore (not pictured) form an appressorium (Ap) to penetrate the root, spreading hyphae laterally through the outer root tissues between the cells (H). Once inside the cortex, adjacent to the vasculature (Va), hyphal elaborations penetrate the cell wall of cortical cells, forming a membrane separated compartment (arbuscules, Ar) inside the cell. The arbuscules rapidly branch to increase the surface area for nutrient exchange, and last for around 9 days before collapsing and being broken down by the host. After this, the cell can be re-invaded and a new arbuscule forms. Vesicles (Ve), large bulbous globular structures, also form between the cells of the root and act as nutrient stores for the fungus.

This association does not take place in isolation. The soil is filled with a wide array of micro-organisms, including many varieties of predators, pathogens and other competing symbionts. Both partners face challenges to unlocking the benefits of the symbiosis. The fungus, only able to obtain carbon from its plant host, must be able to survive the periodic die backs of vegetation that characterise many ecosystems, forming hardened spores that can lie dormant for years[5]. These spores must then know when to emerge, and how to locate the roots of the new host. The plant is constantly challenged by fungi seeking to enter the root, so should seek to

differentiate between them, to allow access to symbionts and defend against pathogens. Both partners have been predicted to sanction 'lazy' partners to enforce mutualism[6], although this is far from proven, and other models of the symbiosis have the plant playing a passive role[7,8], acting more commensally with an surplus supply of photosynthetic carbon, or even modelling the AMF as a very successfully pathogen, that supplies the host with its own surplus nutrients to maintain the host (and thus its own propagation).

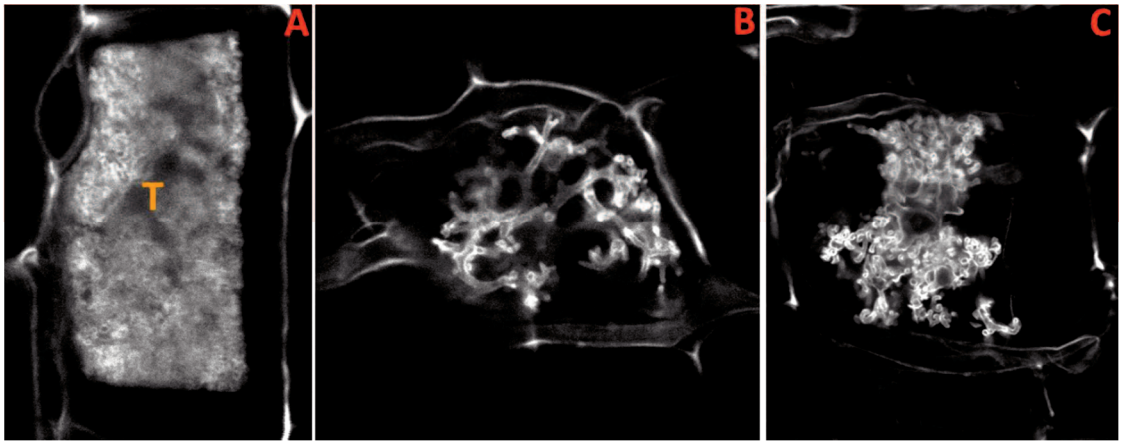


Figure 1.2 – Wild type and mutant *R. irregularis* arbuscules
Confocal microscope images of mature wild type (A) and mutant (B, C) arbuscules stained with fungal cell wall binding WGA-AF488. Observe the faintly staining plant cortex cell wall surrounding the fungal structure, and the core trunk of the mature arbuscule (T) surrounded by the progressively fine branching that gives the arbuscule its name (derived from the Latin *arbuscula*, lit. small tree). Mutant arbuscule B has ceased development at the 'bird's foot' stage, composed only of the first round of branching, while C exhibits signs of degeneration.

The key, then, is communication. The plant and fungus signal to each other, exchanging messages that allow germinated fungal spores to home in on the roots of the potential host. In turn, these fungal hyphae release signals to prime the plant to form the pre-penetration apparatus that allows fungal access to the root, and to damp down its immune system. Once inside the roots, this signal exchange continues, preventing the activation of a defence response provoked by microbe associated molecular patterns (MAMPs) from the fungus[9]. Both partners attempt to regulate the symbiosis to ensure they benefit. The plant excludes the fungus from vulnerable tissues[10] and diverts carbon to the most productive AMF colonies[6]. The fungal side of the partnership is less understood, but they control their own resource delivery to the host[11] and manipulate the host with effectors[12].

1.2 – Why the AMS matters

Fungal associations are an omnipresent part of the flora of all but the most extreme ecosystems. AMF associate with 80% of plant species, and most of the remainder form other associations with the Dikarya mycorrhizae (ecto-, ericoid, and orchid). They provide benefits to their hosts both large (contributing up to 90% of plant phosphate uptake[13]) and wide (tolerating plants to stresses including drought, temperature, metal toxicity, pathogen and herbivore attack as well as delivering other nutrients)[14–18]. The common mycorrhizal networks (CMN) of many different AMF species link plants throughout their ecosystem, forming a conduit for movement of carbon, mineral nutrients and signalling molecules[19–21]. Even in intensively farmed agricultural land, where tilling breaks down the CMN and NPK fertilisation suppresses colonisation, mycorrhizal fungi remain in the soil[22].

Currently, world agriculture is facing many serious challenges. We need to maintain and increase crop yields in line with population growth and increasing standard of living, and to do so in the face of increasing desertification, salinization and constraints on availability of water and fertilisers. Current high quality (i.e. low arsenic content) rock phosphate reserves are dwindling, with the majority of reserves outside of Morocco expected to be depleted by 2100[23], and with all currently exploitable reserves predicted to last around 300 years at current consumption rates[24]. This will drive up input costs, and eventually impact crop yields as soils become P limited[25]. A way to improve crop phosphorus use efficiency would reduce the need for heavy fertilisation, stretching reserves and limiting eutrophication from agricultural run-off.

Supplementation of agricultural land with AMF inocula is challenging given the scale necessary, compounded by our inability to axenically cultivate AMF, but is showing promising results[26]. However, the effects of mycorrhizal colonisation are heavily dependent on both plant and fungal genotypes, and the soil condition. Any single fungal inoculum is unlikely to benefit all areas or all crops. Deployment of such technology to the southern hemisphere, where efficient delivery and application of standard fertilisers is already difficult, is even more of a challenge. However, it is also a huge opportunity to improve the yield and nutrition of crops for subsistence farmers, and improve their resilience against droughts.

While re-supplementing the fungal stock of intensively farmed areas may give benefits, given the current inability to mass-culture or genetically modify the fungus, we believe that the best approach overall is breeding or genetically engineering plants to improve the efficiency of the symbiosis. New plant lines would be easier to centrally test and tailor to specific soil/fungal conditions, and could be distributed using existing supply lines. Understanding the genetic mechanisms underlying plant colonisation will be important to a targeted breeding program, as will a good understanding of the crops, soil biota and environmental conditions of the world's agricultural systems.

In short, in high-input agriculture the symbiosis offers to reduce costs, and in low-input agriculture to increase yields.

1.3 – The common symbiosis signalling pathway

1.3.1 – How rhizobia have advanced mycorrhizal research

The common symbiosis pathway (CSP) arose at the dawn of land plants over 400 MYA[27–29]. Its origins are uncertain, but appear to re-purpose immunity-related[9] and drought response genes[30,31] to form the original symbiosis between primordial plants and the Mucoromycotina[32]. This system was quickly expanded to include the AMS with Glomeromycotina, and then further expanded around 60 MYA to facilitate the root nodule symbiosis (RNS) between legumes and rhizobia. Recent work suggests that it is also involved in ectomycorrhizal[33] and actinorhizal[34] associations as well. While this is not the universal pathway for all root-microbe beneficial interactions[35], it is clearly hugely important to the plant, and provides a common basis for understanding the willing downregulation of defence and large scale cellular rearrangements necessary to accommodate intercellular symbionts.

The RNS is probably the best understood of the plant root symbioses at the molecular level, thanks to its importance in agricultural practice, and the comparative ease of working with the system than with the AMS. Rhizobia, principally *Sinorhizobium meliloti* (symbiont of *Medicago*) and *Mesorhizobium loti* (symbiont of *Lotus*), can be cultured axenically (unlike all known AMF), making maintaining strains in culture far

simpler. Simpler bacterial genomes (the first *Rhizobium* genome was published 12 years before that of the first AMF[36,37]) and the ability to produce stable transformed bacteria strains make it possible to investigate interactions between host genome and symbiont genomes that remain impossible in the AMS. Finally, nodules can be observed with the naked eye on the root surface, whereas accurate accounting for mycorrhizal colonisation requires destructive staining or nucleic acid or lipid extraction. This makes quantification and mutant screening much easier, aided by the relatively lower variation in colonisation seen in the RNS. Use of proxies, primarily visualisation of Ca^{2+} spiking, has helped investigation of the AMS, but caution must be taken with the resulting data as certain mycorrhizal responses are independent of this response (e.g. AM induction of lateral root branching[38]).

One problem found in the literature surrounding the symbioses is that genes often have different names to their homologs in other species. In this thesis, we will be using the gene names found in our model organism, *Medicago truncatula*. To ease understanding, we provide **Table 1.1**, describing the names of homologs for common symbiotic genes across the major model species.

Table 1.1 – Symbiotic gene orthologs across four important plant models.

Oryza sativa	***Arabidopsis thaliana***	***Medicago truncatula***	***Lotus japonicas***	**Function**
SLR1	GAI, RGA, RGL1 & 2	DELLA1 & 2	DELLA1 & 2	DNA binding negative regulator of GA signalling
Duplicated into **CASTOR** & **POLLUX**	POLLUX	DMI1	Duplicated into **CASTOR** & **POLLUX**	Potassium ion channel
DMI2	*missing*	DMI2 *or* NORK	SymRK	LCO receptor co-factor
DMI3 *or* CCaMK	*missing*	DMI3	CCaMK	Calcium/calmodulin binding kinase
CYCLOPS	*missing*	IPD3	CYCLOPS	Transcription factor
CERK1	LYK1	LYK3	NFR1	Symbiotic and pathogenic CO binding; LCO binding
CeBIP	LYM2	LYM2	unknown	Pathogenic CO binding
unknown	*missing*	LYR3	unknown	Myc-LCO binding
RLK2	*missing*	NFP	NFR5	Nod-LCO binding

Abbreviations: lipochitooligosaccharides (LCO), chitin oligomers (CO), mycorrhizal-LCO (Myc-LCO), rhizobia-LCO (Nod-LCO)

1.3.2 – The common symbiosis pathway

The CSP (**Figure 1.3**) is a central plant signalling mechanism that detects signalling molecules released from a symbiont at the cell-surface, passes this information to the nucleus and triggers specific gene expression changes to induce accommodation of that symbiont. In the legumes, where this pathway has been characterised, it was neofunctionalised to act in the RNS as well as retaining its role in the AMS.

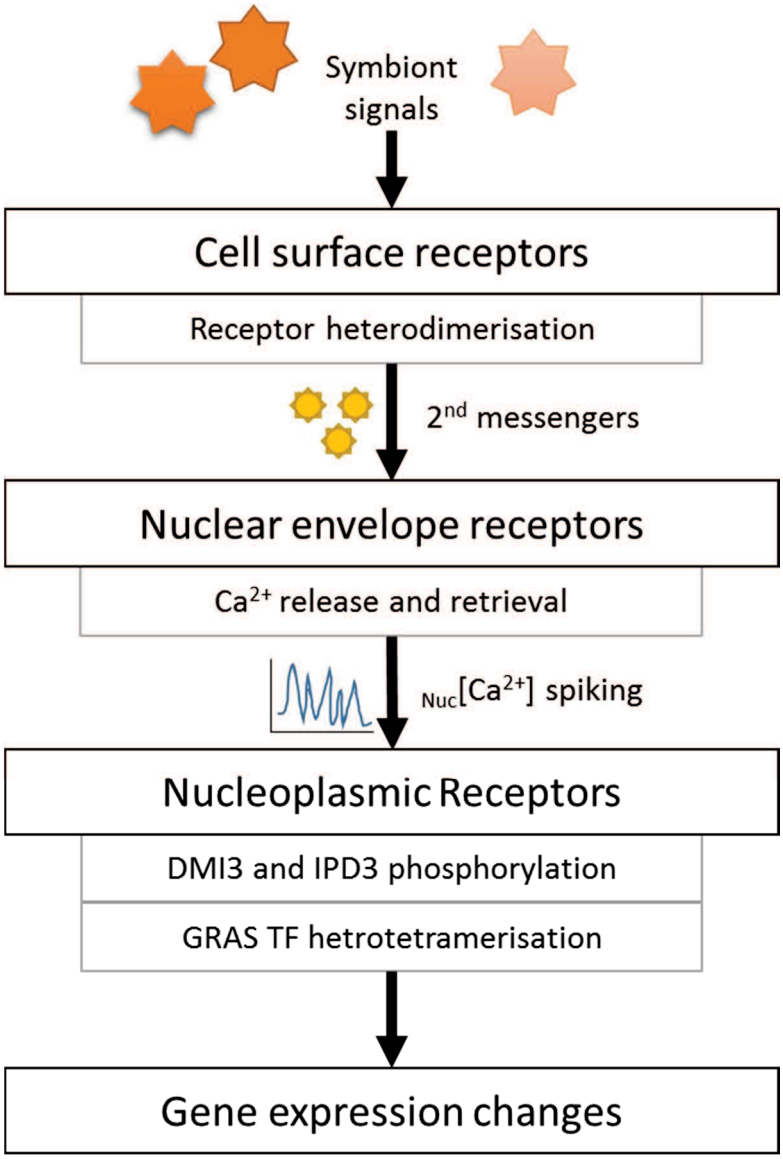


Figure 1.3 – An overview of the common symbiosis pathway

The CSP begins with the detection of chitin-based signals, referred to as Nod factors (in the RNS) and Myc factors (in the AMS). Fungal signals are a mix of lipochitooligosaccharides (LCO)[39] and short (4-5)[40] and long (6-9)[41] chain chitin oligomers (CO), discussed in more detail in **Section 1.4.2**. Rhizobia release only LCO, 3

to 6 N-acetylglucosaminyl residues with a fatty acid tail attached to the first residue via an amide bond, and further modification by N-acetylation or addition of carbon or sulphate ester groups[42]. Plant interactions with the rhizobia, unlike those with Glomeromycotina, are highly specific, and it is these decorations on the LCO that are thought to be responsible for the species specificity in legume/rhizobia interactions.

The CSP begins at the cell surface with Nod/Myc factors binding to the LysM domains of an integral membrane receptor (NFP/LYR3) with an inactive kinase domain[43–45]. This receptor then forms a complex with DMI2, another receptor-like kinase which lacks an extracellular LysM domain but carries an active kinase domain[46]. Since the RNS arose in the ancestor of the modern legumes, this system has been elaborated upon, partially in the Fabaceae, a subfamily that includes both model legumes *Medicago truncatula* and *Lotus japonicas*, as well as most of the agriculturally important legumes. In the Fabaceae, *NFP* has been duplicated, and another LysM receptor kinase, *LYK3*, has also been recruited to the symbiosis[47,48]. This has allowed the co-evolutionary specialisation of these receptors to recognise the species specific Nod-LCOs produced by Rhizobia. Another complication is the apparent recruitment of LNP, a root surface apyrase that is necessary for the function of the CSP in legumes for both the RNS and AMS, and shown to bind LCO[49].

The next step appears to be conserved across the different symbiosis using the CSP. The LysM receptor-DMI2 complex binds the scaffold-like remorin protein MYCREM (duplicated in legumes for the RNS-specific SYMREM), inhibiting the MAPKK SIP2 and promoting the activity of HMGR1, which catalyses the production of mevalonic acid[50–53]. These activities presumably start a second messenger chain that is picked up by the symbiotic nuclear envelope complex, although the exact mechanism of this step remains opaque. However the signal reaches the nuclear membrane, once there it induces the nuclear calcium spiking response that has come to define the CSP. This is triggered by cycles of calcium release from the perinuclear space by three isoforms (a-c) of CNGC15[54]. In direct contact with, and simultaneously activated by, CNGC15 is the potassium channel DMI1, which acts to balance the charge differential created by the calcium movement. The Ca^{2+} ATPase MCA8 creates the down-stroke of the calcium spiking, retrieving calcium from the nucleoplasm[55,56]. Both these proteins are essential

for calcium spiking, as is the nonspecific cation channel AnnMt1 and three components of the outer ring of the nucleoporin complex (NENA, NUP85 and NUP133), the latter presumably required for the localisation of the other integral membrane proteins[57–60].

Inside the nucleus the calcium/calmodulin dependent kinase DMI3 transduces this calcium spiking signal. At resting calcium concentrations, calcium binds the tetrad of EF hand motifs on DMI3, allowing autophosphorylation of the protein. High calcium concentrations cause a calmodulin protein to bind to DMI3, which blocks the autophosphorylation activity. Once now unstable the phosphate group undergoes hydrolysis, the repression of the DMI3 kinase is freed, and DMI3 phosphorylates IPD3[61,62]. In the RNS, this drives expression of the GRAS transcription factors NSP1 and NSP2, which form a heterotetramer to start the genetic reprogramming required for the RNS[63]. In the AMS, GRAS transcription factor (TF) RAM1 is the central regulator[64].

1.3.3 – Differences in the CSP between the symbioses

The CSP is well understood, but already we see differences between the RNS and AMS. Different receptors perceive different (albeit related) signalling molecules, then associate with DMI2 via different scaffolding proteins (MYCREM & SYMREM), and most importantly, produce different downstream effects. How does the plant take the same signals, mevalonic acid and Ca^{2+} spiking, and begin different programs of gene expression to build the radically different terminal structures of the RNS and AMS?

The first possible explanation for this is spatial separation. As a general rule in the Fabaceae, rhizobia enter the root via trichoblasts, whereas AMF form appressoria on the surface of atrichoblast cells. This could allow the CSP to use the same signal cascade to produce different gene expression responses by integrating a cell identity marker. However, AMF spore exudates induce Ca^{2+} spiking in trichoblasts, and Nod-factors induce Ca^{2+} spiking in atrichoblasts in legumes[65]. This suggests both cell types are at least partially responsive to either signal, regardless of where the symbiotic contact point forms. In addition, rice (which does not form the RNS) also exhibits Ca^{2+} spiking in trichoblasts[65]. This suggests that the CSP was not recruited to root hairs by the RNS, but was capable of functioning there prior to the evolution of the RNS. What

a rice trichoblast does on perception of Myc factors is unknown. Since it responds only to CO4, not LCO, this is probably not a part of the root hair induction pathway (which is primarily driven by LCO).

A second possible explanation is that non-LCO signals (see **Section 1.4**) trigger a separate, AMS-specific pathway. Under this hypothesis, LCO perception would induce the RNS, and LCO + CO4 perception would induce the AMS. This also seems unlikely, as responses to both LCO and CO4 are blocked by the same *DMI* mutants, implying they use the same pathway. There is also no clear and repeatable difference in the characteristics of the Ca^{2+} spiking response induced by Myc and Nod factors that would support two signalling pathways using some of the same components. The obvious solution for this would be a secondary pathway functioning alongside the CSP. Possible candidates for such a role have been found by mutant screens in the monocots, including the cytoplasmic karrikin receptor D14L, which would induce the ubiquitination of unknown targets by D3 on ligand binding[66], and NOPE1, a GlcNAc transporter[67]. These genes are both required for AM colonisation in rice and maize. While it is tempting to assign these into the same pathway (with NOPE1 importing a signal for D14L to perceive), this seems unlikely given the differences in structure between the two molecules (**Figure 1.4**). Additionally, nurse plants separated by AMF-proof mesh rescue the *nope1* phenotype, and *nope1* root exudates produce different AMF gene expression responses compared to wild type exudates[67], which supports a role for NOPE1 as an exporter of an unknown signal to the fungus.

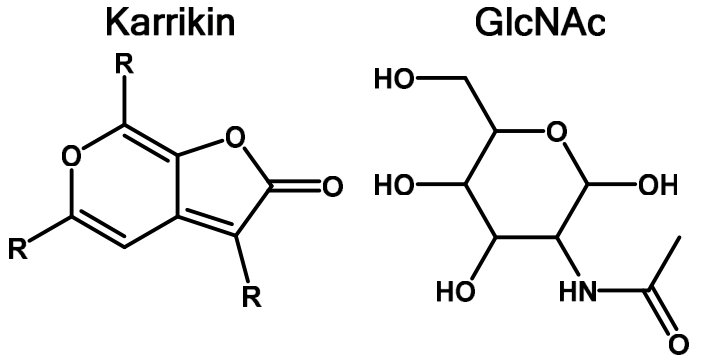


Figure 1.4 – Structure of hypothetical fungal signalling molecules detected by the D14-like pathway
Structures of smoke derived signalling molecule karrikin (R groups are either hydrogen or methyl groups) and GlcNAc, a component of chitin.

In both the AMS and RNS the DELLA-DMI3-IPD3 complex[64] drives expression of a complex of GRAS transcription factors (NSP1/NSP2 in the RNS, DIP1/RAM1/NSP2/NSP1 in the AMS)[68–71]. Both these responses involve NSP1/2, and the role of DIP1 appears to be to link the complex into GA signalling. Thus, it is tempting to label RAM1[72] as a central regulator of mycorrhizal identity, which controls which set of downstream transcription factors (including *CBX1* & *MYB1* in the AMS[73,74] and *NIN* & *ERN1* in the RNS[75]), and thus which specific symbiotic structure is formed. This would be simple enough to test, as it would imply a plant expressing a proNSP2::RAM1 construct would prevent nodules forming by forcing an AMS-like identity onto cells. Regardless, there must be a second signal, either simultaneously acting on the promoters of the GRAS TF or downstream of them, to maintain signal identity aside from Ca^{2+} spiking.

1.3.4 – After the CSP

Once we leave the CSP behind, our understanding of the AMS becomes distinctly patchier, with genes known to be involved in the symbiosis, but largely unconnected to each other. This includes processes like pre-contact signalling, facilitating fungal access to the root, and regulation of nutrient exchange between partners. These genes are likely to be key to any attempt to understand and improve the symbiosis in the field. This project was initiated to locate more of these genes, and try to fill in the gaps in the network of this fungal-only signalling pathway (FSP). Recent years have seen an explosion in our knowledge of plant-fungal signalling, and in the nature and regulation of nutrient exchange at the arbuscule, but many gaps still remain.

1.4 – Picking up the phone: pre-contact communication to establish symbiosis

1.4.1 – Plant signals to the fungi

Plant roots constantly exude a vast range of compounds into the rhizosphere, both deliberately (e.g. the mucus sheath to protect the root tip, defensive secondary metabolites, organic acids and chelating agents to increase nutrient uptake) and ‘accidentally’, as cells rupture and die, leaking their cell contents. The AMF, like all plant interacting microbes, home in on this mixture of deliberate signals and chemical noise from the growing root.

Glomeromycotina spores, large and multinucleate, are the resting, desiccation resistant stage of the AMF's lifecycle. Their germination is increased in the presence of plant root exudates, but axenic germination is both possible and now routine in labs[76,77]. Axenic germination requires damp conditions and favours low to moderate substrate [P]. Temperature and pH also affect germination, although different strains would likely show differing preferences. AMF spores may also produce self-inhibition factors, which require breakdown by soil bacteria or absorption by the substrate to allow germination, but the evidence of this is conflicting, and this may also be strain dependent[76,78]. However, without plant exudates to home in on, the initial hyphae emerging from the germinating spore are arrested after around 10 days[79], re-entering dormancy and re-germinating at a later date. While a lot of work looks at the perspective of spores in the soil, this paradigm is not universal. Some ecosystems (e.g. agricultural fields, those with strong dry seasons or regular fires) see plants regularly re-colonising a soil bare of plant growth (and thus requiring AMF to periodically go dormant). However, many other ecosystems are consistently productive (e.g. rainforests). In these settings, newly germinated radicals descend into a world filled by networks of hyphae from many different AMF species that could persist indefinitely without the need to sporulate, as they will always have active partners. Either way, hyphae emerging from a newly germinating spore, or spreading from a mature extra-radical mycelial (ERM) network, will encounter a rainbow-like series of compounds that it uses to find the root. This pattern is formed by different abilities of these compounds to spread through the soil, dependent on their stability and rate of diffusion (**Figure 1.5**).

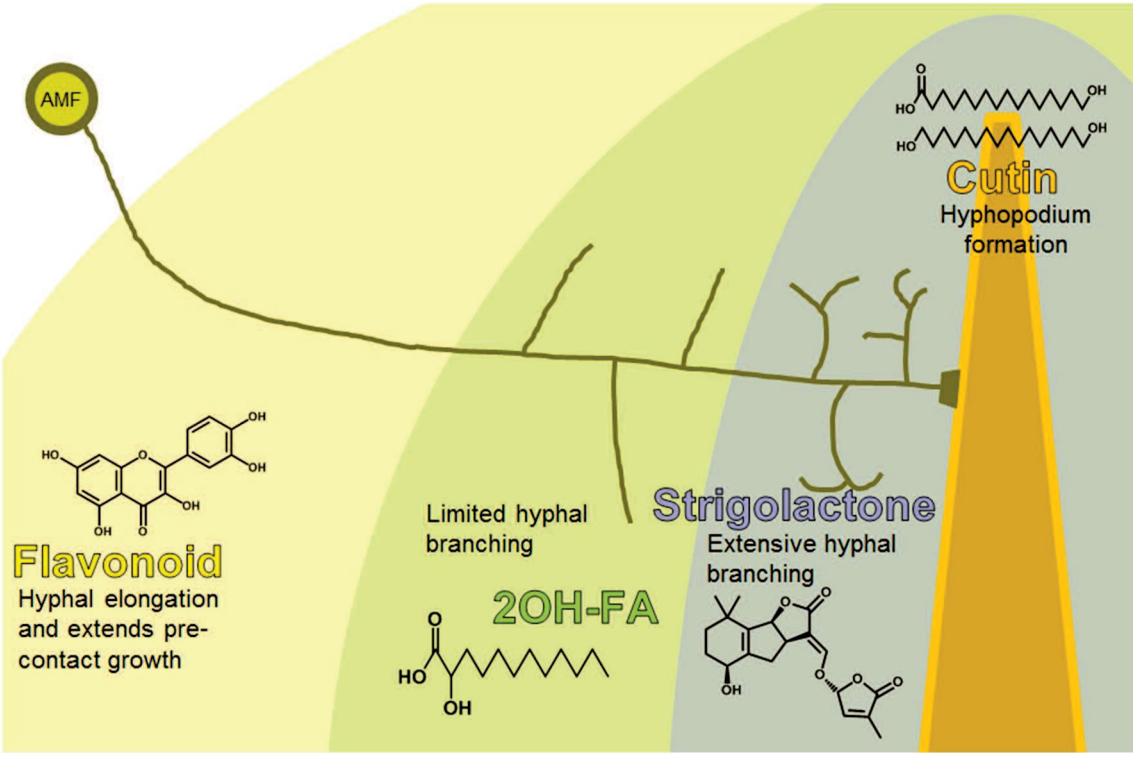


Figure 1.5 – Four plant signals promote AMF colonisation
The four known groups of compounds from plant root exudates known to promote colonisation by AMF, displaying the generalised chemical structure and arranged by the predicted distance from the root the compounds would function at.

The most mobile of these are from the flavonoid family of secondary metabolites. These, particularly quercetin and isovitexin 2''-O-β-glucoside, increase both the speed and duration of pre-contact hyphal growth, and leading to an increase in colonisation of the host plant[79,80]. These effects are not seen with all flavonoid-type compounds, and many have inhibitory effects on AMF growth[81], although there are currently no known mechanisms or structural requirements governing this effect (**Figure 1.6**). Flavonoids increase the chance of the fungus finding the root, but are otherwise not required for colonisation[82]. Less mobile, but still quite soil stable are the 2-hydroxy fatty acids (2OH-FA). Nagahashi & Douds (2011)[83], showed that C12 and C14 2OH-FA induced lateral branching of growing *Gigaspora gigantea* hyphae. In the rhizosphere, this branching would improve the chance of the fungus finding the root of the potential host. This response was not seen with the 3-hydroxy fatty acids, which had a mildly inhibitory effect on hyphal growth, nor was the branching response seen with longer or shorter chain 2OH-FA. Interestingly, this includes the C16 fatty acids that the fungus depends on the host for[84–86]. The response to the different 2OH-FA also differed somewhat between AMF species, which might provide the potential for selection of specific hosts seen in a minority of AM species[87]. We lack evidence for

deliberate fatty acid release, but synthesis of pro-AMF flavonoids is upregulated under phosphate stress, and AMF-inhibitory flavonoids make up the majority of those secreted by non-host plant species, suggesting deliberate modifications to the secretome may help a plant control its symbiosis[80,81].

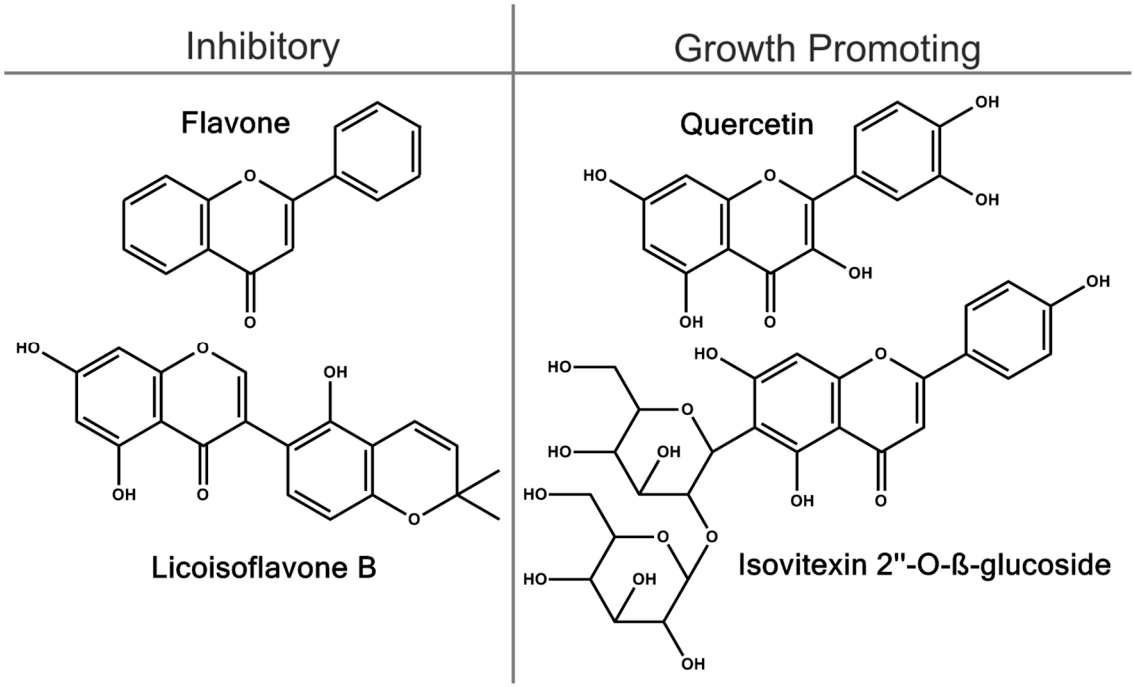


Figure 1.6 – Structure determines AMF response to flavonoid group molecules
Chemical structure of flavonoid compounds shown to promote AMF hyphal extension and growth period, or inhibit hyphal growth (flavone) or branching (licoisoflavone B)

Strigolactones are plant hormones that are synthesised in the root plastids, and regulate lateral root and stem branching[88]. Biosynthesis is strongly upregulated by P starvation, and strigolactones are selectively exported into the rhizosphere by the ABC transporter PDR1[89,90]. The labile ether bond of strigolactone is quickly hydrolysed in the rhizosphere, limiting its spread through the soil, but providing a very steep concentration gradient that many different organisms (e.g. parasitic plants of the genus *Striga*) use to chemotactically locate the root[91]. In AMF, strigolactones induce extensive ramification of the growing hyphae, as well as providing a chemotactic gradient[92]. While the role of strigolactone in controlling plant branching appears to be basal (with strigolactones acting in this role in the pre-AM macroalgae of the order Charales), the function of PDR1 and the strong phenotype of *pdr1* mutants points to neofunctionalisation of the hormone to mycorrhizal signalling early in the history of the symbiosis[88]. As with flavonoid deficient mutants, strigolactone deficient mutants retain normal internal colonisation, indicating that these compounds have a purely

chemoattractive role in colonisation, and are dispensable once surface contact has been achieved[89].

Finally, upon perception of fungal signals, the plant upregulates RAM2, a GPAT enzyme that produces a 20-30% increase in 16 carbon cutin monomers (C16:0 FA & 1,16-hexadecanediol) around the root surface[93]. This stimulates the fungus to form appressoria on the root surface and begin the process of root entry. As with strigolactone, the system has been hijacked by pathogenic fungi (*Magnaporthe grisea*) and oomycetes (*Phytophthora palmivora*), which also require a RAM2 dependent increase in cutin monomers to invade the root[93], presumably achieved via effectors.

1.4.2 – Fungal signals to the plant

The signals released by the AM fungus are comparatively less well understood. Work with germinated spore extracts (GSE) and various pure molecules show that a number of different compounds are responsible for triggering plant responses. Those identified so far fall into three groups, all chitin-derivatives with various levels of modification. These groups are lipochitooligosaccharides (LCO) and chitin oligomers of short (3-5 residues, referred to as CO4s) and long (6-9 residues, called CO8s) chain lengths. It is still a point of contention if the fungus is releasing the 'active' molecules for signalling, or if these are created by enzymatic breakdown of fungal material. Further complicating the situation are conflicting reports of detection via cleavage, and the fact that several pro-colonisation signals are also drivers of immune responses in pathogenic spore extracts[9,94].

The first group of these chitin-based signalling molecules, and the earliest isolated, are the mycorrhizal lipochitooligosaccharides (Myc-LCOs). These are similar in structure to the LCO used in pre-contact signalling for the nodule symbiosis, and were originally thought to be the evolutionary precursors of such bacterial signals. Recent research, however, has suggested that the rhizobial genes required for LCO synthesis are not fungal in origin, and even that LCO perception in the AMS may be unique to the legume clade, not widespread as had been initially thought[41]. In structure, Myc-LCO have different modifications than Nod-LCOs, and illicit the strongest plant response when in a mix of sulphated and non-sulphated forms[39], rather than being primarily

sulphated like the Nod-LCOs of specialist rhizobia species[42]. Nod-LCO are a driver of specific symbiont-host interactions, and different plant hosts produce Ca^{2+} responses to specific Nod-LCO decorations, each of which corresponds to different bacterial operational taxonomic units (OTU). The GSE from different AMF species, however, contain a broader mix of signals, and the exudates of different fungal species induce similar Ca^{2+} spiking responses in the host plant, mirroring the relative generalist nature of the AMS.

The exact identity of the plant Myc-LCO receptor(s) is still unclear, but it appears to be a complex of LysM receptor kinases located on the plasma membrane. The general structure of these proteins is a single transmembrane α helix, with an extracellular carbohydrate binding domain built from leucine rich (LxxL) repeats, and an intracellular serine/threonine kinase domain. However, many of these proteins lack one or the other of these kinase or carbohydrate binding domains. The first receptor to be discovered was DMI2, which is required for nuclear calcium spiking[95,96], but does not directly bind Myc-LCO[75]. In the RNS, DMI2 forms a complex with the heterodimer of with LYK3 and NFP, a pair of LysM-RK that binds Nod-LCO[97,98]. The latter lacks an active kinase, but is phosphorylated by LYK3 upon LCO binding[98]. This complex then nucleates a larger complex to initiate second messenger traffic[75] (see **Section 1.3.2**). In the AMS, a LysM-RK with a non-functional kinase domain, LYR3, has been shown to bind Myc-LCO with high affinity[99] and then form a complex with LYK3 and DMI2[45], presumably the mycorrhizal equivalent of the NFP-LCO-LYK3 complex. A similar protein, LYR1, is also upregulated in the symbiosis[100], but has not been shown to bind Myc-LCO. Further to this, NFP and its tomato and *Lotus* orthologs (SILKY10 and NFR5), but not the rice ortholog (RLK2), are required for Myc-LCO perception, and all appear to have undergone duplication events. AMF produce a diverse range of MYC-LCOs, so this large and diverse collection of receptors may be required to maintain the generalist nature of the AMS, functioning semi-redundantly with each other[75]. LCOs have been shown to induce production of lateral roots via the CSP[39], along with a DMI3-independent pathway activated by a yet unknown fungal signal[101]. These lateral roots are more responsive to fungal signals, and preferentially colonised in many plant species[101]. In the legumes LCOs induce branching in root hairs[44], although this response to Myc-LCOs is probably a relic of the RNS.

The next group of signalling molecules isolated were the short chain (3-5 residues) chitin oligomers (CO4). These are produced by AMF in response to strigolactone, and induce calcium spiking and symbiotic signalling in the plant[40], although they do not promote mycorrhizal-associated lateral root branching[38]. As with the Myc-LCO receptor, the exact nature of the CO4 receptor is unknown, but both LYK3 and DMI2 are required for CO4 to induce nuclear Ca^{2+} spiking[40,102]. It has been suggested that plant chitinases induced and secreted during the symbiosis may also produce CO4 by degrading larger chitin oligomers to limit immune signalling[103], or that this was a product of cleavage by the receptors. However, CO4 can be recovered from GSE, and the requirement to cleave long chain oligomers to prevent a defence response seems unlikely given the recent finding that these long chain chitin oligomers are facultatively pro-symbiotic[41]. This third and most controversial group of mycorrhizal signals are 6-8 residue chitin oligomers (CO8). These are traditionally thought of as the main fungal MAMP, and are a key elicitor of programed cell death, via a $CERK1_2$/LYK4/LYK5 heterotetramer in *Arabidopsis* and a CERK1/CeBIP heterodimer in rice, both activating a cytoplasmic MAPK cascade[94]. Recent work in rice has shown that CO8, at higher concentrations than CO4 (10^{-5} M compared to 10^{-8} M), will induce symbiotic Ca^{2+} spiking. As with CO4, this CO8 response was dependent on *Os*CERK1[41]. How the plant distinguishes between pathogenic and symbiotic signalling in this manner remains unknown[9].

The whole field is currently rather confused, as LYK3 (the *Medicago* ortholog of *Os*CERK1) does not appear to function in plant defence responses against CO8, only in symbiotic detection of CO4[9]. Responses to different fungal signals also appear to differ between plant species (e.g. rice shows Ca^{2+} spiking in response to undecorated CO4, and tomato to $CO4_{\Delta 4'Ac}$, but not vice-versa[41]) and between experimental systems (root organ culture vs whole plant) and different measures of symbiotic potential (Ca^{2+} spiking vs marker gene expression). These differences in technique mean we must take pains to not overstate current knowledge of plant responses to fungal signalling. The molecular make-up of crude AMF GSE is highly complex, and the difference between symbiotic and pathogenic signalling in the host may be due to ratios of the various molecules, rather than a presence or absence.

The existence of different responses to fungal signalling molecules by different host species would appear to run counter to the generalist nature of the symbiosis. AMF may release a wide suite of compounds, and each host only detects a certain semi-overlapping range of these signals. This could arise from random genetic drift from the duplicated LysM-RK receptors, or be counter-selected by red queen dynamics with species-specific pathogens hijacking certain AMF signals, leading to certain hosts losing the ability to detect those signals. There is clearly selective pressure to remain generalist, given the increasing evidence of fungal symbiont communities changing seasonally and across the lifespan of the host[104,105]. The many different proposed LCO receptors suggest extensive gene duplication, resulting in multiple LysM proteins forming a receptor complex (e.g. the Nod-LCO detecting Sym37/Sym10/Sym2 in *Pisum sativum*[106]), to allow for specificity to certain LCO decorations. There is also proteomic evidence for multiple low affinity LCO binding proteins[107], playing an as yet unknown role in the AMS.

1.5 – Life inside the root

1.5.1 – A highway through the root: how does the plant accommodate hyphal growth

Fungal entry requires drastic remodelling of the host cell, an upheaval on the level of formation of dedicated organs like root hairs or guard cells. As the fungus forms an appressorium on the root surface or an entry peg above the cortex, the plant forms a pre-penetration apparatus[108]. This is a total repolarisation of the cell, requiring restructuring of the cytoskeleton and endoreduplication of the DNA to facilitate the massive amounts of protein synthesis needed. The nucleus and ER are moved along this reforming cytoskeleton to the point of fungal contact[108]. The vacuole, normally filling 90% of the cell[109], collapses to make room for the fungus, cell division machinery is repurposed[110] and symbiotic exocytosis markers are switched on[111–113], to build huge amounts of membrane and wall material[114] required to surround the symbiont.

Between the initial entry, and forming arbuscules, the hyphae spread through the root apoplast (**Figure 1.7**). This requires the loosening of the cell wall. In the ectomycorrhizal symbiosis, this is at least partially controlled by fungal CAZyme

effectors[115,116]. AMF have a lower genomic complement of CAZymes than EcMF[117], but genomic data predicts that AMF have a compliment of secreted CAZymes (16 in *Gigaspora rosea*, 25 in *R. irregularis*)[118]. The effector function of these secreted proteins has not been tested, but we would expect some to be involved in host colonisation. Other AMF effectors are known, including SP7, which alters plant ethylene signalling via interactions with ERF19, promoting a receptive host environment[12]. Another effector, Pep1, blocks the function of fungal cell surface receptor GPCR4, which aids colonisation by an unknown mechanism[119]. The plant helps the fungus along, producing its own apoplastic subtilases (*Lj*SbtM1/3[120]) to loosen the matrix of the cell wall, and downregulating (via small GTPase ROP9[121]) a respiratory burst oxidase RbohB[122] that negatively impacts fungal colonisation. RbohB activity improves rhizobia colonisation, suggesting that its function is signalling rather than producing superoxides to directly damage fungal hyphae[122]. An E3 Ub ligase LIN (*Lj*CERBERUS)[123] and carboxypeptidase SCP1[124] haves also been shown to facilitate hyphal elongation inside the root, but the mechanisms of their function are unknown.

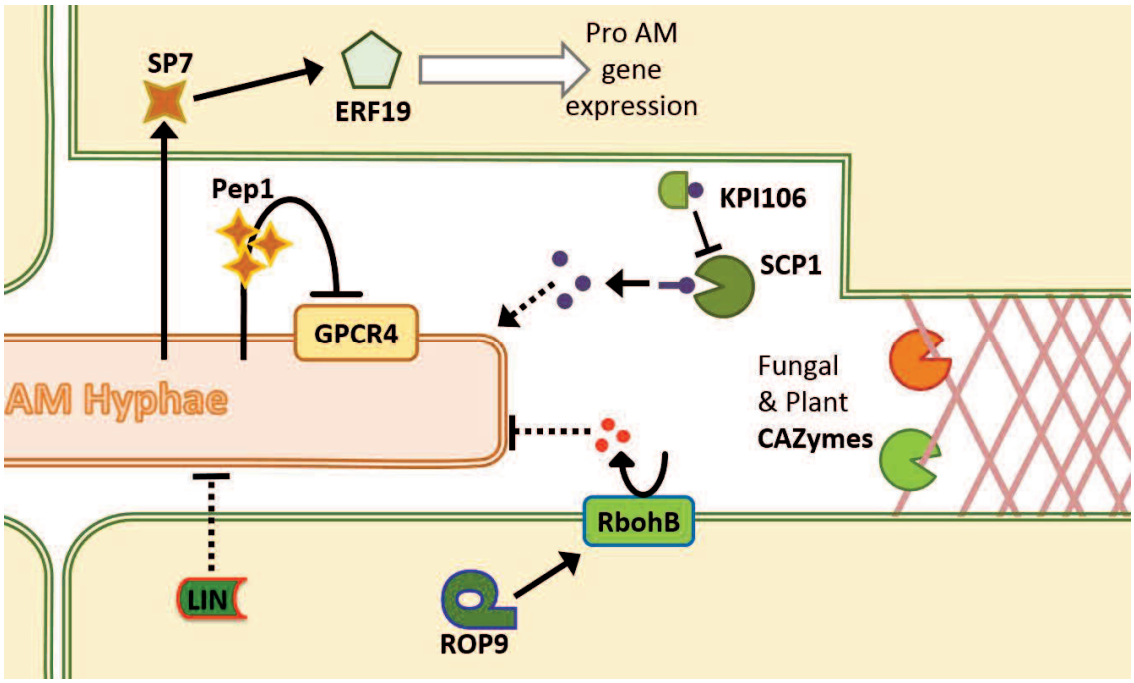


Figure 1.7 – Plant and fungal proteins control intercellular fungal colonisation
A variety of poorly understood pathways affect the spread of fungal hyphae through the intercellular spaces of the root. Both host and symbiont work in concert to suppress plant immunity and to remodel the intercellular space. Other plant factors like SCP1 and LIN regulate fungal colonisation at this stage via unknown means. Fungal proteins are shown in orange, plant proteins in green. Dotted lines indicate pathways of inferred action.

1.5.2 – Minerals for bread and butter: symbiotic nutrient exchange

For a long time it had been assumed that photosynthetic sugar was the carbon source of the fungus, with the presence of sugar transporters on the periarbuscular membrane demonstrated over 20 years ago[125] . However, this view was challenged by metabolic evidence that AMF could not produce fatty acids in the ERM[84], soon backed up by the genomes of the Glomeromycotina, which lacked the FAS enzyme required for fatty acid synthesis[31,126]. Recent work has uncovered upregulation of fatty acid synthesis in arbuscule containing cells. A mycorrhizal-specific plasmid glucose importer[127] and fatty acid elongase[128] drive increased flux through the pathway, preferentially retrieving C16 fatty acids from the plastid[129] and acting with a trio of previously characterised plant proteins (RAM2 and STR/STR2[93,130,131]), to produce and export β-monoacylglycerol to the fungus[128] (**Figure 1.8**). While the fungal uptake transporter has not been found, it seems likely that this pathway contributes not just structural FA, but considerable metabolic carbon to the fungus. The relative contribution of FA and sugar to fungal metabolism is likely genotype dependent. Arbuscule formation requires FA delivery for structure, but a HIGS/VIGS approach to knock-down fungal sugar uptake transporters offers a chance to examine the role of different carbon sources in fungal metabolism.

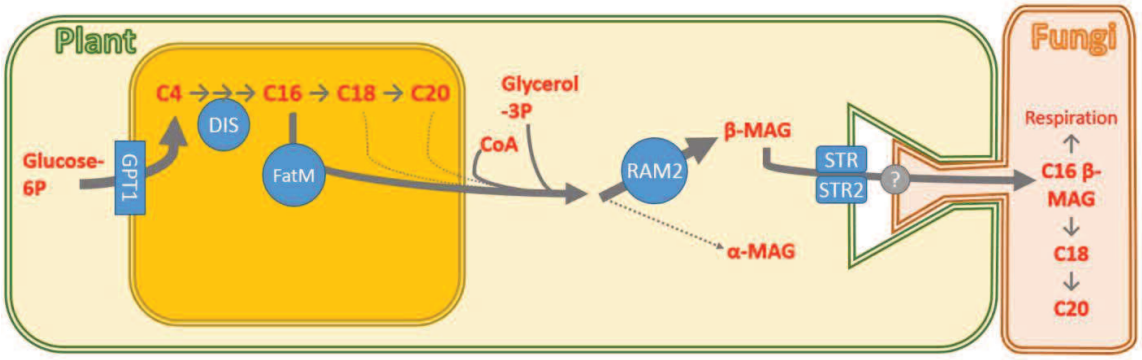


Figure 1.8 – A fatty acid synthesis pathway feeds the fungus
The cutin/suberin-like fatty acid synthesis pathway is recruited to provision the fungus with structural and respiratory carbon. GPT1[127] and DIS[128] (a redundant homolog of the standard fatty acid extension enzyme KAS1) drive increased carbon flux through the plastid FA synthesis pathway, and FatM[129] (a redundant homolog of FatB) and RAM2[93] redirects C16 fatty acids into the β monoacylglycerol pool[85]. This is exported into the periarbuscular matrix by the STR/STR2 transporter[130,131]. While the fungus cannot make its own FA[31,126], it possess the enzymes to extend C16 acids into longer forms[31], and produce the fungal specific ω5 desaturation[132].

The view of sugar transport has also changed, from the old expectation of sucrose export from the plant, either taken up by the fungus directly or after hydrolysis by extracellular invertase, to a new model (**Figure 1.9**). The plant releases glucose from a variety of monosaccharide transporters[100,125], which is taken up by fungal MST2[133], then converted to trehalose to prevent re-release[134]. A large number of different plant exporters have been implicated in different plant species[135], suggesting either a large redundancy in this process, or multiple independent convergent evolution events during the evolutionary history of the AMS. The latter (if provable) would support FA transfer being the primary C delivery to the fungus.

In return for these sugars and fat, the plant gains phosphorus and other nutrients (**Figure 1.9**). The fungus takes up Pi from the soil, converts it to polyphosphate bodies, which are transported to the arbuscule, broken down and released by exocytosis[136]. The plant retrieves this phosphate via secondary active transport[137]. The plant must maintain a proton gradient across the periarbuscular membrane[138,139] to obtain this phosphate. This gradient is also expected to be responsible for uptake of other symbiotic nutrients (see **Section 1.6.2**). In the case of sulphur, it has been shown that S uptake only occurs with Pi uptake[140], suggesting that it is demand for Pi that controls HA1 expression, and other actively taken up nutrients may be of secondary importance to the symbiosis. Nitrogen, transported through the ERM as arginine[141], is broken down in the urea cycle, and taken up by the plant through AMT family gas channels[142,143]. Amino acids are also obtained directly from the fungus, primarily those with a positive or neutral charge[144]. Symbiotic water uptake is also predicted, as both partners present aquaporins on the periarbuscular membrane[145–147]. This may help explain plant toleration to drought (see **Section 1.6.3**), and it has been suggested that the fungus links into the plant transpiration stream to reduce the cost of transport from the ERM to the arbuscule[148].

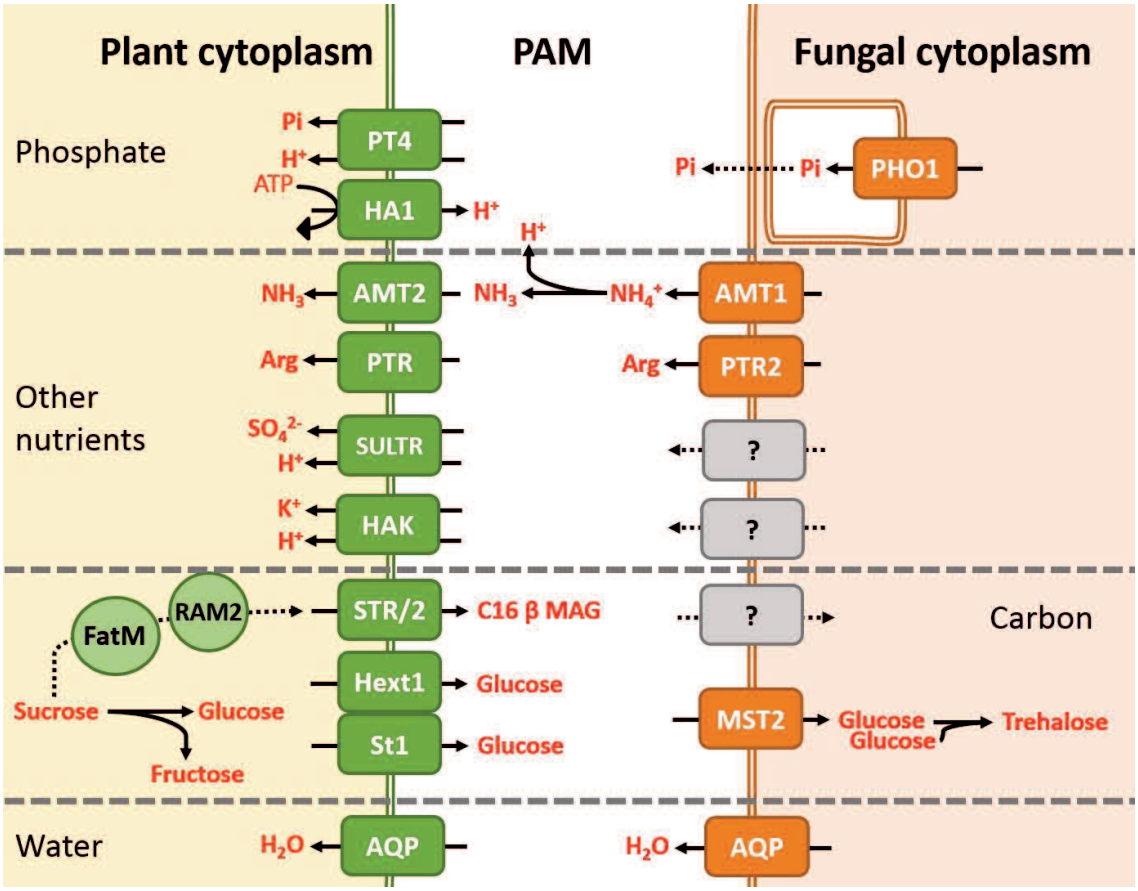


Figure 1.9 – Major nutrient exchanges across the periarbuscular matrix
Plant transporters: phosphate transporter 4 (PT4)[137], H^+ ATPase 1 (HA1)[138,139], ammonium transporter (AMT)[142,143], dipeptide transporter (PTR)[149], sulphate transporter (SULTR)[150], high affinity potassium transporter (HAK)[17,151], stunted arbuscule 1 & 2 (STR & STR2)[130,131], hexose transporter 1 (Hext1)[100], sugar transporter (St1)[125], Aquaporin (AQP)[145,146]. Mycorrhizal palmitoyl-acyl carrier protein thioesterase (FatM) and reduced arbuscular mycorrhization 2 (RAM2) synthesise C16 fatty acids for export[129].
Fungal transporters: phosphate transporter (PHO1)[136], ammonium transporter 1 (AMT1)[152], dipeptide transporter (PTR2)[153], monosaccharide transporter 2 (MST2)[133], Aquaporin (AQP)[147]

1.5.3 – Master switch(es) of the AMS

Many of the genes involved in the mycorrhizal growth program are either repurposed (e.g. *TPLATE, AP2*[154]) or duplications of other genes, with the same protein function and a different timing of expression (e.g. *VAMP721, FatM*[110,129]). Thus, we can infer that in large part the AMS brings together pre-existing building blocks to make its unique structure. This hypothesis would suggest the existence of a number of central regulatory proteins that switch cell fate to the mycorrhizal program. These genes (**Figure 1.10**) present important points for understanding and engineering the symbiosis.

We propose RAM1, a GRAS TF with one of the strongest knock out phenotypes known[72,155], as one of these master switches. The expression of *RAM1* is the first clear difference in the output of the CSP between the response to Nod and Myc factors. There are at least six members of this TF family involved in the AMS (DIP1[156], MIG1[157], NSP1[69], NSP2[39], RAD1[158], TF80[159]). Mutant studies suggest that the other GRAS TF have smaller roles in the symbiosis, as none shows as strong a phenotype as *ram1*. NSP1/2 are also involved in the RNS[63] (and conserved in the non-legumes, suggesting that their ancestral role is in the AMS[160]), where they form a NSP1/NSP2 heterodimer[161]. MIG1 is known to alter cortical cell development, and to facilitate arbuscule accommodation[157]. DIP1 binds DELLA proteins[156], so we suggest it links the GRAS TF array into the GA/DELLA network (see **Figure 1.10**). NSP1 is required for arbuscule degradation[73] and for lateral root branching[39], and may have other functions. RAD1 binds RAM1 and NSP2 (which also interact directly), and is involved in delivery of early arbuscule markers to the periarbuscular membrane[159]. TF80 has only been described as being controlled by RAM1 in expression studies, and is a binding partner with RAM1[159]. Park et al (2015)[159] also describe *TF124*, a close homolog of *RAM1* that appears to have an inverted expression pattern to RAM1 (i.e. when *RAM1* is upregulated, *TF124* is downregulated), which could also be important to maintaining the pro-mycorrhizal gene program. The specific function of NSP2 in the AMS is also unknown, but clearly important, as it is required for responses to Ca^{2+} spiking[39], and is targeted by miR171h[10] to exclude the fungus from the growing root tip. In conclusion, in the AMS we would expect to see the GRAS TFs interacting in complexes of two or more protein species, with RAM1 as the central component, switching partners to drive specific downstream responses.

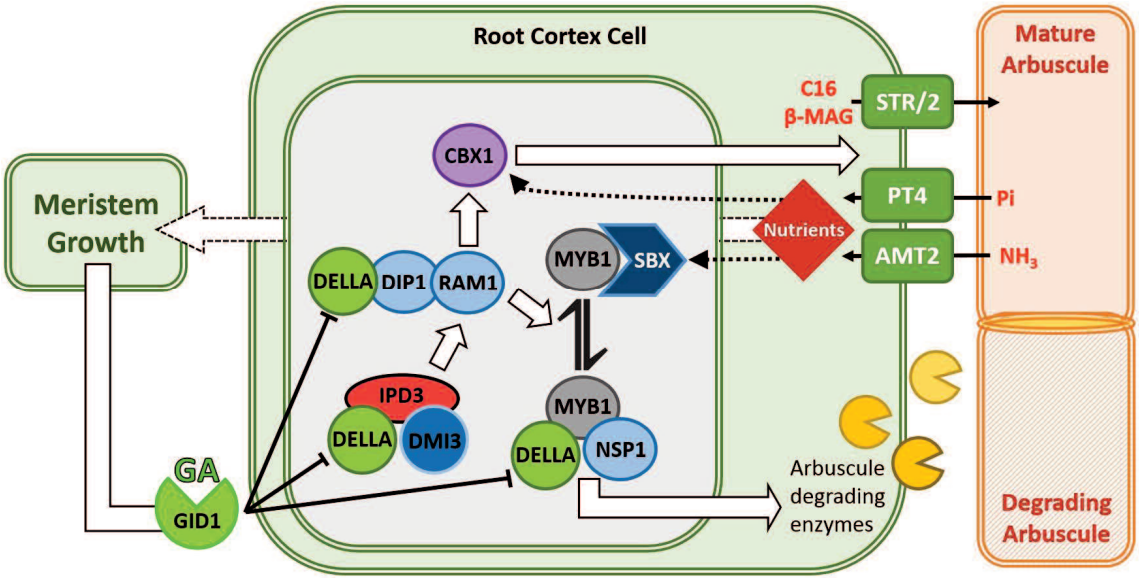


Figure 1.10 – Four master switches control the AM symbiosis
Plant and fungal proteins control intercellular fungal colonisation. RAM1 drives expression of the mycorrhizal growth program. CBX1 coordinates nutrient exchange. MYB1 integrates cell nutrient status to determine when to trigger arbuscule degradation. DELLA integrates whole plant nutrient status, to increase symbiotic expenditure when nutrients are required. Dotted lines indicate the hypothetical feedback from symbiotic nutrient import to promote CBX1 and suppress MYB1.

Downstream of RAM1 (**Figure 1.10**) another transcription factor, CBX1, appears to be a master regulator of reciprocal nutrient exchange[74]. It binds to a consensus sequence found in the promoter of the genes responsible for carbon provision (*FatM, RAM2*) and for phosphate uptake (*HA1, PT4*). It is not currently known how CBX1 is regulated to sense the onset and magnitude of phosphate delivery. Some species have separate phosphate sensors on the periarbuscular membrane, but in *Medicago* PT4 is believed to be involved in both bulk phosphate import and concentration sensing[162]. There may be some small expression of PT4 initially, that then feeds back into a 'full-blown' CBX1 response, but this is speculation. If CBX1 is feedback regulated by phosphate, then it provides a compelling argument for the plant to be able to discriminate between fungal symbionts on a cell by cell basis, greatly supporting the biological market hypothesis (see **Section 1.7.2**).

Characteristic of the arum type AMS is the relatively short lifespan of the arbuscule. A few weeks after formation, the arbuscule begins to change. The periarbuscular membrane is populated by t-SNARE protein SYP131II, and at this point the splice form of SYP131II changes[112], indicating a change in vesicle trafficking around the arbuscule. Visually, the arbuscule soon forms septa around the point of cell entry, then begins to collapse, before being digested by the plant cell as it returns to a general cortical cell

fate. A third transcription factor, MYB1, has been suggested as the master regulator of the switch from arbuscule maintenance to the arbuscule degrading state. It complexes with DELLA proteins and NSP1 to promote the expression of a host of hydrolases, chitinases and other genes associated with arbuscule degeneration[73]. What controls the timing of this switch to the arbuscule degrading state is currently unclear. PT4 expression is important to maintain normal function[163] (thus linking back to CBX1-mediated reciprocal exchange), but both MYB1 and NSP1 are expressed throughout the lifespan of the arbuscule. Floss et al (2017)[73] suggest that before the cortical cell switches to the arbuscule degrading state, MYB1 may be inhibited by binding other another protein, or by post-translational modification. A candidate for the former role would be a SPX domain protein, a family that is known to bind to and block the DNA binding site of MYB TFs at high nuclear Pi concentrations[164]. There are multiple overlapping pathways at work here however, as in *pt4*, AMT2 suppresses arbuscule degradation[142], and arbuscule degeneration still occurs normally in *pt4 myb1*, supporting the existence of a 'back-up' for MYB1[163].

The final master control point (**Figure 1.10**) is the aforementioned DELLA proteins[165]. In *Medicago*, there are three of these negative regulators of GA signalling, which act redundantly in the AMS. Their involvement was first shown by Floss et al (2013)[166], where a double knock-out gave a phenotype of reduced number of arbuscules while retaining wild type initial colonisation and intraradical mycelium (IRM) growth. Since then, they have been shown to act at all levels of the arbuscule development program, interacting directly with many important proteins. DELLA proteins forming a complex with DMI3 & IPD3 to active RAM1[64] and RAD1[167] expression and AMS specific signalling. They bind MIG1[157] to influence cortical cell development, DIP1[156] to promote arbuscule formation and MYB1[73] to enhance expression of arbuscule degrading enzymes. DELLA proteins are degraded by GID1 binding GA, so their presence acts to integrate plant growth and nutrient status. We would expect a fast growing (thus nutrient replete) plant produces a lot of GA, degrading DELLA proteins, limiting arbuscule formation to prevent the unnecessary carbon expenditure when cheaper P uptake opportunities are abundant. Addition of high concentrations of P has been shown to prevent new arbuscule formation, but we lack evidence on the effect on lifespan of pre-existing arbuscules.

The clear question to come out of all of this is: Why do arbuscules degrade? High levels of Pi and N in the cell would be expected to improve arbuscule lifespan, but whole plant nutrient status appears to push on both sides of the scale at once. DELLA proteins promote both arbuscule formation and degradation. Arbuscule degradation must be expensive to the fungus, even if it does retrieve much of the cytoplasmic contents prior to forming septae. To the plant it is even more expensive, since plant photosynthesis is the ultimate source of all the carbon burned to build the structure, and then make enzymes to digest and recycle the chitin. That a nutrient poor plant actively promotes the degradation process tells us that it is a key part of the symbiosis. Two hypothetical reasons can be suggested. First, that degradation and digestion of the arbuscule remnant is required for the plant to obtain something from the fungus (as is seen in the orchid mycorrhizal symbiosis (OMS)), although the nature of such a substance is unclear. Second, *pt4* mutants suggest that rapid turnover is a method of punishing non-contributing arbuscules, so the arbuscule turnover of may be required for selection for commensal behaviour, or for arbuscules with the most efficient links to the ERM[163].

1.6 – Houston, we have landed: the fungi that fed the conquest of land

Plants appear in the macrofossil record in the Silurian period, around 430 million years ago (MYA), with microfossil evidence for the spores of embryophyta-like organisms dating back to the mid Ordovician (475 MYA)[168]. The first Glomeromycotina-like microfossils appear 460 MYA[169], with the first physical evidence of mycorrhizal associations dated to the early Devonian (407 MYA)[27–29]. Work on protein mutation clocks suggests that the major fungal linages had diverged from their common ancestor by 960 MYA, with the Glomeromycotina splitting from the Basidiomycota/Ascomycota linage around 1400-1200 MYA[3]. The same method puts the emergence of the eukaryotic green algae (from which the embryophyta emerged) at around 1060 MYA, with the emergence of the tracheophyta from the embrophyta[170] around 700 MYA[3], although other work suggests dates much closer to that of the fossil evidence (470 MYA)[171]. Regardless of when exactly the genetic diversification that allowed the colonisation of land occurred, climatic factors may have precluded

significant expansion during most of the era before fossil evidence of sporopollenin. From 725 to 580 MYA, the earth's climate was dominated by the Sturtian, Marinoan and Gaskier glaciation periods, as oxygen and volcanic ash cooled the earth, presenting a great challenge to any would-be photoautotrophic explorers. Yet, it is during the short interstitial periods of global warming in this era that we first find evidence of 'true' multicellular animal life (i.e. that with differentiated and specialised cells)[172], so it is possible that the same shallow seas provided a similar nursery for terrestrial plants.

Despite the uncertainty of dating the conquest of land, it is clear that fungal symbiosis with algae and plants played a key part of this globe-changing event. The first fungal-algae associations likely occurred in the oceans of the late Tonian era (around 750 MYA), with free-swimming fungi associating with oceanic algae, surviving saprotrophically or pathogenically on pectin from those algal cell walls[4]. From this, various fungal clades became symbiotic, forming the lichens[173], which along with free-living fungi and early liverworts broke down the rocky surface to form the first soils. As soils appeared, the land plants, believed to have emerged from a single freshwater colonisation event[174], diverged and formed autotrophic rhizoids, long, thin tubular structures, lacking roots or leaves. These, like the liverworts, formed associations including modern-looking arbuscules and Mucoromycotina-like coils[29]. It is currently debated whether the 'original' plant root symbiosis was with the Glomeromycotina or the Mucoromycotina, as both symbioses clearly arose early in the plant linage. The Mucoromycotina symbiosis appears to be more efficient than the AMS under higher atmospheric CO_2[32], as was the case 500-400 MYA, suggesting it would have been more beneficial for early plants. But, given the paucity of fossil evidence, a definitive answer may never be known. These fungal associations likely provided necessary inorganic nutrition and water to the early plants, with this role continuing as roots evolved to provide stability to allow higher stems to outcompete an increasingly crowded land. The Mucoromycotina colonised the liverwort, hornwort and later the fern linages[175], while the Glomeromycotina also spread to the newly emerging spermatophyta, including the early conifers, remaining the dominate form of mycorrhiza until the emergence of the ectomycorrhizal symbiosis between Basidiomycota and Pinaceae between 180 and 50 MYA[174,176]. Both AMS and ectomycorrhizal symbiosis (EcMS) spread to the Angiosperms that rose to dominance by the late Cretaceous, and soon

after evidence of the nitrogen fixing *Rhizobium* and *Frankia* bacterial symbiosis appears in the fossil record[174], with the other Dikarya symbioses arising early in the evolution of their host taxa some 80 to 50 MYA for the Orchidaceae[177] and 120-110 MYA for the Ericaceae[178].

The early terrestrial earth was an inhospitable one, with less sunlight[179] to power growth, and no soil to support plants, nor store water between rainfall, and little access to mineral nutrients. While AM associations would not have been able to provide significant physical support, their great surface area would have aided in capturing water and retaining it between rains, supporting the poikilohydric early plants. AMF are still important for water provision under drought stress in modern plants, despite the evolution of roots and root hairs to fill the same role[180]. The early AMF associations also considerably boosted the ability of the plant to weather rock[181], the primary means of gaining nutrition in the early terrestrial environment. It is unlikely that we will ever know what the exact contribution of the AMS to the evolution of modern land plants was. But, given the conditions early plants faced, and the ubiquity of the symbiosis within the fossil record and extant embryophytic life (only the mosses do not appear to have mycorrhizal associations[2]), we can infer that this contribution must have been a hugely significant one.

1.7 – An ancient symbiosis: why has the AMS remained stable across evolutionary time

1.7.1 – The ever-present threat

Any cooperative relationship between organisms is vulnerable to 'lazy' or 'cheating' partners, selection favouring those gaining maximum benefit with minimum expenditure. In the ectomycorrhizal basidiomycetes, these competing pressures play out in an unstable mutualistic-pathogenic dynamic[182], with fungi evolving to maximise uptake of carbon from the plant, while the plant tries to eliminate non-beneficial fungi via immune responses. Despite this expectation, there is as yet no good evidence for non-symbiotic members of the Glomeromycotina, despite their half a billion year history. One reported exception to this is *Glomus macrocarpum,* reportedly the cause of tobacco stunt disease[183], but more recent work has shown it to be mutualistic in

other species[184]. It is not unusual for growth reductions under mycorrhizal conditions to be reported[185,186], so the report of Modjo and Hendrix (1986)[183] may simply be an extreme example in that particular genotype/environment niche. Exceptions aside, it does appear that the Glomeromycotina have retained mutualism throughout their evolutionarily history with land plants, putting them at odds with expectations and evidence from other fungal lineages.

The current consensus is that this stability is brought about by 'biological market dynamics', a structured reciprocal exchange of resources between partners. However, a lot of the experimental evidence supporting this comes from split pot or root culture experiments, with rarely more than three individuals involved. In a natural context, the symbiosis is a multi-party system, with a single hyphal network linking multiple plants of different species, and a single plant accessing multiple genetically distinct networks. This would appear to give both partners access to a biological marketplace enabling even the obligate biotrophic AMF to selectively deliver nutrients to the 'highest bidder'[21]. Experiments have shown that both plant[6] and fungus[11] can preferentially redirect internal carbon/phosphate flux to increase delivery to partners that provide the most phosphate/carbon in return.

1.7.2 – Control of reciprocal exchange

The level of control of nutrient exchange the partners are capable of is currently a topic of extensive research. Before the discovery of the FA transfer pathway, it had been suggested that the fungus obtained carbon via hexose symporters on the intracellular hyphae[8]. The increased local [P] from arbuscule delivery would lead to the plant allocating more photosynthate to that region of root. Sugars leaking from cortical cells would be taken up by the fungus. While we do not yet know the relative contributions of hexoses and FA to fungal metabolism, CBX1 provides a direct link between P/C exchange[74] on a cell by cell basis. Despite this, a very fine control of P/C exchange (e.g. 20 units of C for one of P) is almost certainly unrealistic in a biological system. At what scale the plant can measure Pi concentration at will have major implications for the biological market hypothesis, and for the ability of the plant to distinguish 'cheaters'. Direct measurement of Pi transport activity would allow for fine control of C delivery, but potentially make the system more vulnerable to fungal

effectors faking this signal. If the plant measures cytoplasmic Pi concentration, it would likely be influenced by other cells in the vicinity. In this case, since multiple species of AMF can be found in close proximity in the root in a wild system, a less beneficial AMF could 'hide' within a region of root containing highly beneficial fungi, ensuing a good supply of carbon for little cost[7].

The same holds true in reverse. It has also been observed that the C cost of maintaining a shared network falls harder on the dominant plant species in the CMN, with less input from subordinate species, even when both derive equal P benefit from the symbiosis[187]. From the fungal perspective, there is little incentive to disincentive this behaviour, so long as the dominant vegetation continues to be 'willing' to transfer sufficient C to support the mycorrhizal network. The mechanism of fungal partitioning of Pi from the ERM to its different IRM is currently unclear. It has been suggested that AMF link into the transpiration streams of their hosts in order to shuttle polyphosphate to the IRM[148]. However, this would provide another avenue for plant 'cheating', running high rates of transpiration to draw in Pi from a greater area. It also should be reasonably easily tested, as it would also suggest Pi delivery should mainly be to those plants nearest a Pi patch. In an ideal system, the fungus should directly sense carbon delivery, and be able to redirect cytoplasmic/vascular traffic to shuttle Pi for greater distances to bypass a less cooperative but closer individual.

1.7.3 – An updated model?

The market of the symbiosis is single blind; a plant cannot tell what another plant is trading and alter payment accordingly, and a fungus cannot observe other fungi. Thus, for the market to maintain mutualism, both partners must be promiscuous generalists who can accurately measure input and redirect supply accordingly. One updated model for the symbiosis is the concept of 'luxury goods'[7]. In this model a plant may not be incentivised to regulate nutrient exchange too closely if it is not carbon limited. This is likely to happen in the conditions that maximise mycorrhizal benefit, high N and low P and no climatic limitations on photosynthesis. The plant can produce large amounts of RuBisCO, but cannot usefully consume the C produced due to a lack of P to expand the biomass. In such a state, it is of no significant cost to the plant to maintain even unproductive mycorrhizal associations.

But, many plant species show a depressive mycorrhizal growth response (MGR)[186], suggesting that the C funnelled towards the fungus is very much a limiting resource. Why do the plants continue to invest in the network? One solution is to invoke the as-yet undiscussed elephant in the room, the nebulous non-phosphate benefits of the AMS (see **Section 1.8** for a brief overview) as sufficiently important to plant fitness (i.e. survival to breed and propagate young) that this loss of potential biomass can be an acceptable sacrifice. This may be an issue of experimental design, as most work is done in monoxenic culture, often with a single plant, and with relatively young monocarpic plants. It is unclear if this observed reduction in relative growth rate (RGR) is predictive of a loss of fertility and offspring viability, and how the lack of other individuals to interact with changes the symbiosis. Direct coupling of C/N exchange occurs, but not universally[11,188], although it is the norm in the EcMS. Other benefits are harder to establish – enhanced protection against drought or certain pathogens are unlikely to be a benefit to every plant generation, but how often do these need to be a threat, and how much of an advantage does the symbiosis need to provide before the cost of maintaining it through the good times is worth it? Obviously, plants cannot plan ahead evolutionarily, so some will lose the symbiosis, but the selection pressure must be strong enough that these individuals cannot reproduce when the hypothetical stress occurs.

This 'pay now, collect later' approach also appears on the fungal side. AMF with sufficient carbon maintain (presumably costly to build and maintain) colonisation of partners with low carbon returns[11,187], and even more strikingly will support achlorophyllous plants[184]. This growth-mode of seedling mycoheterotrophy immediate reminds one of the basidiomycete/orchid symbiosis, where an even more extreme version is clearly stable in evolutionary time[177]. However, orchids associated with OMF show a far greater tendency towards specialisation than AM plants[189], normally a hallmark of parasitic behaviour, although it is probable that over the lifespan of the plant, net carbon flux is into the fungal network. Additionally, both OMF and AMF would be simultaneously associated with other plants[190,191], so the fungus is never without a net carbon source and does not have to bank on investments at some point in the future.

1.7.4 – For every carrot, a stick

Another prediction of the biological market hypothesis is that the partners, as well as rewarding mutualistic behaviour, should be able to sanction non-cooperative partners. In plants, the clear expression of this is the rapid turnover rate of arbuscules that fail to deliver significant P or N[142]. If an AM associated plant is saturated with P (as might occur with a new fungal association with higher P delivery rates than its current partners), it prevents the formation of new arbuscules, further limiting the poorly mutualistic symbiont from carbon access. While most AM plants are highly generalist, they rarely associate with all AM taxa[192]. This could be due to an even more extreme sanctioning, the exclusion of AMF species whose particular benefits do not complement the plants evolutionary niche, although this could also be explained as the loss of ability to communicate via evolutionary drift (as is seen in many speciating insects).

Since less is known about fungal nutrient partitioning (see **Section 1.7.2**), it is harder to estimate a fungus' ability to sanction. The exclusion of certain plant/fungus combinations could be a response initiated by the fungus, but we think this is unlikely. In experimental systems, so long as the fungus has C contributing partners, they maintain high colonisation of less beneficial partners, even though nutrient transfer to these partners is reduced[11]. This even extends to colonising *Arabidopsis thaliana* in the face of a strong defence response from the non-mycorrhizal Brassicaceae[193]. Mycoheterotrophic plants (see **Section 1.7.5**) further cast doubt on the ability of a carbon replete AMF to sanction plants, which could also explain the benefits seen by low contribution subdominant vegetation[187]. If starved for C, a fungus does become selective, as can been see by the cycles of germination AMF spores undertake, germinating in good abiotic conditions, extending a hypha, and then retracting and returning to dormancy in the absence of compatible root exudates. This response is seen with non-hosts like *Arabidopsis*, which will not be colonised if there are no C-contributing hosts to support the fungus.

1.7.5 – Is carbon flux monodirectional?

Mycoheterotrophic and mixotrophic plants (deriving all or some of their carbon from mycorrhizal fungi) are well documented. These are classically members of the Orcidaceae, but an increasing number of AM mycoheterotrophs have been found[194,195]. These plants are parasites on the CMN, and must reverse the normal flow of carbon, while retaining P & N delivery. We would expect that these plants would be under the same selection pressures as pathogens (avoiding rejection or immune response, while moderating virulence to avoid overrunning the hosts), becoming highly specialised on a few species of fungi, and to show evidence of co-evolutionary selection with these. This pattern of increasing specialisation is seen in the orchids and in some AM mycoheterotrophs[194,196], but Merckx et al (2012)[195] found a large number of quite generalist AM mycoheterotrophs. They report thant 14 of 34 mycoheterotrophic flowering plants or fern gametophytes were colonised by at least three different Glomeramycotina virtual taxa, while 20% of these fungal taxa were also consistently found across the range of plant families. Given the plants were sourced from three continents, the generalists species are probably indistinguishable in their fungal range from most AM autotrophs.

While a reduction in fungal structural complexity has been observed in some mycoheterotrophic species[184] (suggesting carbon is obtained not by fungal export, but via necrotrophic digestion of fungal intercellular coils), carbon traffic between assumedly autotrophic species has been reported[197]. This phenomenon is currently little studied, so the extent it occurs in nature is unclear, as is any physiological relevance. Transfer of ^{13}C label applied to a neighbouring plant could occur either via digestion and recycling of fungal cell wall material in the degrading arbuscule, or via the carbon skeleton of arginine exchanged as part of N transfer. In these cases, the C flux is an insignificant part of the necessary C cycle of the symbiosis. It has been speculated that arginine catabolism plays an important role in carbon uptake by orchid protocorms[198]. This system (**Figure 1.11**) could also provide a neat bypass for the problem of reversing carbon flow in the AMS, although this would require the mycoheterotrophic host to induce a change in location of the urea cycle in AMF.

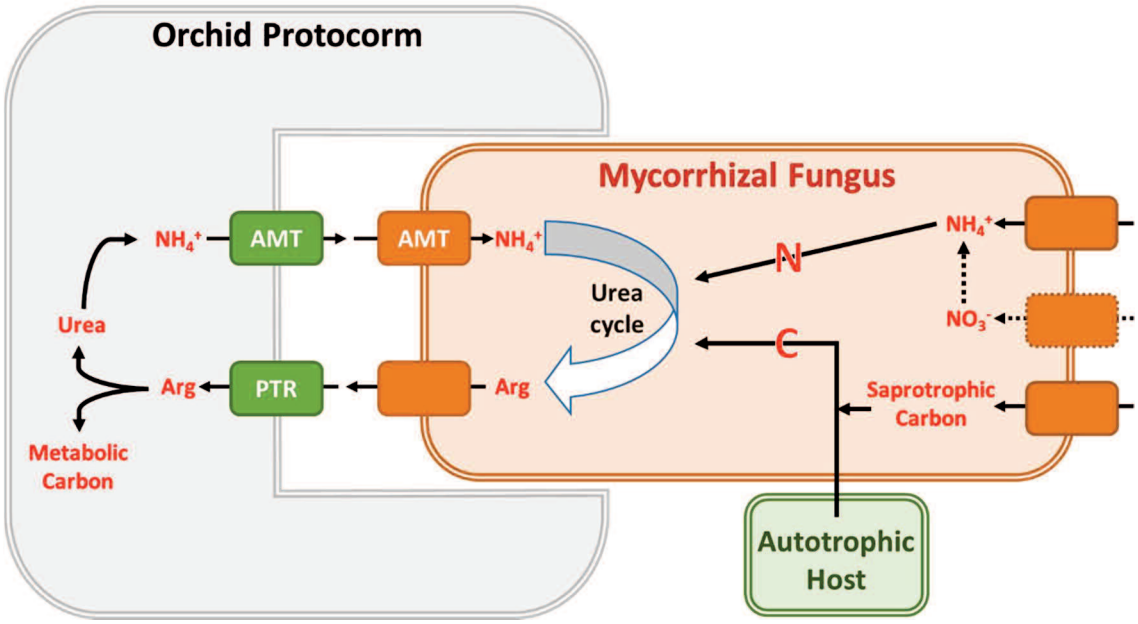


Figure 1.11 – Carbon acquisition by orchid protocorms from their fungal symbionts
A proposed route of carbon acquisition by the protocorm stage of orchids from their fungal symbionts, specifically *Serapias vomeracea* and symbiont *Tulasnella calospora*[198]. The orchid takes up arginine from the fungal interface, and utilising the carbon for metabolism and releasing ammonia. This creates a large concentration gradient between host and symbiont (the opposite direction to that seen in autotrophs), so ammonia passes through AMT gas channels back to the fungus. The fungus then runs the urea cycle to rebuild amino acids using carbon obtained from other plants or for saprotrophic litter decay. *T. calospora* lacks nitrate uptake capacity, but this is likely a local adaption rather than the rule for OMF.

Fungal carbon transfer is better documented in the EcMS, where significant transfer of carbon from non-shaded to shaded *Pinus* has been observed[197], suggesting that EcMF may 'fine-tune' the flora in their CMN to promote total survival as an insurance policy. We are not aware of reports of this behaviour in the CMN of AMF, but see no reason why it should not also occur there.

Another question posed is why the fungus has never gained (or retained) the capacity for saprotrophic carbon uptake. This is likely now a dead end, genome reduction having stripped the fungus of any litter decay functionality[37]. Simply expressing monosaccharide uptake transporters in the ERM would lead to insufficient carbon to sustain the fungus, given the intense competition by soil heterotrophic microbiota. Such as change, deliberately made, does have merit for facilitating axenic culture of AMF for use as a biofertiliser, especially if expression of the gene could be linked to supplementation of a xenobiotic compound to prevent expression in the field.

1.7.6 – Conclusions

In the legume RNS, the biological market has driven specialisation for specific partners with the most co-operative genotype[199], something not recapitulated in the much older AMS. Unlike the rhizobia, the Glomeromycotina are obligate biotrophs. By remaining generalist, with many plant species feeding into their CMN, the fungus is buffered against the loss of its host, to disease or seasonal changes in metabolism. This is then reinforced by plants, who we would also expect to remain generalists, given the different benefits observed in different fungal genotypes, leading to different AMF being optimal in certain environments. Many associations allow the plant to choose only the most productive fungi to support, and the fungal networks also show seasonality[105], so a wide range of associations allows the plant to consistently gain nutrition.

We know of evolutionarily stable plant 'cheaters', but none on the fungal side of the AMS. This could be an artefact of the lower intensity of study of subterranean biota, or show the comparatively stronger selection pressure imposed by the plant host on its fungal symbiont. Through times of plenty, stringency of the plants reward and sanction behaviour may be relaxed (as predicted by the luxury goods model), allowing commensal or parasitic life-history strategies to emerge in the Glomeromycotina, but as the load of these organisms on the plant increases, or the plant is otherwise stressed by changes in the environment, it would force selection for a return to high stringency behaviour, killing off or forcing a return to mutualism for the fungal lineage, at which point the cycle would begin again. Escape from this cycle is plausible, but would likely require gain of an effector suite that would allow the fungus to maintain arbuscules and carbon delivery without plant benefit, and more actively suppress plant immunity. Such a change would need to be sudden to escape the island of stability the AMF currently reside in, probably requiring whole genome duplication and a lack of counter-selection, or an extensive horizontal gene transfer event.

Another possibility for fungal cheating would be the existence of 'leech' fungi, which colonise the root in very close proximity to another mutualistic AMF, and gain their C by taking up C delivered to the fungal mutualist[8]. As described in **Section 1.7.2**, this would be easier in a broader, less cell-specific model of plant measurement of nutrient

exchange, and likely focus on gaining C from sugars, rather that C16 FA, the delivery of which appears more tightly constrained to the arbuscule. So long as such a 'leech' fungus had a low enough metabolic C demand, it would be possible that it would not impose a significant cost on the host plant (especially under the luxury goods model, which assumes photosynthate to be a non-limiting resource to the host). The cost of the association would be felt more strongly by the fungus, but again only if C was limiting in the system. We would expect such a life-history strategy to be most likely to occur in a stable ecosystem with year-round AMS associations, as the low rates of C uptake would make regular need to sporulate and re-colonise the host prohibitively expensive to the 'leech'. It may be possible that some of the endosymbionts observed in many AMF species (see **Section 1.9.5**) are in fact 'leeches' in this manner, and indeed their location in the hyphal wall matrix or cytoplasm of the AMF would give them access to the periarbuscular matrix and thus a greater nutrient availability than the main intercellular spaces of the root. Again, if the nutrient demand of such 'leeches' was low enough, it would not be worth the expense of eliminating them for the plant and AMF hosts.

1.8 – Looking past phosphorus: what are the benefits of the AMS?

1.8.1 – A wide array of published benefits

The AMS is ascribed an extraordinarily wide range of benefits, both in the scientific literature, and in the recent explosion of horticultural interest in plant inoculation. As well as its classically defined role in phosphate nutrition, the AMS has been reported to promote uptake of most of the other macro- and micro-nutrients[149]. Other reports ascribe roles in tolerance to almost every abiotic and biotic stressor imaginable. This has given rise to AMF inocula being sold as something of a cure-all to horticulturalists and amateur gardeners. Are any of these claims substantial, or are AMF inocula nothing more than snake-oil?

A closer examination of the literature reveals that most of the 'benefits' of the symbiosis are highly context dependent, an interplay of plant and fungal genotypes with the abiotic environment. Plant/fungus combinations from the same local

environment are more likely to be beneficial than isolates of the same species from different locals[200], where selection will have produced AMF pools with costs and benefits adapted to the local flora. There is evidence that the different benefits ascribed to the symbiosis are to some degree a trade-off which each other. There will not be a single fungus isolate that benefits the host plant in all circumstances, and while initial reports of crop inoculation are promising[26,201], gaining the maximal benefit from AMF in agriculture will require considerable investment in environmental sampling and the development of predictive models of the local limiting factors. Inoculation with a range of AMF species has been suggested as a short-cut, but there is evidence that this does not overcome the problem[202,203], although longer term (decade-scale) selection by plant hosts might self-select for locally beneficial AMF populations if such a population could be maintained in arable soils.

1.8.2 – Plant nutrition

As with P, the symbiosis is likely most beneficial to plant nutrition when soil concentrations of other nutrients are limiting. There is some evidence that N transfer can be coupled to C delivery[204] and give plant growth benefits[205], but also plenty of evidence to suggest that this benefit is dependent on genotype and environment[206,207]. The greater soil mobility of N and higher fungal structural demand (relative to P) both reduce the efficiency of symbiotic over direct uptake[208]. Other nutrients reported to be transferred by the symbiosis include iron[209], sulphur[140], potassium[17], magnesium[210], nickel[211], zinc and copper[212]. While these may be depleted in certain areas, or be desirable for human nutrition, in general NPK, insolation, temperature, and water are the main limiting factors to agricultural growth. Thus, any growth benefits from these nutrients would be hard to entangle, and may be accidental – evidence suggests that mycorrhizal provision of S to *Medicago* plants occurs only if P was also being transferred[140]. Additionally, *Rhizophagus* is unlikely to have many methods beyond more efficient exploration of the soil volume to obtain nutrients inaccessible to the plant. For example, *Rhizophagus* has a siderophore uptake system but lacks the ability to synthesise them, meaning it would be reliant on bacterial siderophores for its iron uptake[212]. AMF also lack the capacity seen in the EcMF to use Fenton chemistry to access recalcitrant soil nitrogen otherwise inaccessible to the plant[213].

Rapid arbuscular turnover is the sanction imposed on fungi (see **Section 1.7.4**) with low PT4-linked Pi uptake, and to our knowledge this phenotype is rescued only by AMT2-linked N uptake[142]. This would imply that any benefits of the symbiosis aside from P and N nutrition are secondary at best, although it must be caveated by the fact these studies look at laboratory culture, so may not account for changes that occur under a/biotic stress.

1.8.3 – Abiotic stress tolerance

AMF have been extensively studied as a way to reduce the linked problems of drought and salt stress, and have been shown to maintain plant growth under such conditions. Again, this is dependent on the AMF/plant species used, and the growth measure being considered[202]. For example, *Claroideoglomus etunicatum* and *Rhizophagus clarus* appear better than *Funnelliformis mosseae* and *R. intraradices* at maintaining shoot K^+/Na^+ ratios under salt stress, while legumes benefit more than grasses. The latter may be due an artefact of dominance mechanics however, as dominant vegetation normally shows less benefit from the symbiosis than subordinate species[214]. For drought stress alone, the choice of endpoint appears to drive experimental findings[200,215,216]. Many endpoints can be measured (e.g. gene expression, accumulation of proline or ABA, stomatal conductance) but it is unclear which of these are drivers, and which symptoms. Direct fungal delivery of water is definitely plausible, with aquaporins found on the periarbuscular membrane[145,146], but the effects of hormone homeostasis as a by-product of fungal manipulation of the JA and ABA pathways may also be important[16,217]. More experiments are needed, but are complicated by many factors, including that changes in soil moisture disproportionately affect N mobility relative to P (altering plant nutrient status) and accounting for changes to water use by the fungus, and the alterations in soil structure induced by glomalin[218].

Another major avenue of interest is in bioremediation of heavy metal contaminated sites[219]. AMF inoculation has been shown to promote tolerance to a wide range of heavy metals (including cadmium[220], aluminium[221], copper[222,223], asrenic[224], lead & zinc[225]). Once again, this strategy is complicated by the prevalence of genotype interactions (i.e. sensitive plant varieties become more tolerant with AM inoculation,

but tolerant varieties gained no benefit[223], and AMF species gave different levels of benefit[222]) and by experimental design (i.e. AM grasses on contaminated soil showed higher rates of photosynthesis than NM plants, but this did not translate into a growth benefit[226]).

1.8.4 – Mycorrhizal induced resistance to pests and pathogens

Mycorrhizal induced resistance (MIR)[227] is the catch-all term for the priming of the JA response pathway seen in AM plants. This improves plant resistance to many types of insect herbivore[228,229] and to necrotrophic pathogens[230,231], while having less effect on sucking insects or biotrophs[16]. As with abiotic stresses, this response varies with genotype (e.g. AM tomato has a large increase in chemical defences (relative to the NM plant) when challenged with insect herbivores but no difference without the herbivores[232], whereas AM nightshades show a defence increase when unchallenged, but AM and NM plants have little difference once herbivores were applied[233]. In bean and tomato, only some varieties exhibited MIR toward bacterial and fungal pathogens[234,235], and this depended on AMF species[236]). Multiple causes have been suggested for MIR, and the final cause is probably a combination of these. Systemically, AM plants tend to have better nutrition, so can direct more resources towards defence, although a positive MIR is still observed when phosphate nutrition is accounted for[234]. The fungus also manipulates defence hormone signalling to prevent the SA response employed against biotrophs. AMF induce production of OPDA (precursor to JA[237,238], an antagonist of SA signalling) and effectors like SP7 (which binds ERF19 to manipulate ET defences[12]) further alter host immunity. MIR can also be seen as a priming response, not directly producing expensive secondary metabolite defences, but increasing the abundance of receptors and intermediate signalling components, thus driving a much stronger response when a pathogen or herbivore is detected[239]. For resistance specific to root pathogens, other effects such as mycorrhizal induced changes to root architecture[101], physical blocking of root entry by the fungus and release of antibiotics by the fungus or its co-symbiont bacteria (which can in turn act as elicitors of immunity)[240] have also been proposed as causes of MIR.

Moving beyond pot experiments, recent work on CMN linked plants suggest that this too can play a role in biotic stress tolerance. Signals from plants under aphid attack

have been shown to be passed via the CMN (although the exact mechanism is currently uncertain) to other plants, priming their own defences against aphids[19]. These networks also aid the spread of plant allochemicals, facilitating suppression of competing plants[241], a function that may also play a role in invasion and other events characterised by the meeting of two distant plant/fungal communities[242,243].

1.9 – Moving beyond monoxenic culture

1.9.1 – Competition

Compared to the tightly controlled and uniform monoxenic cultures used in lab experiments, natural soils present a riot of information, challenges and opportunities from their complex communities. In this environment, both fungus and plant face steep competition for nutrients and space in the rhizosphere. This means that an organism in the field will often experience greater nutrient stress than one in monoxenic lab culture at the same abiotic nutrient concentrations, as other organisms compete for uptake. This leads to an increased demand on active uptake systems, increasing the carbon cost per unit nutrient returned, therefore increasing the relative cost to the plant to support the mycorrhizal network. However, this is likely offset by the similar competition-induced increase in cost for the direct uptake pathway. AMF also offer improved ability to obtain nutrients. AMF can explore the soil volume more quickly and with a lower metabolic cost than the plant[244], rapidly proliferate short lived hyphae in nutrient rich patches[245] and are even able to retract their cytoplasm from unproductive hyphae[246]. The Dikarya mycorrhiza have a variety of pathways (e.g. nitrogen extraction from litter via the Fenton reaction[213]) that allow them to access nutrients in forms that the plant cannot access. To the best of our knowledge, AMF lack these, mainly facilitating increased plant uptake via their more efficient soil volume exploration. They show some evidence of methods to improve uptake in competitive environments. *R. irregularis* expresses a siderophore uptake transporter to steal iron mobilised by soil bacteria[212], and it has been suggested that some AMF form mutualistic relationships with other nutrient mobilising bacteria[247]. They also facilitate inter-plant competition[241] (**Section 1.8.4**), effectively compelling other plants to form AM associations to remain competitive in the 'arms-race' with their neighbours.

1.9.2 – Less to go around: Pathogen and fungivore activity in the field

In a native soil community both plant and fungus are exposed to, and normally consistently challenged by, pathogens and predators. AM colonisation aids plants in defence against many of these pathogens and herbivores (**Section 1.8.4**), but inevitably these pressures will still impose a cost on the plant. These losses will increase the importance of mycorrhizal uptake, to replaced lost nutrients but also reduce the surplus carbon the plant has to pass to the fungus. If the limiting factor in plant growth changes, relative to lab experiments, this could radically alter the effects of the symbiosis (**Section 1.7.3**).

Pathogens and fungivores damage the hyphal network, requiring carbon be redirected to defence processes and to replace losses, to rebuild and maintain links across soil volume already exploited, rather than exploring novel volume to gather nutrients that could be exchanged with the plant. Both have the overall effect of increasing the carbon cost per unit nutrient that can be transferred to the plant host. Naturally, as the plant represents the sole carbon source to the fungus, this either leads to an increased carbon drawdown on the plant, or a reduction in nutrient transfer.

1.9.3 – A trade-off between defence and symbiosis?

All these pathogens and predators need to find their target. Thus, the signalling molecules released by the symbiotic partners to allow them to find each other (**Section 1.4**) can become problematic in the field. Parasitic plants of the genus *Striga* are the most famous example of this, germinating in response to strigolactone exudates of plant roots[248]. Other fungi and oomycetes have hijacked mycorrhizal gene RAM2 to provide root entry cues[93] and metabolic carbon[85]. Thus, selection must balance maintenance of pro-symbiotic programs and avoidance of hijackers.

This points to the existence of a yet poorly understood pathway, or series of pathways, that balance the hormone processes of the plant to integrate the demand for nutrients with the need to maintain defences against pathogens, while suppressing them to facilitate mycorrhizal colonisation (**Figure 1.12**)[85,156,166,249].

Another unknown in our understanding of the symbiosis is the overlap of pro-symbiotic signalling molecules with defence responses. For example, the cutin monomers produced by RAM2 are known to induce SA-linked defence responces[250] and chitin derivatives (primarily CO8) is well documented as a primary anti-fungal MAMP[94].

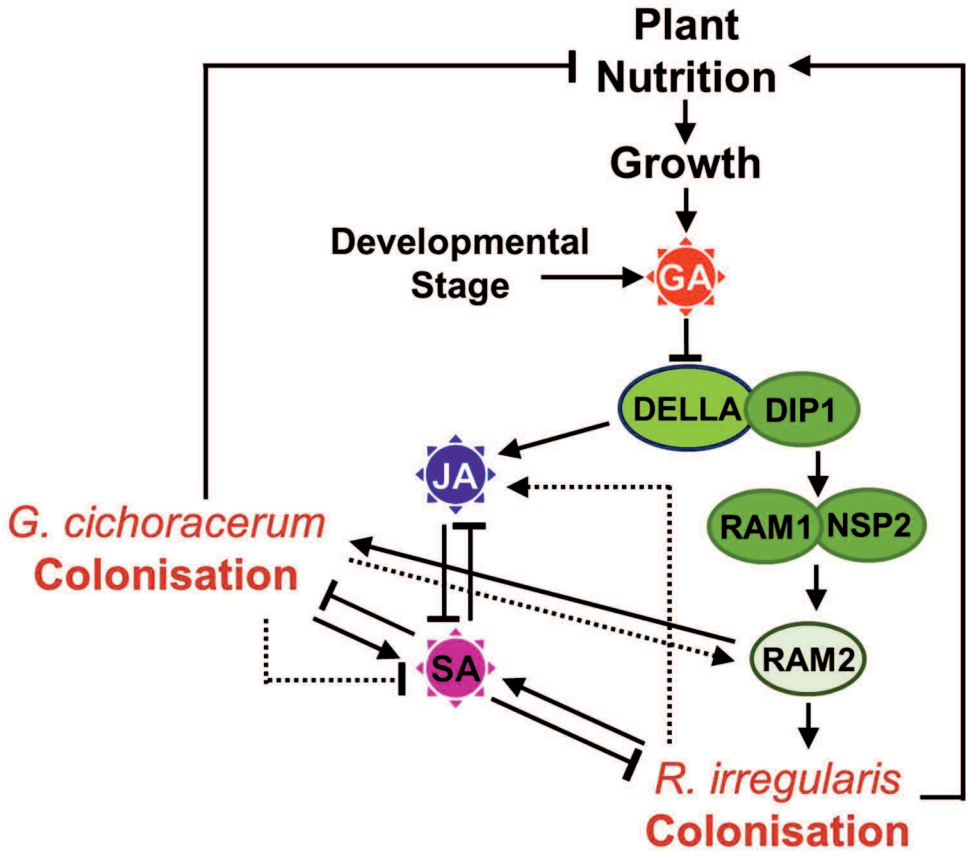


Figure 1.12 – The interaction of plant hormone networks is critical for symbiont and pathogen interactions
Hormone regulatory network of a mycorrhizal plant under attack from the 'hijacking' pathogen *Golovinomyces cichoracerum*. Supposed effector control is represented by dotted lines. A plant growth-limited by low nutrients limits GA production, leading to DELLA accumulation, facilitating RAM1/2 expression and promoting the JA pathway, suppressing SA-linked defences. As the plant gains nutrients, it degrades DELLA, limiting mycorrhizal growth. However, when attacked by *G. cichoracerum*, the plant must maintain SA defences in the face of falling nutrients (stolen by the pathogen or used to build the defence response). This suggests that while JA and SA have well known systemic effects, the optimum response would be to compartmentalise defence, and allow certain sections of the root to interact separately with mutualist and pathogen. Other hormones, such as ET or ABA, are also known to be involved in defence and symbiosis, so may facilitate this control.

1.9.4 – Friends as well as foes

Mycorrhizal fungi are, of course, far from the only 'friend' the plant has in the rhizosphere. Plants interact with many organisms under the umbrella of plant growth promoting rhizobacteria (PGPR) and fungal endophytes, covering a wide range of interactions and exchanges. Many PGPR may simply gain C from root tip sheath mucus or decaying plant cells, and benefit the plant by mobilising recalcitrant soil nutrients or breaking down organic stores, which the plant can then use. Many of these microorganisms are reasonably specific in their interactions, and further inter-plant competition by impairing the growth of their host's rivals and the rivals own symbionts[251].

The more elaborate symbioses (mutualistic endophytes, nodulating bacteria etc.) present more of a challenge to the AMF. They generally gain their C from the plant and can provide an overlapping range of benefits (e.g. N nutrition). Thus, while such interactions are purely beneficial to the plant host, they are direct competitors with the AMF for essential plant photosynthate. Does this relationship become antagonistic? Some nodulating bacteria have been shown to have anti-fungal properties, but these, like the plant defences, appear to be suppressed when in contact with AMF[252]. We know that mycorrhiza and nodules are spatially separated on the plant root, which may imply a plant system to keep them apart. This separation may be a necessary part of the maintenance of mutualism, preventing one organism leaching C intended for the plant's choice of symbiont. The plant genes RbohB[122] (which promotes the RNS and inhibits the AMS) and miR171h (which degrades NSP2 to limited fungal spread, but does not affect nodules[10]), suggest that this separation is under the control of the plant rather than a 'no man's land' between two warring symbionts.

1.9.5 – A fractal symbiosis: AMF have their own symbionts

Much of the literature involving the AMS concerns itself with *R. irregularis*, the first AMF to be sequenced. However, *R. irregularis* is unusual, as it lacks the endosymbionts seen in many other AMF lineages. These are best documented in the genus *Gigaspora*, which hosts the vertically transmitted *Candidatus Glomeribacter gigasporarum*[253], and *Funneliformis mosseae*, host to a wide range of bacteria and yeast species[254,255], along with cytoplasmic bacteria in many other AMF[256–258]. Cured *Gigaspora margarita* has been produced, and shows slower growth, reduced sporulation and decreased ATP production[253,259]. *Ca.G. gigasporarum* is totally dependent on its host. Selection for endosymbiotic life has clearly been occurring for a long time, with the bacteria only able to synthesise six amino acids, and dependent on amino acid catabolism for ATP as it lacks a key glycolytic enzyme[260]. *Ca.G. gigasporarum* may play a role in the plant-fungal interaction, as type III secretion system transcripts are strongly induced in the symbiotic mycelium[260], but its function to the fungal host is not essential[253]. For *F. mosseae*, the symbiosis appears to be more facultative, with its various symbionts producing siderophores, IAA and GA, and causing a significant decrease in plant mass when *F. mosseae* was cured[255]. Symbiotic yeasts live in the cell wall of *F. mosseae* spores, and aid phosphate accumulation[254]. Endobacterial associations are further complicated by the presence of a large number of viruses in AMF[261,262], particularly in *Gigaspora margarita*[263], though whether these play any role in the AMS is unknown.

We see a range of growth modes of AMF symbionts, from a lack of associations, through wide ranges of looser facultative associations to 'true' permanent vertical endosymbiosis. Other 'helper' bacteria associated with the hyphal surface can further improve colonisation and plant symbiotic benefit[264]. The different types of symbionts observed suggest that there may have been a single endosymbiosis event early in the Glomeromycotina lineage with a Mollicute bacterium (subsequently lost in *R. irregularis*), and later a large series of associations with other organisms in some lineages (e.g *Funneliformis*).

1.9.5 – What is 'a' fungus?

One problem with defining these interactions in the field is an inability to tell where one organism ends and another begins. We can genotype soil core samples, but have no way of knowing if two samples with the same genotype are a single connected individual or related neighbours. AMF hyphae can fuse with themselves, so multiple disconnected hyphae may link up, and with the discovery of mating type loci in *R. irregularis*[265], we know that this fusion can also occur between genetically distinct fungi to form a heterokaryon[266]. In the lab, these are apparently stable, so may well exist for long periods in the field, without an unknown signal to quickly drive karyogamy and meiosis. Can the AMF, which is multinucleate, aseptate and potentially heterokaryotic, as well as able to lose and re-attach parts of itself, be considered a single organism? Another view would look at the individual nuclei as the 'organisms', living cooperatively in a shared cytoplasm but competing for mitotic opportunity. The nuclei move through the hyphae, and have been observed to re-enter the spore before it returns to a dormant state[267]. However, closer observation has indicated that the nuclei are at least somewhat differentiated[267]. Spherical nuclei move freely through the hyphae, while another population of ovoid nuclei adhere to the cell membrane, at sites soon to undergo major events like branching. DNA staining indicates that at least some of these nuclei are then degraded when a hypha is abandoned. Thus, a spore re-entering a dormant state loses nuclei in the senescing hyphae, and some of those nuclei that return to, or remained with, the spore undergo mitosis to return the spore's nuclei count to the resting state. This largely uniform population, where some nuclei terminally differentiate and die without reproduction, is reminiscent of organisms on the cusp of multicellularity, such as the sentinel and stalk cells of the slime mould *Dictyostelium discoideum*[268] or the somatic swimming cells of many basal *Volvox* species[269]. This seemingly primitive system remains poorly understood, but efforts at AMF transformation offer promise[270]. If a marker could be produced (either directly fluorescent or a barcode for other fluorescent tagging) that would remain vertically associated with a small number of nuclei in the spore, then population dynamics and potential inter-nuclear competition could be examined.

1.10 – A forward genetic screen for new mycorrhizal genes

Current knowledge of the AMS has been mainly derived from the legumes, focused on the CSP. We aim to fill in the gaps in this knowledge, focusing on the FSP. We define this as the gene/protein networks required for the AMS that have not been cross-functionalised into the RNS, and conceive of four main parts to this pathway. First, signalling before and during hyphae-root contact. Second, the signalling pathway that is inferred to function alongside the CSP, allowing differentiation of the same Ca^{2+} spiking signal into differing gene responses, some of these independent of the *DMI* response. Third, the systems required to allow growth of the fungal IRM through the intracellular spaces of the root, and the construction of structures like the arbuscules. Finally, regulation of the symbiosis, how the plant measures fungal performance and rewards or sanctions its partners.

To expand our understanding of the FSP we carried out a forward genetic screen in a transposon mutant population of the model legume *Medicago truncatula*. This collection was screened for plants exhibiting a phenotype of reduced colonisation by *Rhizophagus irregularis* while retaining wild type levels of colonisation by *Sinorhizobium meliloti*. This should ensure that the CSP remains intact, such that any gene characterised in the study is important to the function of the AMS, as part of the FSP.

Arabidopsis thaliana, the plant model of choice, does not form associations with AMF. We chose to use *Medicago truncatula* (common name; barrel medic) for this study, as it is smaller and has a shorter life cycle (3-4 months) than the crop species often used in mycorrhizal research (e.g. maize, rice and soybean). It has a reasonably small (<500 Mb) genome, with one accession (A17) published and another (R108) publically available pre-press. Substantial genetic resources are available, including T-DNA and transposon insertion mutant collections. *Medicago*, a native to the Mediterranean[271], is also a legume (unlike rice or maize), allowing us to screen for the response to *Sinorhizobium meliloti* to remove CSP genes from our screening efforts.

The screen used a transposon insertion mutant collection produced and curated by the Samuel Robert Noble Foundation[272]. This collection uses tnt1 as the mutagen, a stress activated[273] 5334 bp *copia*-like retrotransposon isolated from tobacco[274]. This transposon was introduced into a parental line in the readily transformable R108 background by *Agrobacterium*-mediated transformation, and the stable offspring passaged through callus culture to induce transposition[275]. Tnt1 shows no site specificity and a preference for euchromatic regions[275], increasing the rate of gene disruption. The Noble Foundation estimates that ~25,000 lines in the collection contain an average of 4-40 new insertions, and that the transposon should remain stable in the progeny[276]. Many lines in the collection have had their insertions sequenced, and entered into an online database[277], which can then be used to obtain separate alleles of mutated genes.

This project takes a number of lines from the primary screen, confirms and describes the mycorrhizal phenotype and attempts to find the cause of these phenotypes. This will allow us to characterise more genes involved in the AMS, improving our knowledge and ability to manipulate the symbiosis.

Chapter 2 – Materials and Methods

2.1 – Media & materials

See **Appendix 4** for a description of the contents of all substrates and other growth media, fertilizers and stock chemical solutions, and for plant, bacterial and fungal stocks used in this study.

2.2 – Seed preparation

Medicago seedpods were collect from netted and dried plants and stored in paper bags at room temperature. Seeds were extracted by grinding seedpods between two sheets of corrugated rubber, then scarified by gentle rubbing with high grit sandpaper. Next, the seeds were sterilised by shaking in a 15% solution of commercial bleach for 10 minutes, then washed five times with autoclaved deionised water to remove remaining bleach. Seeds were germinated in Petri dishes containing sterile tissue paper soaked in autoclaved deionised water, and left for 24-48 hours in the dark at room temperature before transfer to the growth substrate.

If seedpods had not been stored for at least 3 months, they were vernalised for 7 days at 4°C in darkness prior to germination. Plants intended for split-plant or root culture experiments were germinated directly onto their growth substrate in petri dishes.

Allium seeds were placed in the growth substrate in 2 cm divots and placed into the greenhouse to germinate (see **Appendix 4.3** for conditions).

2.3 – Fungal material

Arbuscular mycorrhizal fungi used in this study were obtained from PlantWorks Ltd (Kent, UK), which provided various AMF species in either their commercial four species mix (RootGrow Professional; containing *Claroideoglomus claroideum, Funneliformis mosseae, F. geosporum* and *Rhizophagus irregularis*), or independently in single species inocula of the same species. This inoculum was formed of spores and dried

root material from the host plants (clover and maize) in a 1:1 zeolite/pumice substrate. Further to this, *Rhizophagus irregularis* DAOM-197198 (the strain sequenced by Tisserant et al (2013)[37]) was obtained in monoxenic culture on carrot hairy roots from our collaborators in the Murray lab (John Innes Centre, Norwich).

2.3.1 – Production of root organ cultures to propagate AMF

Agrobacterium rhizogenes Arqua-1 from glycerol culture was streaked out on TY plates, and grown at 25°C for 2 days.

To produce carrot hairy roots, a carrot was cut into 0.5-1 cm thick slices. These slices were surface sterilised by immersion in 95% EtOH for 10 seconds, 10% commercial bleach solution for 10 minutes, then washed four times in autoclaved diH_2O. The sterilised slices were placed distal side up on water agar plates and streaked with *A. rhizogenes*. After 3 weeks in the dark at 28°C emerging root tips were removed with a sterile scalpel, and transferred to MW+Suc plates (see **Appendix 4** for media) containing 500 mg/L carbenicillin. These roots were allowed to grow, subculturing every 3 weeks until no bacterial growth was observed.

To produce *Medicago truncatula* hairy roots the root apical meristem of a 2 day old seedling was removed with a sterile scalpel, and the cut section of the seedling's root was touched to a lawn of *A. rhizogenes*. The seedlings were transferred to MW plates, and incubated in the growth room (see **Appendix 4.3** for conditions) for 3 weeks, keeping the plates at a 45° angle (root downward). Roots that emerged from the cut were removed and transferred to a MW+Suc plate containing 400 mg/L augmentin. These roots were allowed to grow, subculturing every 3 weeks until no bacterial growth was observed.

2.3.2 – AMF production by hairy root culture

AMF were maintained by subculturing, transferring 2 cm^2 sections of hairy root culture to new 12 cm square Petri dishes (Greiner Bio-One, Kremsmunster, Austria) containing M+Suc media. This was repeated every 4-6 months to ensure growth and spore production were maintained.

To provide AMF material for experimentation, hairy roots culture sections were transferred to 9 cm round split plates (Thermo Scientific, MA, USA) containing 25 ml M media on one side, and 25 ml M+Suc media on the other. An autoclaved 1x4 cm section of filter paper (Whatman PLC, Maidstone, UK) was placed over the central divider to form a paper bridge[278] while the media was liquid. The 2 cm^2 section of stock plate was placed on the M+Suc media, and the fungus could grow across the paper bridge to colonise the M media, which lacked the sucrose required to support carrot hairy root growth.

After 4 months of growth, the M media compartment of plates visibly containing AMF mycelium and spores was removed and stored in two 50 ml polypropylene tubes (Sarstedt AG & Co, Nümbrecht, Germany) at 4°C. When a defined spore solution was needed, 40 ml of citrate buffer (pH 6.0) was added to each 50 ml tube, then the tubes were gently shaken overnight at room temperature to dissolve the agar medium. Spores were precipitated by centrifugation at 2000 g, and then washed twice with autoclaved diH_2O. The number of spores was quantified with a Fast Read disposable counting chamber (Immune Systems Ltd, Devon, UK).

2.3.3 – AMF production by stock pot culture

For routine colonisation experiments stock pot inoculum was used, as it was cheaper and easier to produce than a defined spore solution. In addition the inclusion of root fragments led to a more vigorous inoculum compared to isolated spores. Stock pot cultures were established in 2 litre pots with a sand:Terragreen substrate planted with either leek (*Allium ampeloprasum* var. Albana) or chive (*Allium schoenoprasum*) as the host. The stock pot culture was established by watering 4 week old host plants with a defined spore solution of *R. irregularis* DOAM-197198 or by mixing 100 ml of the PlantWorks RootGrow Professional inoculum into the substrate before planting. Pots were grown for at least 3 months, watered daily with tap water and fertilised weekly with 200 ml R solution ([279], modified by [278]).

When inoculum was required, a stock pot was devegetated, the roots chopped into sections <2 cm long and returned to the mixed substrate, which was mixed into the sand:Terragreen substrate used in colonisation experiments. One quarter of the stock pot substrate was used to establish a new stock pot.

2.4 – DNA extraction, amplification and sequencing

2.4.1 – DNA extraction

DNA was extracted by either the DNeasy Plant Kit (Qiagen, Hilden, Germany), per the manufactures instructions, or by the CTAB/chloroform method ([280], modified by [281,282])

An young folded *Medicago* leaf was removed from the plant with fine forceps, which were washed in 70% ethanol between plants to avoid contamination, and placed into a 1.5 ml microcentrifuge tube containing 200 µl 2x CTAB. The leaf was homogenised using a vertical benchtop drill and microcentrifuge pestle, vortexed briefly, then incubated at 65°C for 30 minutes to denature the released enzymes. This solution was placed on ice and 200 µl chloroform was added. After vortexing for 15 seconds, the solution was centrifuged at 20,000 g for 5 minutes. 120 µl of the upper phase was transferred to a new microcentrifuge tube containing 300 µl 100% ethanol, vortexed briefly to mix and incubated for 30 min at -20°C. After this, the sample was centrifuged at 20,000 g for 15 minutes to precipitate the DNA. The supernatant was discarded and the pellet incubated for 30 minutes at 35°C to remove any remaining ethanol, then resuspended in 50 µl elution buffer (Qiagen) and stored at -20°C.

2.4.2 – Nucleic acid quantification

Nucleic acids were quantified with a NanoDrop 1000 or NanoDrop 8000 (NanoDrop Products, DE, USA) following the manufacturers recommended protocol.

2.4.3 – Triplex PCR

Triplex PCR was used to genotype tnt1 insertions, using two primers specific to the flanking sequence of the insertion locus, and a universal primer specific to one end of the tnt1 insertion (see **Figure 2.1**). This gave two possible amplicon sizes between 200 bp and 1000 bp, depending on the primer set. One amplicon indicated the absence of tnt1, and the other the presence of the insertion. The presence of both amplicons in a single reaction indicated a heterozygous tnt1 insertion.

For the PCR, a 10 µl reaction mix was used, containing 0.5 µl of DNA extracted by the CTAB/chloroform method, 0.2 U of Phire HotStart DNA Polymerase (Thermo Scientific), the provided polymerase buffer, 0.5 µM of each primer and 200 µM of each dNTP, with the excess made up with nuclease free H_2O.

Figure 2.1 – Triplex PCR to locate tnt1 insertions
Location of primers (red arrows) for tnt1 insertion genotyping by triplex PCR

PCR used an initial 5 minute melting step at 98°C, followed by 35 cycles of a 10 second 98°C melting step, a 15 second annealing step (temperature between 58 and 62°C, depending on the primers used) and a 30 second extension step at 72°C, with the final extension step extended to 5 minutes to ensure completion of all current syntheses. The primer annealing temperature was calculated using the IDT OligoAnalyzer tool[283], with the following concentrations: 0.05 µM primers, 50 mM Na^+, 1.5 mM Mg^{2+} and 0.8 mM total dNTPs.

7 µl of the reaction mix was mixed with Orange Loading Dye (Thermo Scientific) and run on a 1.5% agarose gel at 110 V for 40 minutes, with SYBR-SAFE as the visualisation agent. Gels were pictured and edited using a Computar H6Z0812 camera (CBC Group, London, UK) and GeneSnap software (SynGene, Haryana, India).

2.5 – Microscopy

2.5.1 – Ink staining of mycorrhizae

Standard phenotyping was performed on ink/acetic acid stained roots[284] with dissecting or light microscopes. 5-8 week old plants were harvested, roots washed in tap water to remove remaining substrate, and placed into 50 ml polypropylene tubes containing 35 ml of a 10% solution of KOH. The tubes were incubated at 80°C for 50 minutes to clear the cells, then washed three times with diH_2O to remove the KOH. The roots were drained and stained with 10 ml of a solution of 5% acetic acid and 5%

Pelikan Brilliant Black Ink (Pelikan, Hannover, Germany) and shaken gently at room temperature for 15 minutes, then destained by washing three times with a 2% solution of acetic acid over two days. For thicker/older root tissue that stained or destained poorly, repeating the process helped give clearer samples without the destruction of young tissue occurring with a longer 80°C incubation. Roots were then viewed whole under a dissecting microscope (Nikon SMZ800; Nikon, Tokyo, Japan), and colonisation was assessed by counting the presence or absence of fungal colonisation or the various fungal structures (hyphae, vesicles and arbuscules in a hierarchical order) at every point the roots crossed a grid overlay. Alternately, the roots were cut into 3 cm sections and mounted on a slide, and colonisation assessed at distances along the root sections using a graticule and light microscope (Nikon Optiphot 2), with pictures captured with a AxioCam MRc5 (Carl Zeiss AG, Oberkochen, Germany).

2.5.2 – WGA-AF488 staining of mycorrhizae

To take pictures of the fine structure of arbuscules, Wheat Germ Agglutinin-Alexa Fluor 488 nm conjugate (Molecular Probes Inc., OR, USA) was used to visualise the fungal structures for fluorescence confocal microscopy, as described in Lum et al (2002)[285]. 5-8 week old plants were harvested, roots washed in tap water to remove remaining substrate, chopped into 3-10 mm sections and placed into 15 ml polypropylene tubes containing 10 ml FAA solution. These tubes were shaken overnight at room temperature to fix the roots, which were then washed twice with diH_2O, and cleared by a two hour incubation in a 50°C water bath in 10 ml of a 10% solution of KOH. Samples were then stained by overnight incubation in 5 ml of a 10 µg/L solution of Wheat Germ Agglutinin – Alexa Fluor 488nm (Invitrogen, CA, USA) in PBS. After washing twice in diH_2O, roots were counterstained for 5 minutes in 5 ml of a 10 µg/ml solution of acid fuchsin in PBS. Roots were mounted on a slide, and visualised using confocal microscopy (Zeiss LSM 710, Carl Zeiss AG).

Chapter 3 – Phenotypic analysis of *Medicago truncatula* lines with aberrant responses to AM fungi

3.1 – Introduction

Forward genetic screens, the assigning of function to genes by screening a randomly mutagenised population for a phenotype of interest, has long been one of the core approaches to gene function discovery. With the advent of large scale expression databases, such as the Noble Foundation's *Medicago* gene expression atlas[286], and price decreases in RNAseq, reverse genetic screens have been rising in popularity. These approaches look for genes upregulated in a chosen condition, and then obtain mutants for these genes, and examine the resulting phenotype to identify function. Both approaches have different strengths and weaknesses. Reverse studies require either significant investment in production of mutants, either specifically for the study, or by identification of mutations in large mutant collections. Additionally, it can only identify genes whose function is controlled by changes in expression, so cannot locate genes primarily regulated by post-transcriptional/translational regulation or genes (such as many receptors) expressed at constant levels. Forward screens also require large mutant collections (although one can selectively genotype those of interest), which can lead to repeatedly finding the same gene. Since they initially look for a phenotype, they will miss genes with redundant function.

Because of these weaknesses, any attempt to understand a plant process like the mycorrhizal symbiosis require both types of study to ensure parts are not missed. We used forward screening due to the availability of the large Noble Foundation mutant collection. The forward genetic approach is also more 'blind' on the part of the experimentalist, so limits the danger that incorrect assumptions about which genes could be involved might bias which genes are studied.

There are several different methods that have been used to quantify success of the symbiosis, including visual inspection of fungal structures (WGA, ink or acid fuchsin staining, fluorescence labelling of marker genes), genetic (*PT4* RNA abundance) or

chemical (C16:1ω5 abundance)[284,287]. We chose to use ink/acetic acid staining as the main method of quantifying root colonisation as it was quicker and significantly less expensive than genetic or chemical methodologies, allowing for larger sample sizes for screening of mutants and backcrossed lines, and involved less toxic substances than acid fuchsin staining. We supplemented this main screening technique with quantitative real time PCR (qPCR) for various marker genes and with WGA-AF488 staining to better describe the mutation in our mutant lines.

The *M. truncatula* lines used in this study were described as reduced AM colonisation mutants during a previous screen in 2005-2008[272]. This screen grew 3237 *Medicago truncatula* lines from the Noble Foundation tnt1 mutant collection with Endorize TA (a commercial inoculum containing a mixture of AMF species produced by the now defunct company Biorize) and the nitrogen-fixing bacterial symbiont *Sinorhizobium meliloti*. After 6 weeks of growth, roots were visually inspected with a dissecting microscope to look for the presence of pink nodules and absence of visible extraradical hyphal networks between the roots, then promising candidates were stained and checked for internal AM colonisation. 105 *Medicago* lines were preliminarily identified as having reduced AM colonisation, whilst retaining the ability to form mature nodules. This screening protocol should target genes important to the AM symbiosis that function outside the CSP, for example in the perception of Myc factors or building of structures like arbuscules unique to the AM symbiosis.

3.2 – Methods

3.2.1 – Plant growth

Initial colonisation experiments were performed in 80 ml pots (P24) with a sand:Terragreen substrate and RootGrow Professional as the inoculant. The initial experiment was grown for 5 weeks with 2.5 ml of inoculum mixed into the substrate, and the second and third used 5 ml of inoculum delivered into the hole made for the 2 day old seedling, and were harvested after 6 weeks in the greenhouse (see **Section 2.3** for the inoculum and **Appendix 4** for the growth conditions). These changes were made to increase the overall level of colonisation at harvest, as the wild type was only

15% colonised in the first experiment. Ink/acetic acid staining was used to find general phenotypes and WGA-AF488 staining to visualise arbuscule fine structure.

3.2.2 – Quantification of gene expression

Plants were grown in the greenhouse for 5 weeks, in sand:Terragreen, with a 1/3 volume inoculum from a *Rhizophagus irregularis* stock pot with leek nurse plants. RNA was extracted using RNeasy Plant Kit (Qiagen), per the manufacturer's instructions. 200 mg of root tissue was flash frozen in the supplied buffer and then homogenised with a Retsch MM300 Ball Mill (Verder Scientific, Haan, Germany). First strand cDNA synthesis was performed on 2 µg of this root RNA with M-MLV reverse transcriptase (Promega) according to manufacturer's instructions. The cDNA product was diluted 10-fold in nuclease free H_2O and 2.5 µl of this diluted cDNA added to 12.5 µl master mix (Applied Biosystems Fast SYBR-Green; Thermo Fisher Scientific) and 2.5 µl of 1 µM of two primers for each gene of interest. These primers gave an amplicon of between 80-150 bp and spanned an exon-exon boundary (see **Appendix 2** for primer sequences). These were used in qPCR in a StepOnePlus RT-PCR machine (Applied Biosystems, Thermo Fisher Scientific).

3.3 – Results

3.3.1 – Confirming the reduced colonisation phenotype

The lines obtained from the Noble Foundation had been screened for phenotype of reduced mycorrhizal colonisation and normal nodulation by Schultze and collaborators. In this project, we performed a more detailed phenotyping to confirm and describe the mycorrhizal phenotype.

Initial experiments were carried out to confirm the reduced colonisation phenotype. First, three lines (NF3209, NF3438 and NF472) were chosen, and plants of each genotype were grown in sand:Terragreen with PlantWorks four species AMF inoculum (at 10 times the recommended gardener's concentration). The R108 ecotype of *M. truncatula*, from which the Noble Foundation lines were derived, was used as the MYC^+ control. After 5 weeks of growth the plants were harvested and the percentage of root length containing AM structures was assessed.

In the first experiment, all three mutants tested showed significantly reduced AM colonisation relative to the wild type (**Figure 3.1**). NF3209 and NF3438 exhibited almost no colonisation. NF742 had significantly more colonisation than NF3209 and NF3438, but significantly less than the wild type. However, the overall level of colonisation in the first experiment was much lower than expected from the literature, as the wild type R108 plants did not exceed 25% of root length colonisation, well below the 50-80% level seen in well-established symbioses[10,138]. In addition, colonisation of the wild type was highly variable ($\bar{x}$=15.2% with a $\sigma_{\bar{x}}$ of 4.72%; subsequently presented in this chapter as 15.2±4.7%). The variability is likely due to variation in the number of infectious particles of AMF different plants were exposed to. If the experiment was assessed before the AMF had had time to reproduce and populate the root sufficiently to overcome this initial bias and reach the maximum colonisation capacity, different levels of colonisation would have been observed. This problem of highly variable results was reduced by changes to the inoculation strategy of future experiments, favouring fresher and stronger inoculum derived from monoxenic leek/*Rhizophagus irregularis* stock pots over the PlantWorks commercial product, but the variability problem plagued the study throughout.

To better assess the phenotype of the various mutants, conditions were optimised to improve the colonisation in the wild type, extending the growth period to 6 weeks and increasing the amount of inoculum applied. The experiment was repeated twice in these conditions, and two additional lines (NF1436 and NF443), which had also been identified in the preliminary screen, were added to the subsequent experiments. In these experiments, carried out in February and April 2014, overall wild type colonisation was closer to the theoretical maximum, reaching 42±13.2% in February 2014 and 39.9±11.9% (**Figure 3.1**) in April 2014. NF3438 demonstrated essentially complete abolition of colonisation in February 2014, with 1.5±1.95% colonisation in the April 2014 experiment. This phenotype ranks as one of the strongest seen in MYC^- plants, which in contrast to NOD^- mutants rarely show more than a reduction in colonisation, rather than an elimination of fungal structures. NF3209 and NF472, on the other hand, both showed an increase in colonisation relative to the control compared to experiment 1. NF3209 showed less colonisation than the wild type (22.5±8.4% and 21.9±6.5%), although this difference was not significant in April 2014,

whereas NF472's phenotype was similar to wild type in both subsequent experiments (41.3±10.6% and 32.4±9%). The difference in the NF3209 phenotype between experiments could be explained as the mutation delaying colonisation, but that given time and/or a stronger inoculum, colonisation will increase, but still lag behind the wild type. For NF472, the mutation could be giving rise to a brief delay in colonisation, which with enough time returns to a roughly wild type state. However, even if we assume the different phenotypes seen in experiment November 2013 and February/April 2014 are an accurate measure of the population mean colonisation, with such a minor change in timing of colonisation it would be very difficult to conclusively ascribe a mycorrhizal phenotype to this line.

NF1436 shows a similar phenotype to R108 and NF472, with only a slight and non-significant reduction in colonisation relative to the wild type (32.4±7.9% and 27.8±12.5%). In the February 2014 experiment, NF443 appears to exhibit the second strongest reduction in AM colonisation phenotype, after NF3438. However, in the April 2014 experiment the colonisation of NF443 is highly variable (12.7±16.2%), which complicates describing the phenotype.

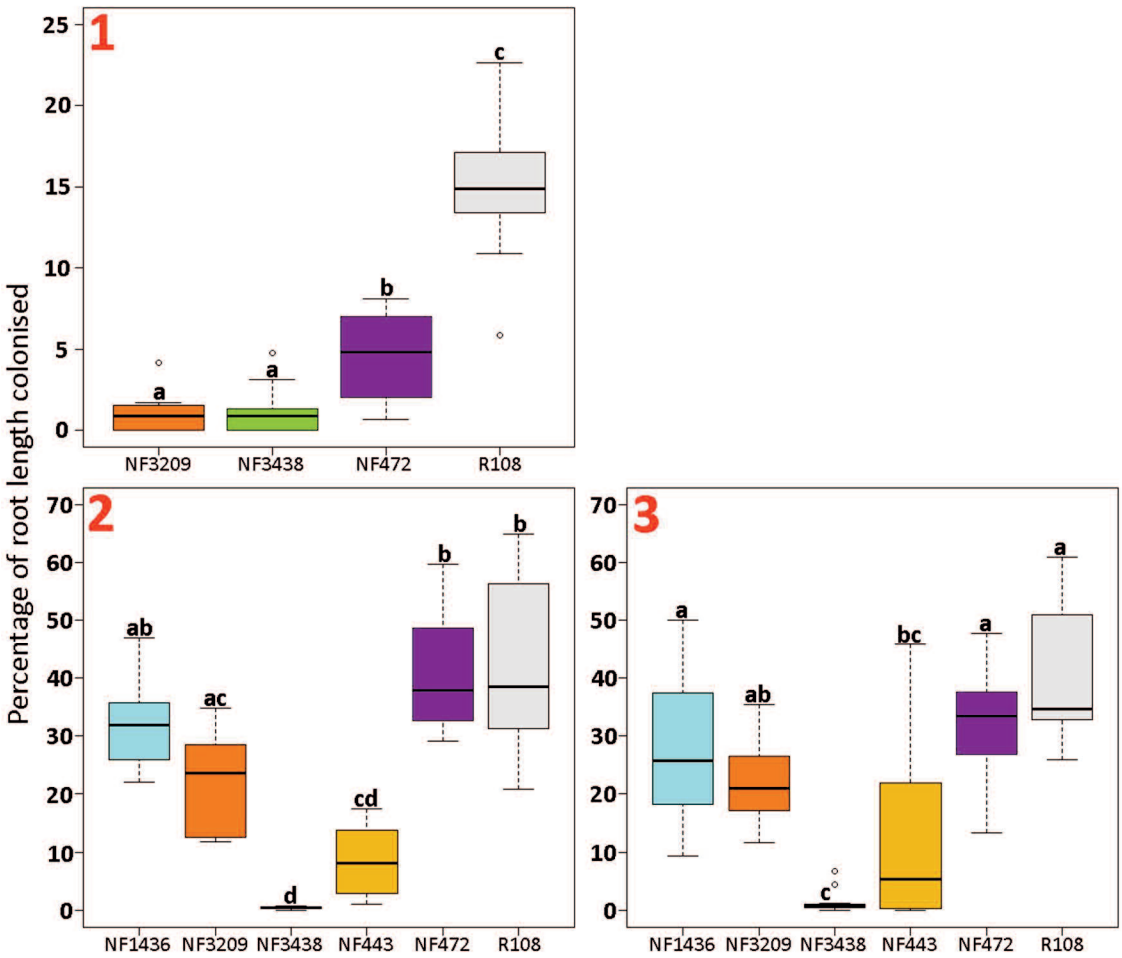


Figure 3.1 – Colonisation for five mutant lines
Percentage of root length containing fungal structures in five Noble Foundation lines identified as MYC$^-$ mutants in the preliminary screen. Plants were grown for five weeks Nov-Dec 2013 (Experiment 1), six weeks Feb-Mar 2014 (Experiment 2) and six weeks April-May 2014 (Experiment 3). Letters indicate significant differences (95% confidence by Kruskal-Wallis and Benjamini-Hochberg-adjusted Dunn test) with 12 (experiments 1 and 3) or 6 (experiment 2) replicates per genotype, each in a separate 80 ml (P24) pot.

Comparison between the three experiments are made harder by the different levels of colonisation observed in the wild type. To better compare the results from the different experiments, the data were normalised, such that the $\bar{x}$ of the wild type control was 1 (**Figure 3.2**). In the normalised data, the February and April 2014 experiments (2 and 3) are essentially the same, but the data from the November 2013 experiment (1) shows considerable difference, with NF3209 and NF472 showing much less colonisation, and NF3438 more colonisation, than in subsequent experiments. For NF3438, this relatively high colonisation in the normalised data is likely an artefact of low colonisation of the wild type plants in experiment 1. In terms of raw percentage root colonisation, NF3438 plants in experiment 1 ranged from 0% (5 of the 12 plants tested) to 4.8%, which is consistent with most subsequent work performed on the line.

For line NF3209, the difference between experiments 1 and 2/3 can be explained as part of the delayed colonisation phenotype, as subsequent work (**Section 3.3.3**) repeatedly showed that NF3209 plants, if given a sufficiently long growth period, or a strong enough inoculum source, will reach a peak level of colonisation around that of the R108 wild type. This peak colonisation level is probably slightly lower than the wild type, but given the high variability of colonisation, any decrease was not significant. Earlier in the growth cycle (around 4-5 weeks with the stronger stock pot inocula) colonisation was significantly lower than in the wild type. The reason for the NF472 results in experiment 1 is unclear, as this difference from wild type was never recapitulated, and close examination of arbuscules (**Section 3.3.2**) showed no difference from wild type.

NF1436 and NF443, the two lines added in the February and April 2014 experiments, showed a consistent phenotype between the experiments with a roughly 1/3 and 3/4 (respectively) reduction in colonisation relative to the wild type. However, they were both more variable in the April 2014 experiment than in February 2014, especially NF443.

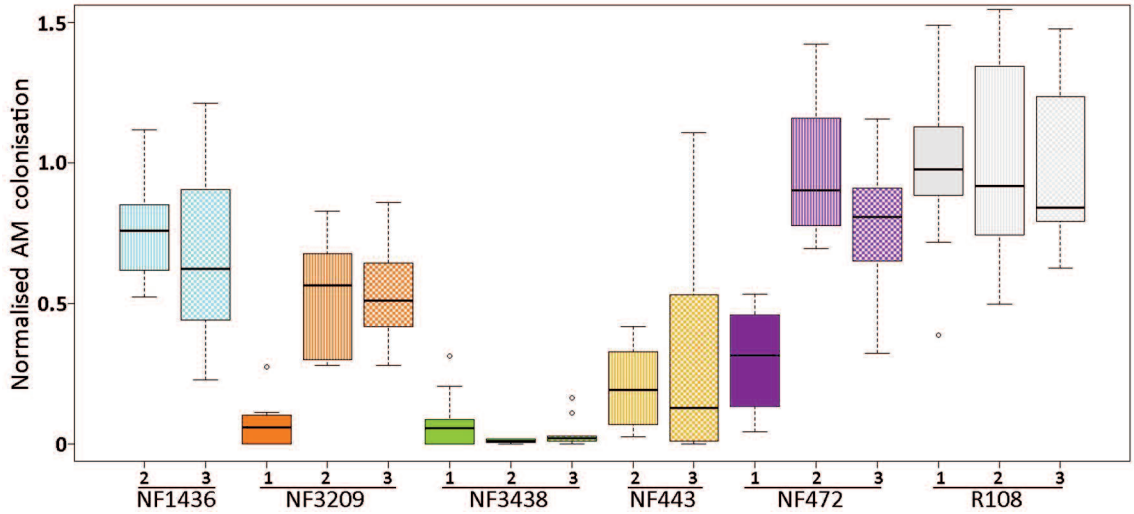


Figure 3.2 – Normalised colonisation for five mutant lines
Root colonisation data from experiment 1(November 2013), experiment 2 (February 2014) and experiment 3 (April 2014), with the percentage of root length containing fungal structures of each plant normalised, such that the $\bar{x}$ of the R108 wild type control of each experiment is equal to 1. 12 (**1**&**3**) or 6 (**2**) replicates per genotype.

Examination of the raw data for April 2014 shows the wide variance observed in NF443 colonisation is due to a clear non-normal distribution of the colonisation of the replicates. The majority of replicates (9 of 12) cluster below 11%, and three replicates around 40% (**Figure 3.3**). The reason for these outlying points is unclear, but they may have been accidently mislabelled wild-type seed. These three experiments were all performed before we had produced selfed seed from parents with verified mutant phenotypes, and there were occasions where some of the seeds planted as the R108 control were found to be A17 (which can be visually identified by its more prominent leaf chevron). If a similar mix up had occurred with the source of NF443 seed as well, it could explain why three individuals from this experiment gave a wild type phenotype, which was not subsequently recapitulated.

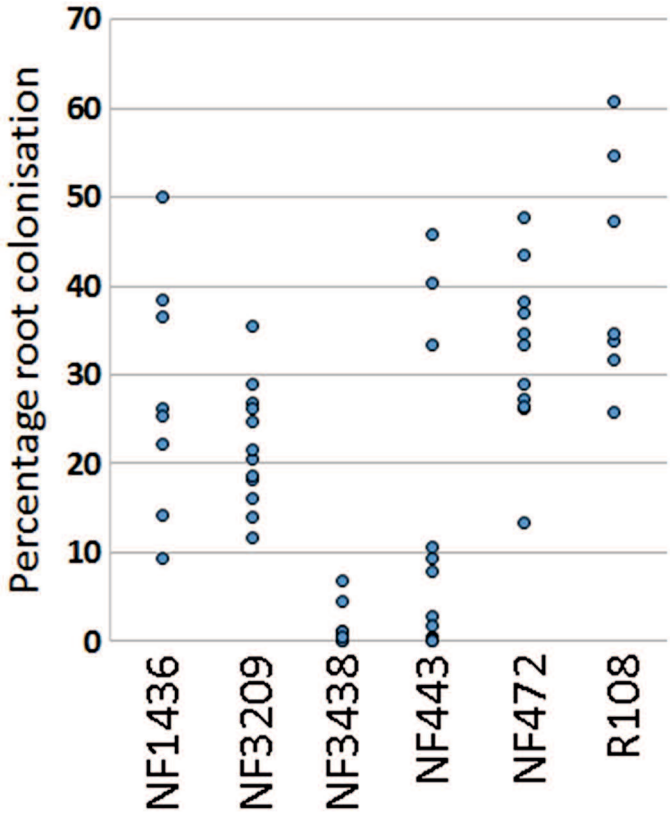


Figure 3.3 – NF443 colonisation from experiment 3 is not normally distributed
Data from **Figure 3.1**(**3**) plotted as a dot-plot to demonstrate non-normal distribution of NF443 data.

If these three outliers are removed (**Figure 3.4**), the mean of NF443 colonisation for the April 2014 experiment becomes 3.7±4.1% rather than 12.7±16.2%, and it is significantly different from all other lines except NF3438. This brings it into line with subsequent experiments, where NF443 consistently showed a phenotype between that of NF3438 and NF3209, with around 1/4 of wild type colonisation.

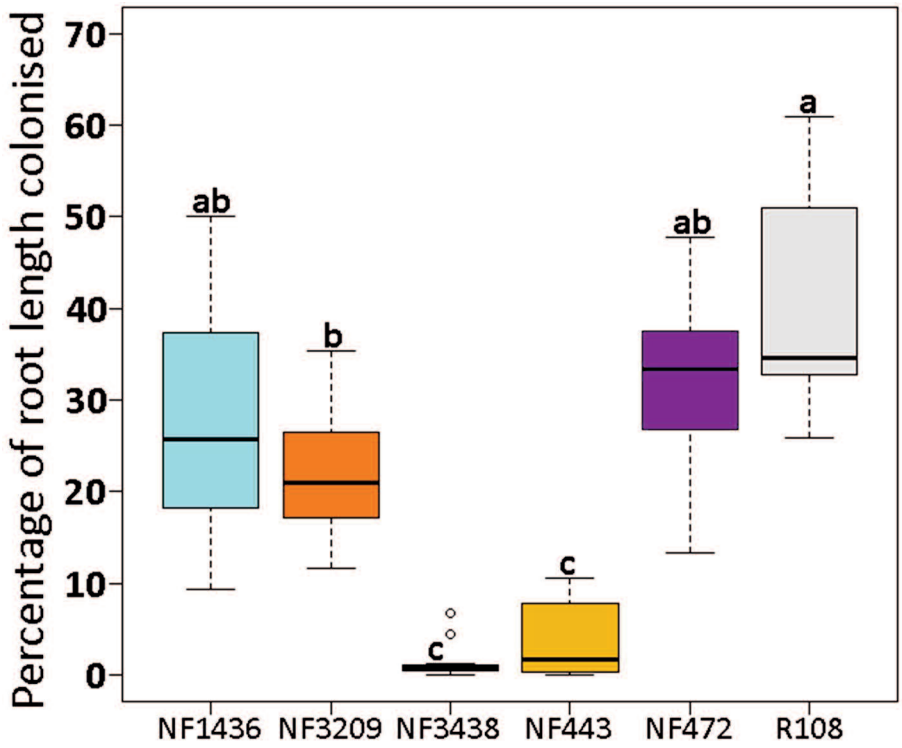


Figure 3.4 – Correction for non-normal data
Data from **Figure 3.1**(**3**) with three NF443 individuals that exceeded 30% colonisation removed. Letters indicate significant differences (95% confidence by Kruskal-Wallis and Benjamini-Hochberg-adjusted Dunn test), with 12 replicates per genotype.

Since the results shown in this section are percentage data (and are therefore constrained to within values 0-100), the low colonisation mutants are vulnerable to skewing into a log normal distribution, which breaks the default assumption of ANOVA modelling (normal data). To account for this, we tested an arc-sin transformation of the data to linearize it, then applying a linear model, or using a non-parametric Kruskal-Wallis test. Both methods gave similar answers, so the data presented above uses Kruskal-Wallis for ease of presentation.

3.3.2 – Arbuscule structure

AM colonisation of the wild-type R108 plants was quite variable throughout the experiments. While the overall colonisation varied, the R108 plants all developed large sectors of colonisation spreading laterally along the root from the appressoria (**Figure 3.5**). Arbuscules were large and cell filling, forming neat files along the axis of growth. With confocal microscopy, we observed a series of developing arbuscules (**Figure 3.6 B**), and the expected pattern of dense colonisation, with arbuscules filling cells of the cortex, but no fungal intrusion into the vasculature (**Figure 3.6 A**).

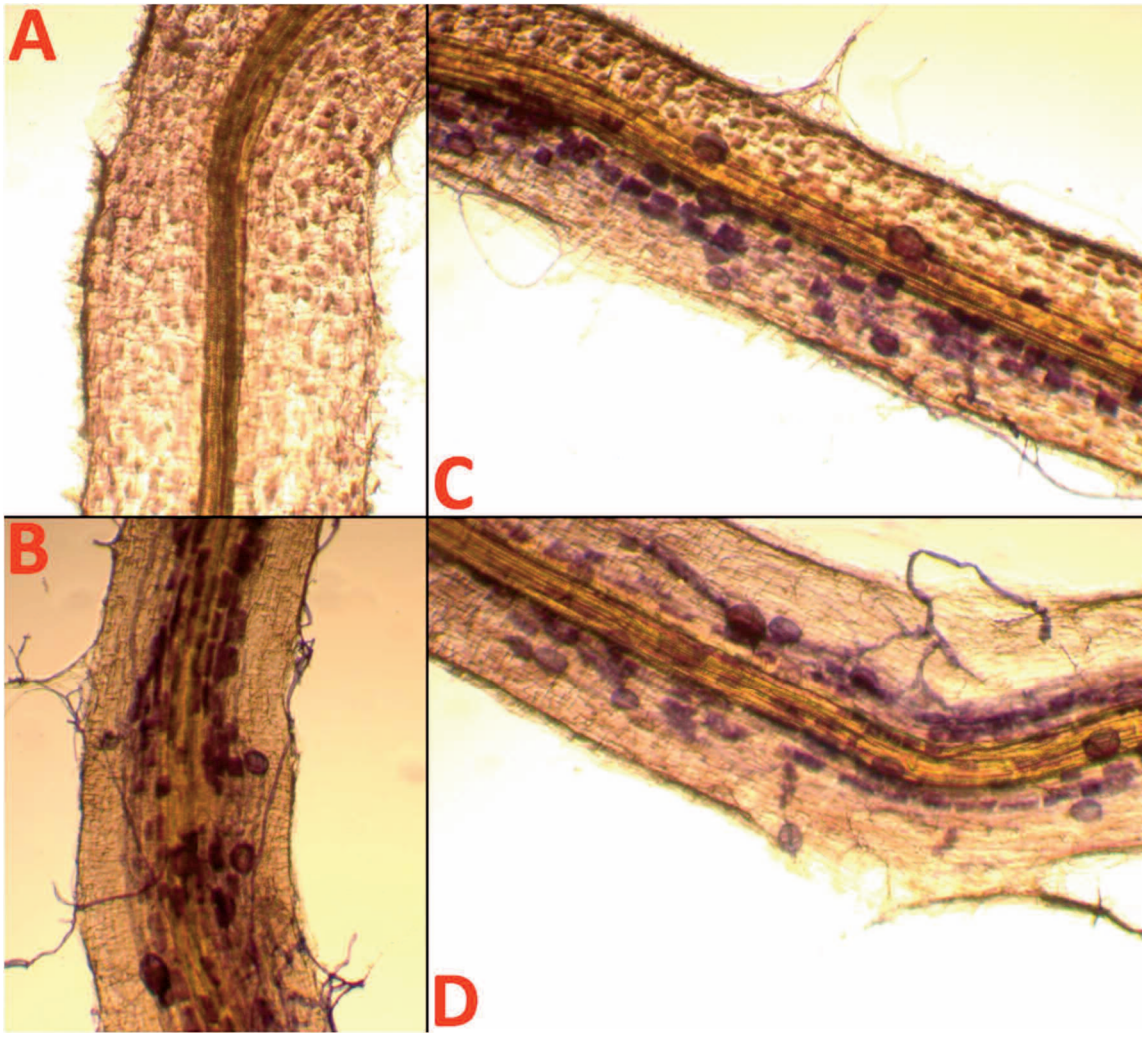


Figure 3.5 – Colonised wild type root
Ink/acetic acid stained root sections of wild type (R108) roots grown without AMF (A) or with PlantWorks four species inoculum (B-D). All visualised by light field microscopy at 10x magnification. See **Section 2.5.1** for the method.

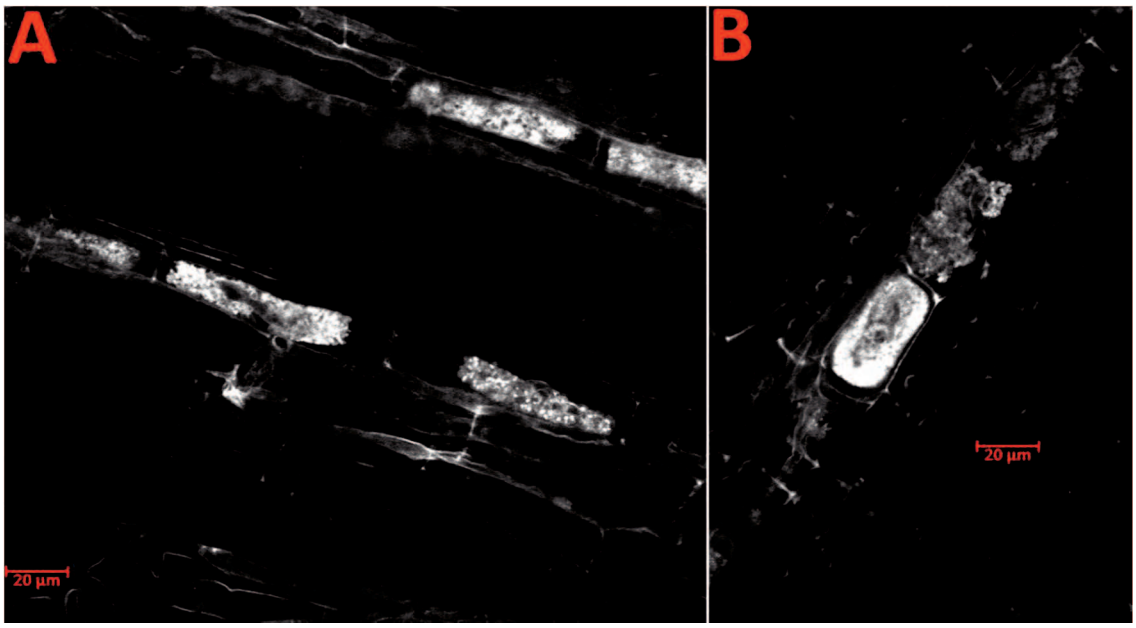


Figure 3.6 – Structure of arbuscules in the wild type plant
R108 roots colonised by *Rhizophagus irregularis* and stained with fungal cell wall binding agent wheat germ agglutinin conjugated to 488 nm excited Alexa-Fluor dye (WGA-AF488; see **Section 2.5.2**) and visualised with a confocal microscope. Image (A) shows a longitudinal section of the root, from root hairs on the root surface (bottom), through the cortex (with strongly staining fungal arbuscules), then the uncolonised vasculature (blank middle) and the colonised cortex on the other side (top). Image (B) shows a series of 3 arbuscules at different stages of development, from the thick branching (far right) through to mature (middle of image). Scale bars are 20 µm.

NF443 showed a large reduction in overall colonisation relative to the wild type. Appressoria are rare, and root internal colonisation was largely just hyphae (**Figure 3.7** and **3.8 A**) and a few small vesicles. We saw little evidence of arbusculation in colonised root sections, and no evidence of fungal sporulation on this plant genotype. The few arbuscules that did form did not develop past the bird's foot stage (**Figure 3.8 B**), lacking any fine branching.

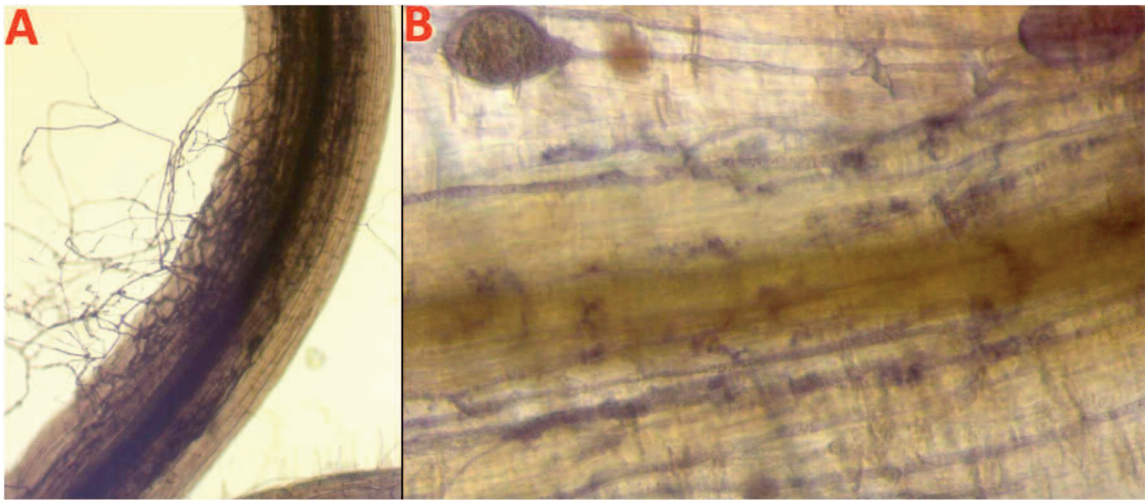


Figure 3.7 – Colonised NF443 root
Ink/acetic acid stained sections of NF443 roots grown with leek stock pot cultivated *Rhizophagus irregularis* inoculum. 4x magnification (A) and 20x magnification (B).

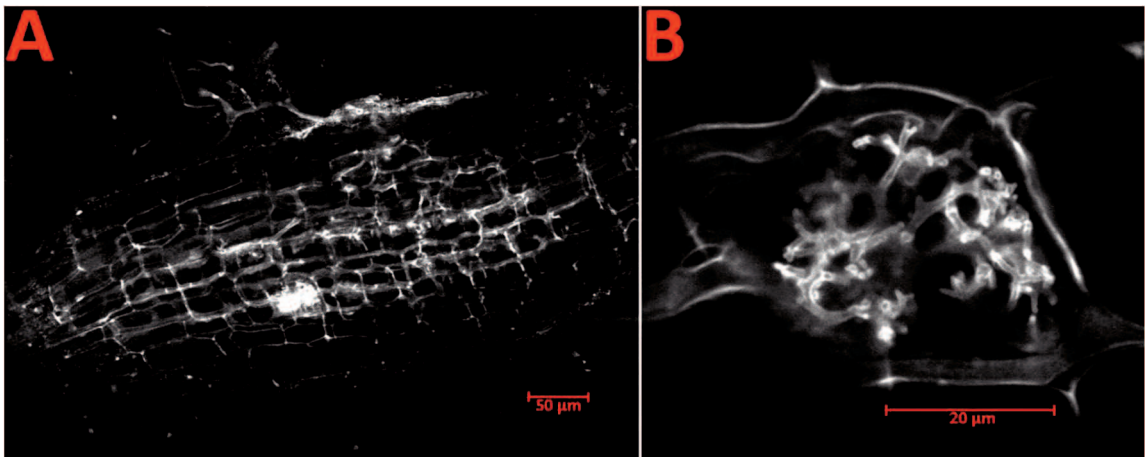


Figure 3.8 – Structure of arbuscules in NF443
NF443 root colonised by *Rhizophagus irregularis* (A) showing appressoria and hyphal spread throughput the root, with characteristic lack of mature arbuscules. Image (B) shows an individual arbuscule, illustrating the cessation of arbuscule development at the bird's foot stage. Stained with WGA-AF488. Scale bars are 50 µm (A) and 20 µm (B).

NF472 plants were well colonised, and showed arbuscule development soon after root penetration by the appressoria (**Figure 3.9**), unlike in NF443. Arbuscules formed a dense, cell filling file along the horizontal axis of hyphal growth (**Figure 3.9** and **3.10 A**), and plenty of vesicles and spores were observed, indicating successful carbon uptake. The arbuscules appear wild type, as they were large, square and cell filling (**Figure 3.10 B**), and we clearly observed the detail of the fine branches.

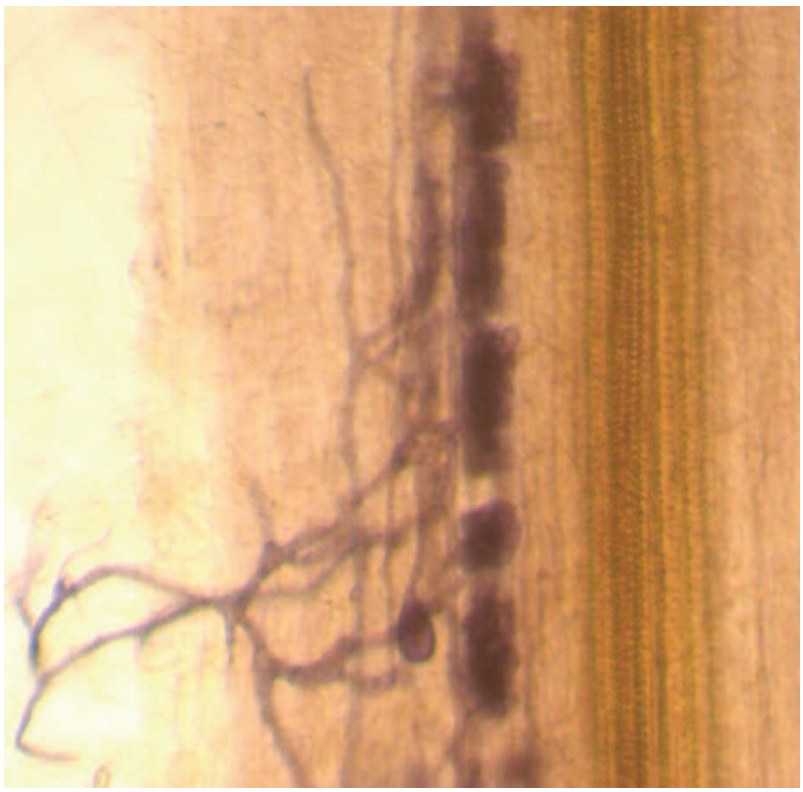


Figure 3.9 – Colonised NF472 root
Ink/acetic acid stained sections of NF472 roots grown with PlantWorks four species inoculum, 40x magnification.

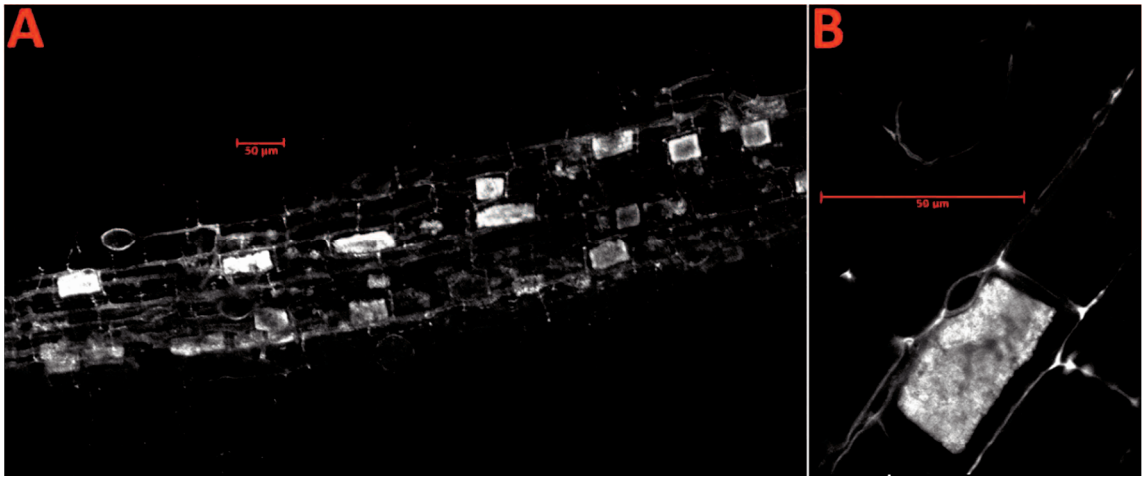


Figure 3.10 – Structure of arbuscules in NF472
NF472 root colonised by *Rhizophagus irregularis* (A) and individual arbuscule (lower B) and vesicle (upper B). Stained with WGA-AF488. Scale bars are 50 µm.

Overall root colonisation levels in NF1436 were largely indistinguishable from the wild type (**Figure 3.11 A** and **Figure 3.12 A**). Many appressoria form, and fungi spread laterally through the root. However, arbuscule formation was patchier, with some sections (**Figure 3.11 B**) lacking arbuscules. On close examination of the sections containing files of arbuscules, these arbuscules were largely degenerate, with bulbous central trunks and poorly organised branching (**Figure 3.12 B**), rather than the neat cell filling arbuscules seen in R108 or NF472 (**Figure 3.6** and **Figure 3.10**).

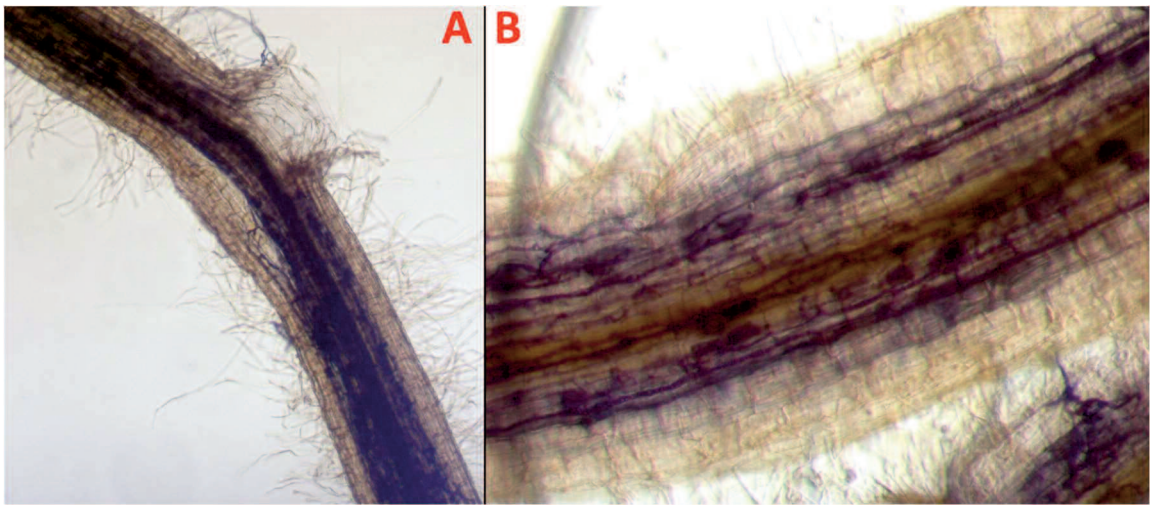


Figure 3.11 – Colonisation of NF1436
Ink/acetic acid stained sections of NF1436 roots grown with leek stock pot cultivated *Rhizophagus irregularis* inoculum. 4x magnification (A) and 10x magnification (B).

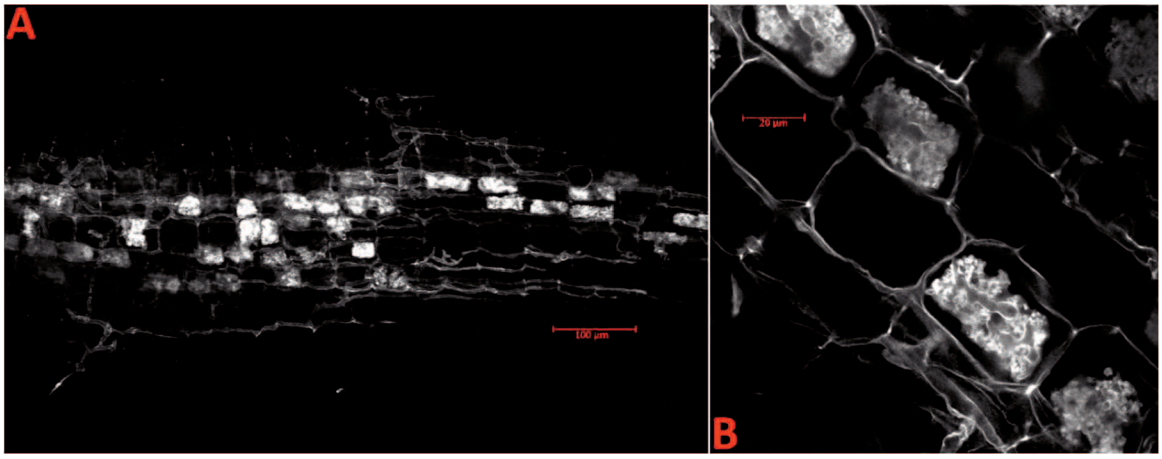


Figure 3.12 – Structure of arbuscules in NF1436
NF1436 root colonised by *Rhizophagus irregularis* (A) and individual arbuscules (B). Stained with WGA-AF488. Scale bars are 100 µm (A) or 20 µm (B).

The phenotype of NF3209 was harder to quantify. Colonisation events were rarer than in the wild type, and the colonisation level differed more between plants than in the other mutant lines. Many of these colonisation events showed little development, but others formed dense files of cell filling arbuscules (**Figure 3.13**), similar to that expected in the wild type (**Figure 3.6**). Even in the less developed regions, vesicles were observed, indicating a substantial carbon uptake by the fungus. Sporulation occurred on these lines, supporting the successful gain of carbon. Close observation of the arbuscules (**Figure 3.14**) suggested a lower number of colonised sections than the wild type, but little difference in development, other than a slight loss in fine branching.

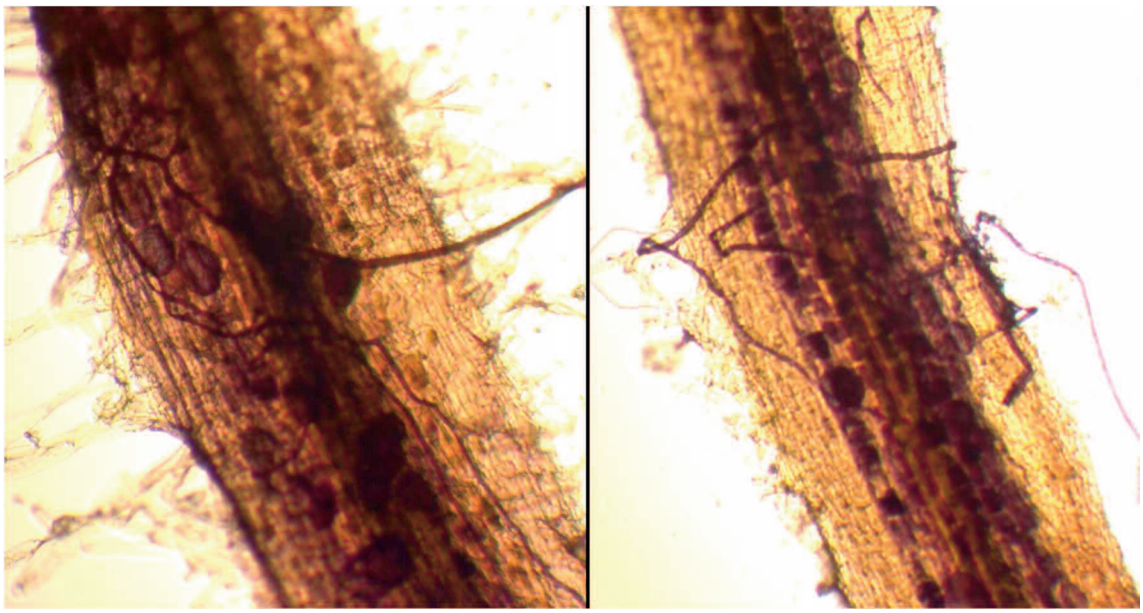


Figure 3.13 – Colonised NF3209 root
Ink/acetic acid stained sections of NF3209 roots grown with PlantWorks four species inoculum, 10x magnification.

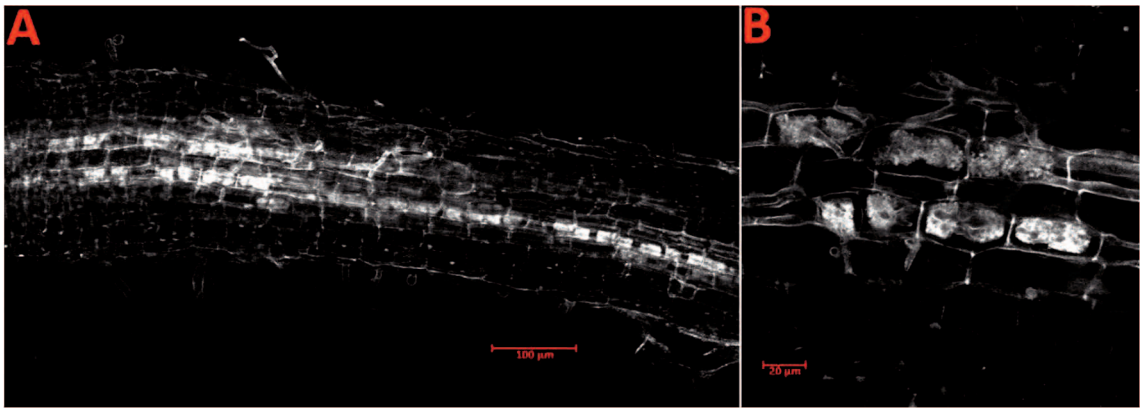


Figure 3.14 – Structure of arbuscules in NF3209
NF3209 root colonised by *Rhizophagus irregularis* (A) and closer detail on arbuscules (B). Stained with WGA-AF488. Scale bars are 100 µm (A) and 20 µm (B).

Even at high inoculum pressure root penetration events were very rare in NF3438. In those root sections where penetration did occur (see **Figure 3.15** and **3.16 A**), the hyphae spread through the cortex, and penetrated cells, but failed to form any defined arbuscules, leaving only a mass of disordered fungal wall material (stained in **Figure 3.16 B**). This confirms NF3438 as showing the strongest reduction in colonisation of our mutant lines.

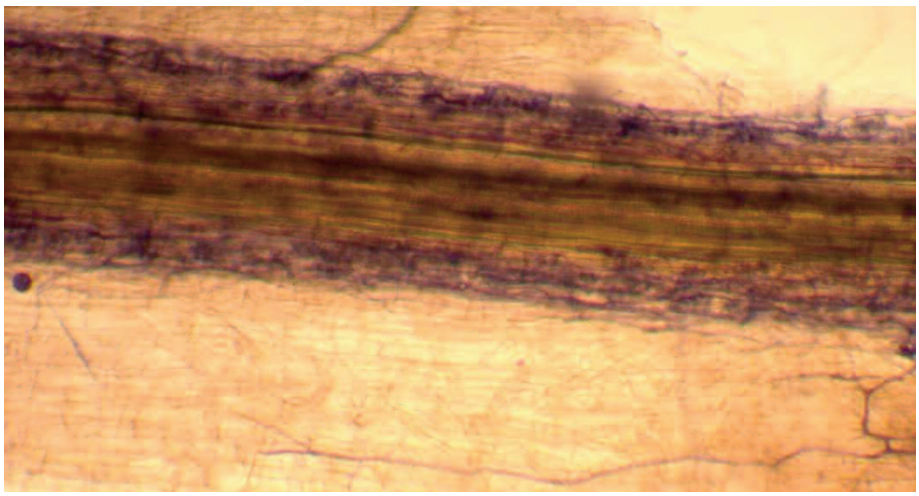


Figure 3.15 – Colonised NF3438 root
Ink/acetic acid stained sections of NF3438 roots grown with PlantWorks four species inoculum, 10x magnification.

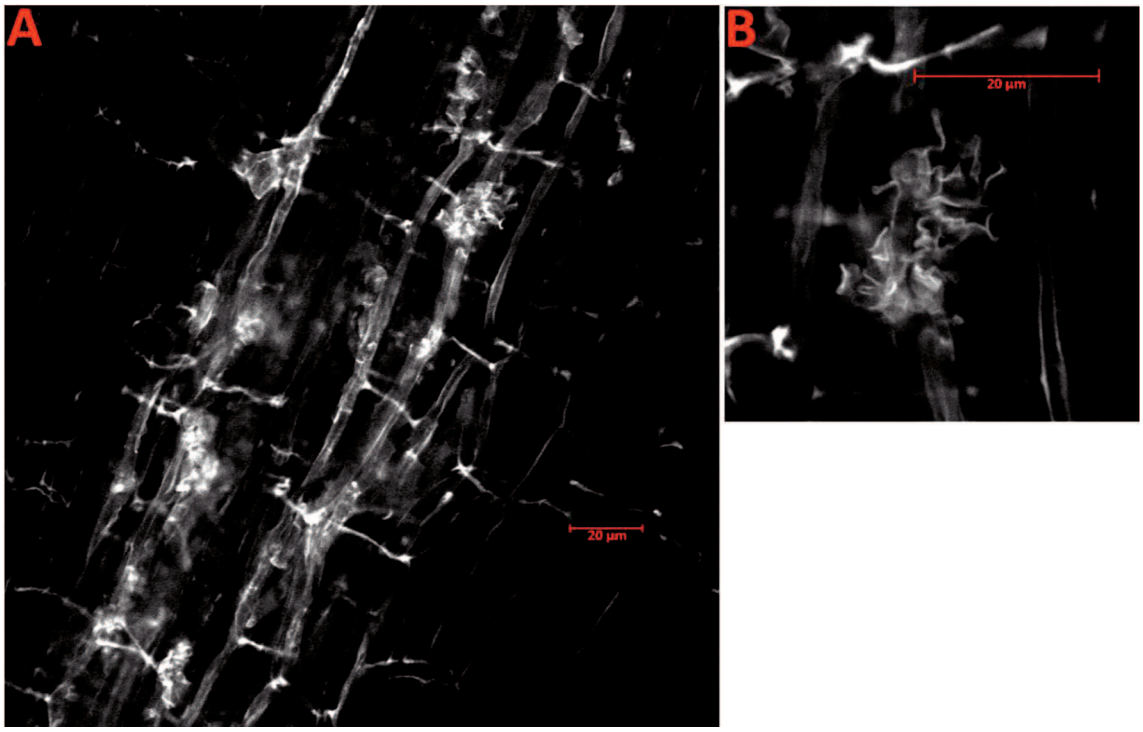


Figure 3.16 – Structure of arbuscules in NF3438
NF3438 root colonised by *Rhizophagus irregularis* (A) and individual arbuscule (B). Stained with WGA-AF488. Scale bars are 20 µm.

3.3.3 – Reduced colonisation or delayed colonisation?

A key problem from the data presented in **Section 3.3.1** was the lack of consistency in some of the mutant phenotypes, most especially in NF3209. While repeatability was an issue throughout the study, we argue that when taken as a whole, the data on NF3209 showed a consistent phenotype. The NF3209 phenotype is best described as a delay in colonisation with somewhat malformed arbuscules. Over a short inoculation period, the phenotype was a severe reduction in colonisation compared to a wild type,

but as the inoculum strength increased, the phenotype regressed towards that of the wild type in terms of total colonisation. A greater proportion of NF3209 arbuscules lack fine elaboration, even in well colonised roots.

Inoculation strength has many variables, such as growth period, number of infectious particles and nutrient availability, and presumably other environmental conditions like temperature. In **Chapter 3**, growth time is consistent with inoculation time, since all seedlings are planted directly into inoculated media. Taking data obtained from the initial mutant phenotyping screens, and the controls for the F_2 segregation screens (**Chapter 5**), a significant and consistent relative colonisation phenotype over time is revealed for NF3209 (**Figure 3.17**). At 5 weeks, NF3209 plants showed only ~10% of the level of AM colonisation of the wild type, but by 7 weeks, they are almost indistinguishable by total colonisation.

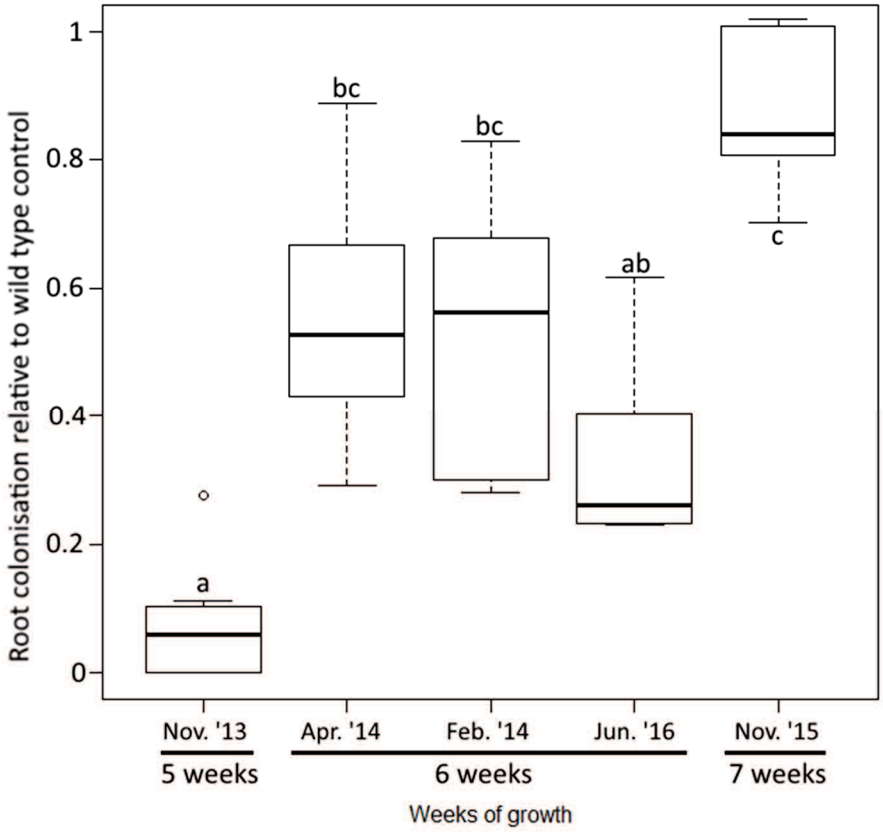


Figure 3.17 – NF3209 relative colonisation increases with growth time
Colonisation data for NF3209 plants grown for different lengths of time, with total AM colonisation normalised to a R108 wild-type control grown side by side. Letters indicate significant differences (95% confidence by Kruskal-Wallis and Benjamini-Hochberg-adjusted Dunn test).

Does the same delayed colonisation phenotype hold true for NF472? This phenotype is harder to ascribe, since we lack any data from a third time period for this line. The large decrease in relative colonisation seen in the 5 week long November 2013 screen appears compelling, but was never recapitulated. An attempt was made to produce a proper time course screen, but the problems with the PlantWorks inoculum meant that the wild type control was never consistently colonised. However, even before 5 week old, NF472 plants were never noticeably different in colonisation than R108. Combined with the wild type appearance of the arbuscules (**Section 3.3.2**), NF472 cannot be considered as presenting a mycorrhizal phenotype. Therefore, it was dropped from further analysis in this project.

3.3.4 – Does reduced colonisation impair plant growth?

In theory, the AMS in a P starved environment should improve the P nutrition of plants, the limiting factor on its growth, and thus increase the size of the plant. A mutant with a reduced amount of AM colonisation was expected to have a lower surface area to exchange nutrients, thus expected to show less of a growth benefit from the symbiosis. However, there is a wide body of literature[288,289] showing growth benefits, no growth change but improved nutrition, no effect and even growth penalties from association with AMF, with results presumably depending on environmental conditions and genotype x genotype interactions between the symbionts, and few of these effects are yet predictable.

To discover what, if any, benefit the mutant *Medicago* plants would fail to obtain from the AMS, several experiments were performed. Three mutant lines were grown alongside the R108 wild type, in media supplemented with bone meal, which should be more accessible to the AM fungi than the plant[290]. There was no significant difference in fresh weight or root:shoot ratio between the mutant and the wild type (**Figure 3.18**). Since no non-mycorrhizal control was grown, we cannot say if there was a change relative to that, but if there was a growth benefit to the wild type, it was the same as in the mutants. The NF3438 line was, as always, barely colonised, so we would expect that it gain no benefit from the AMS. Therefore, we did not observe a significant growth benefit under these conditions for the wild type plants due to AM colonisation.

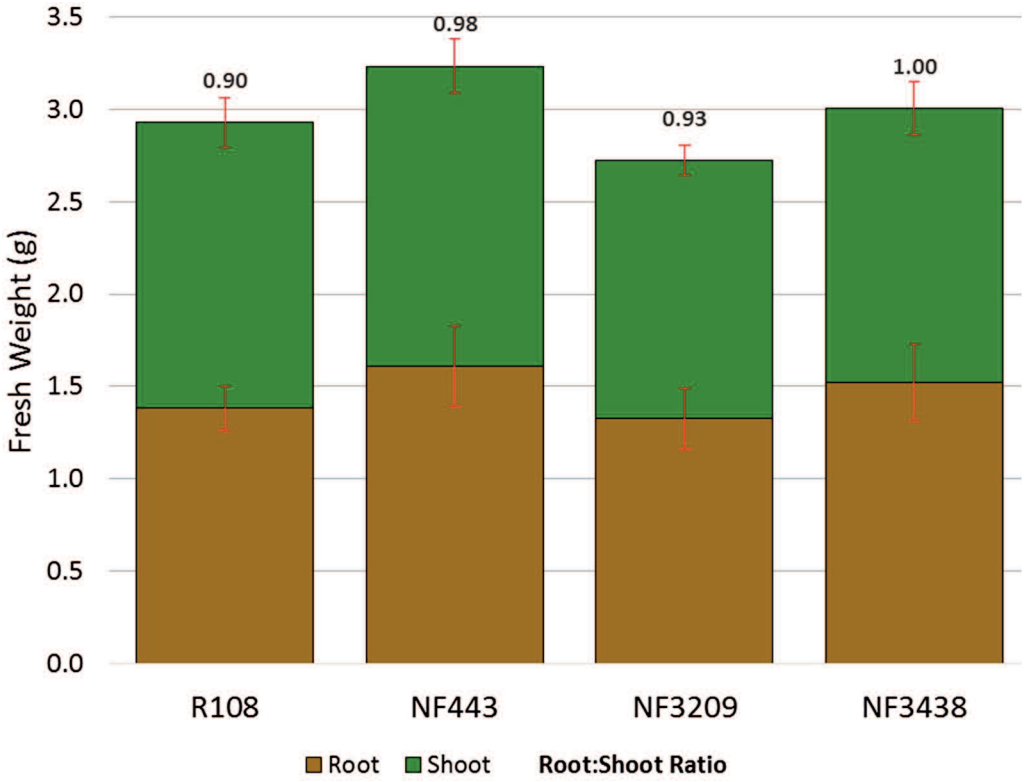


Figure 3.18 – Plant mutants have no effect on growth at 6 wpi
Fresh weight of plants grown for 6 weeks (May 2015) with R no P fertiliser and bone meal supplementation in sand:Terragreen, with Leek stock pot *Rhizophagus irregularis* inoculum.

Another attempt to elucidate a growth effect of colonisation had been made with a competition experiment. We hypothesised that if multiple plants were made to compete for the same limited P supply, then the benefit of the AM symbiosis might be amplified, as it would give a competitive advantage to those plants able to form the symbiosis, as they could exploit a larger substrate volume. If single pot experiments did not show an effect because the plant could exploit the entire volume, then larger areas should increase dependence on the AMF.

24 plants (half R108 and half mutant) were planted in a 6x4 grid in a large, shallow tray, in randomly determined positions. They were treated weekly with R solution, and the substrate supplemented with bone meal. Two trays per mutant were grown, one with and one without PlantWorks inoculum (at 20 times the recommended concentration). The roots had grown together into an intermingled mass that defied attempts to separate it by plant, so only shoot mass data is presented, and genotype colonisation data remained unknown.

Adding inoculum led to a non-significant increase in R108 fresh weight relative to NF3209, but no effect in dry weight (**Figure 3.19**), and the opposite effect on R108 and NF443, with an increase in both (but NF443 slightly more) and a significant increase in both fresh weight ($p<0.01$; ANOVA) and dry weight ($p<0.001$). A linear model suggested that while inoculum addition led to a significant increase in weight overall, there was no significant effect of genotype or the genotype x AM treatment interaction.

This might suggests the wild type plants are driving a nurse plant effect, supporting mycorrhizal colonisation and phosphate delivery to the reduced colonisation mutant NF443, rather than gaining preferentially Pi delivery from the fungus. However, a confounding effect related to extra nutrient availability due to the addition of inoculum cannot be ruled out, as no autoclaved inoculum control was used.

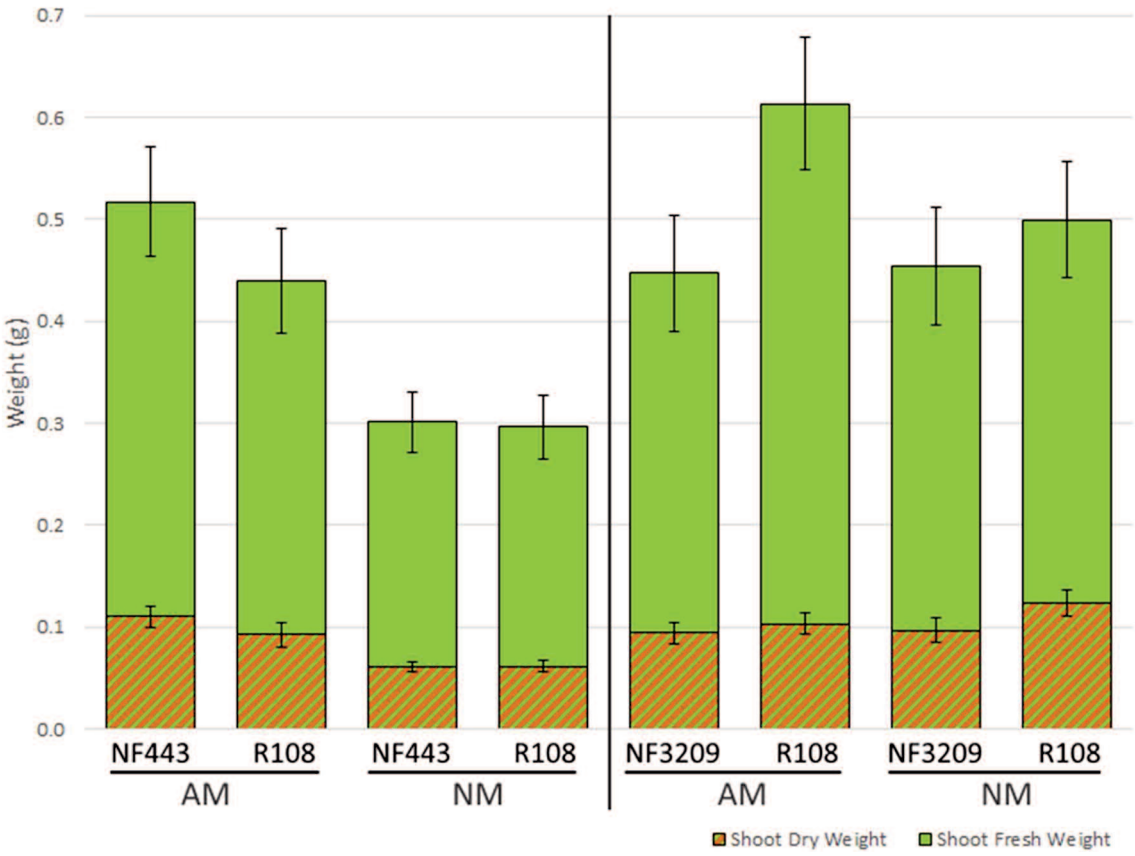


Figure 3.19 – Mycorrhizal inoculation has a significantly effect on plant competition in some lines
Average shoot weights for 6 week old plants (12 R108 and 12 mutant; March-April 2014) grown in a common bone meal supplemented substrate and either with PlantWorks four species inoculum (AM) or without any AMF (NM). Fresh weight includes the 'dry weight' category and error bars show standard error.

The difference in effect on NF3209, which showed a less extreme mycorrhizal phenotype, and the preliminary nature of this experiment (with only one biological replicate) make it hard to draw a reliable conclusion from this data. Since no consistent difference was seen between these experiments, or between the mutants in the various F_2 screens (see **Chapter 5**) undertaken concurrently, it was decided to pursue more productive research paths, rather than try to nail down an effect that was likely very condition dependent.

3.3.5 – Changes in gene expression in the mutant lines

qPCR was used to assess the expression of PT4 and HA1 in the four mutant lines. These proteins are localised to the periarbuscular membrane and drive symbiotic phosphate uptake (see **Section 1.5.2**), and are often used as markers of a successful AMS[73,130,291,292]. We would expect *HA1* and *PT4* levels to correlate with the number and development of arbuscules in our mutants (highest in the wild type, then in NF1436, NF3209 and NF443, with NF3438 being the lowest).

NF443 and NF3438, the stronger MYC mutants, showed very low levels of *HA1* and *PT4* expression, giving little genetic sign that AMF were present (**Figure 3.20**). This suggests that no symbiotic phosphate exchange is occurring in these lines. NF3209 showed a decrease in expression of both genes, although not to the same extent as NF443 and NF3438, and showed much greater in-sample variation. This fits with the delay in colonisation phenotype expected for the line. With the reduced plant Pi uptake suggested by the expression data, we would expect the plant to deliver less C to the fungus, slowing fungal growth. In NF1436 we see no significant difference in *PT4* expression, and a non-significant increase in *HA1* expression that likely does not have much of a physiological impact.

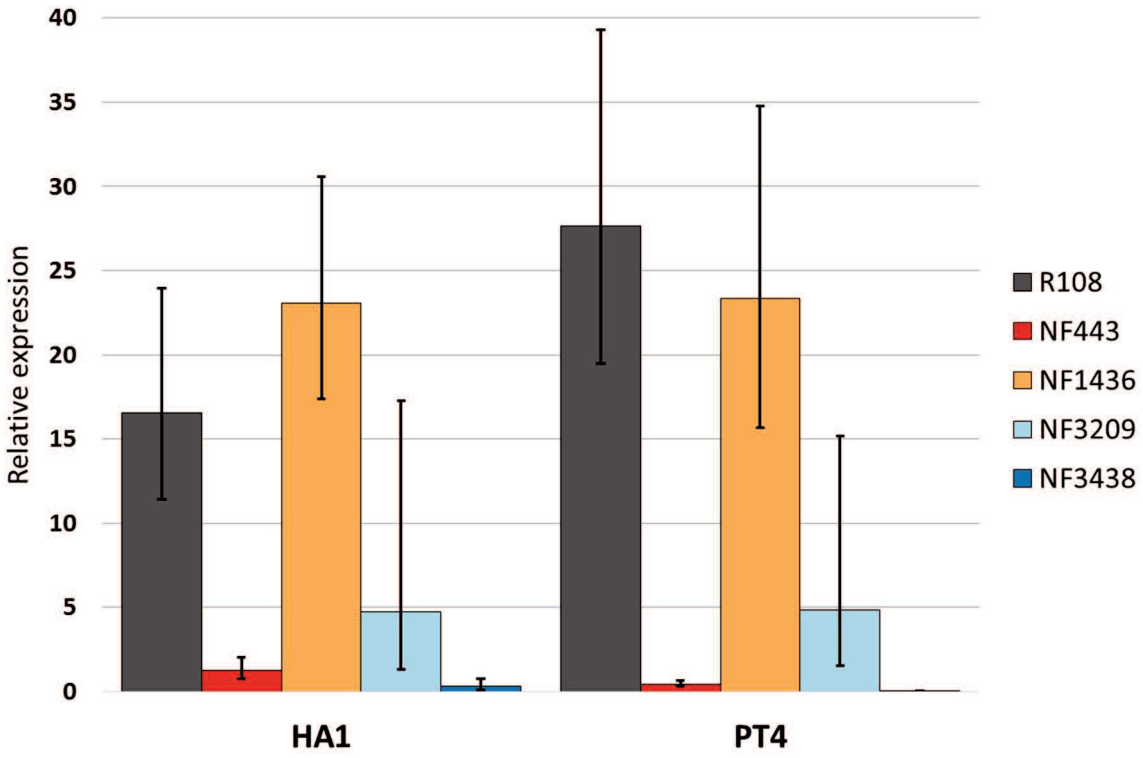


Figure 3.20 – Expression of genetic markers of nutrient exchange in the four mutant lines
Gene expression of mycorrhizal associated genes in the bulk roots of four *Medicago truncatula* mutant lines and the wild-type control (R108), grown with AMF (*Rhizophagus irregularis*). Three biological repeats per line, each from a single plant (except NF443, which has 2 biological repeats). The expression was normalised to the expression of housekeeping gene RNA Helicase 1[293]. Error bars show standard error.

3.4 – Discussion

3.4.1 – The mutant phenotypes

The data presented above allow us to describe in detail the mycorrhizal phenotype of each of the mutant lines studied.

First, we consider NF472 to have been a false positive from the preliminary screen. We have seen no significant deviation from the wild type in our work. NF1436 is a more complex issue. It shows a small and non-significant decrease in overall colonisation, and wild type *PT4* expression, although close visual inspection suggests an increased proportion of degenerating and/or malformed arbuscules. Thus, the mutation is most likely to be in a gene involved in the arbuscule degradation pathway, reducing the lifespan of the otherwise productive arbuscule in the knock-out.

NF3438 has a strong mycorrhizal phenotype with near complete abolition of colonisation, and lacks functional arbuscules. NF443 shows another clear reduced colonisation phenotype, with a considerable reduction in overall colonisation, and the lack of *PT4* expression suggests there is no symbiotic phosphate transfer. NF3209 also showed a significant change in fungal colonisation, and showed some reduction in *PT4* expression at 5 wpi. However, repeat experiments showed a variable phenotype (**Section 3.3.4**), which suggests that it represents a delayed colonisation phenotype.

Overall, most of the lines identified in the prior work[272] and examined in this study have been confirmed to show mycorrhizal phenotypes. A considerable range of different phenotypes have been found, and these can be used to suggest plausible roles for candidate genes in the symbiosis. In subsequent chapters, we present our efforts to identify the tnt1 insertions in these lines, and use co-segregation to discover which insertion is driving our observed phenotypes.

3.4.2 – Comparison of different methods for measuring colonisation

qPCR to examine expression of genes encoding arbuscule-localised proteins, primarily *PT4*, has often been used as a measure of colonisation success[73,130,291,292]. We assume that without PT4, a plant cannot take up Pi via the symbiotic pathway, and thus low levels of *PT4* indicate a deficiency in the AMS. From our results we see that this does not always give the same assessment of colonisation as visual measures of fungal abundance. NF443 has considerably more of its root length containing fungal structures, yet like NF3438, which has almost no visible colonisation, NF443 lacks expression of *PT4*. This likely indicates that while symbiosis can progress further than the block imposed in NF3438, it is probably also non-functional, at least as far as phosphate exchange is concerned. As we know (**Section 1.8**), there are many other benefits to the fungus' host beyond phosphate transfer, so a symbiosis that lacks PT4 may still be mutualistic. Likewise, a mutant may be able to express *PT4*, but be unable to gain phosphate for other reasons (although a mutant in *HA1*, another parts of the plants symbiotic Pi uptake pathway lead to downregulations of *PT4* expression[139]). NF3209 shows an intermediate AM colonisation phenotype, between NF443/NF3438 and wild type in both visual and genetic measures, supporting a tentative description of the mutation as being for a non-essential pro-symbiotic factor. NF1436 shows wild-

type expression of *PT4* and *HA1* despite a large proportion of degraded arbuscules present visually, suggesting there are mycorrhizal genes that will affect the success of the symbiosis without perturbing expression of the plant Pi uptake machinery. Thus, we would suggest that PT4 expression alone should not be used as a screen for mycorrhizal mutants, as it would likely miss many genes involved in the non-phosphate aspects of the symbiosis. It does however present a more accessible approach for estimating symbiotic Pi exchange than isotope tracing experiments.

Chapter 4 – Sequencing of *Medicago truncatula* lines exhibiting reduced AM colonisation

4.1 – Introduction

We have described four *Medicago truncatula* lines that were initially identified in a forward genetic screen, and show reduced mycorrhizal colonisation phenotypes. This screen was carried out to discover new genes involved in the mycorrhizal symbiosis. To complete this aim, we need to discover which gene is responsible for the phenotype in each of these lines. The Noble Foundation estimated these lines contain between 5 and 40 tnt1 insertions, with an average of 25 per line[276]. We planned to use co-segregation analysis to discover the causal insertion. Not all tnt1 insertions were expected to have an equal chance of causing the mycorrhizal colonisation phenotype. While preferential insertion of tnt1 into euchromatin regions is reported[272], we would expect not all tnt1 insertions to be near genes. Insertions away from genes would be unlikely to have any effect on the expression of these genes, and while an insertion into a promoter or intron might change the rate of expression, it is less likely to cause a phenotype change than an insertion that disrupted the coding sequence would (via insertion into an exon). Secondarily, once we had identified the gene likely disrupted by the tnt1 insertion, we could exclude those that are known not to be expressed in the root from the first pass of co-segregation genotyping, and prioritise genes that showed an increase in expression during the AMS.

In previous work[72,138,294] with the Noble Foundation's *M. truncatula* tnt1 collection[272], tnt1 insertions were located by sequencing of flanking sequence tags (FST) generated by thermal asymmetric interlaced PCR (tailPCR) or inverse PCR (invPCR). Other methods such as Southern blotting used to quantify the number of insertions present in each line. Both invPCR and tailPCR methodologies suffer from two drawbacks. First, DNA primers (tailPCR) or restriction enzymes (invPCR) require sequence specific binding to function. However, since the location of the tnt1 sequence is not yet known, we cannot predict what the target sequence needs to be. Therefore, degenerate primers with low stringency annealing conditions (in tailPCR) or multiple

sets of restriction enzymes (for invPCR) are used to target these unknown sequences. However, both methods lead to preferential amplification of certain genomic locations, and we cannot be certain that every insertion will contain sequence that can be identified by even these less stringent methods. The second problem arises from the fact that these methods may need a large excess of sampling to ensure all insertions are located. If we had sequenced the colony PCR amplicons produced by these methods completely at random, locating all the insertions in a line with 25 transposons would require analysing an average of 96 invPCR sequences[i]. We can reduce the need for this somewhat, as invPCR amplified fragments can be run on an agarose gel to separate them and give an idea of the number of insertions present. However, gel electrophoresis cannot distinguish between multiple fragments of the same approximate length. Additionally, fragment length depends on restriction enzymes or primers used, so experiments with different sets of enzymes or primers cannot be compared by fragment size alone. Preferential amplification of some sequences by PCR will further complicate attempts to find all the insertions. Finally, tail-PCR normally finds only FST from one side of the tnt1 insertion, but this should not be a major issue in a species like *M. truncatula* with a published genome[295]. So long as the amplified fragment contains one end of the tnt1 sequence, the location should be establishable to the base pair.

Significant work on insertion discovery using invPCR had been carried out on line NF3209 prior to this project, and to a lesser extent on the other lines in this study. When co-segregation analysis of NF3209 (described in **Chapter 5**) showed that none of the insertions located by invPCR were candidates to cause the mycorrhizal colonisation phenotype, we were faced with a dilemma. We could have continued with the invPCR efforts, but fragment analysis following invPCR (**Figure 4.2**) suggested that the previous work had located insertions corresponding to each of the different amplification bands with the restriction enzyme set used. Repeating the invPCR with different sets of enzymes would likely be extremely time consuming, as most discovered sequences would be repeats of those already found. We decided to take a different approach,

[i] Modelled as a coupon collector's problem (as per the 'simcollect' R package). We can further estimate that an average of 11 sequences would be needed to find all insertions in a line with 5, and 280 sequences in a line with 60 insertions.

taking advantage of the rapidly falling cost of next generation sequencing to locate tnt1 insertions by shotgun sequencing of our mutant lines. This method, while expensive, was likely to find all the possible tnt1 insertions if sequenced to enough depth to properly cover the genome. This was price competitive with PCR based methods when the factor of redundant or excess sequencing of PCR generated FSTs is taken into account. FSTs could then be located by simply looking for reads that contained both tnt1 and *Medicago* DNA. If necessary the genome could be assembled against the published genome to establish if we had fully sequenced the line. Additionally, the large reduction in lab work, and the fact we could automate much of the data processing if it was done routinely, made the whole genome sequencing (WGS) approach quicker than PCR based approaches for transposon discovery.

4.2 – Methods

4.2.1 – invPCR

InvPCR was used to discover the location of tnt1 insertions in *Medicago truncatula* lines obtained from the Noble Foundation (**Figure 4.1**). 1.5 µg of DNA was digested overnight at room temperature with 10 U of either EcoR1 and Mfe1, or Ase1 and Nde1, in a 25 µl solution with NEB Buffer 4 (EcoR1 & Mfe1) or 3 (Ase1 & Nde1). These pairs of restriction enzymes created fragments with compatible overhangs. A 5 µl sample was run on an agarose gel to assess the digestion. After heat inactivation of the restriction enzymes in the remaining 20 µl solution, the fragments were circularised by incubating the digested DNA sample overnight at room temperature with 3 U of T4 DNA Ligase (Thermo Scientific, MA, USA) in a solution made up to 100 µl in the T4 Ligation buffer. Two rounds of PCR with nested primers targeting sites near the end of the tnt1 transposon were used to amplify the plant DNA flanking the insertion. 10 µl of the ligated DNA sample was taken, and added to 0.5 U of Ex Taq proofreading polymerase (Takara Bio, Otsu, Japan), 5 picomoles of primers LTR3 & LTR5 (see **Appendix 2**), 4 nanomoles of each dNTP and made up to 20 µl in the provided polymerase buffer. PCR was performed as follows: after an initial 2 minute denaturing step at 94°C, 30 cycles of 20 seconds at 94°C, and then a 3 minutes annealing and extension step at 72°C were carried out, all followed by a final extension

step of 6 minutes at 72°C. 5 µl of this PCR product was run at 110 V for 40 minutes on a 1% agarose gel stained with SYBR Safe (Thermo Scientific). The remaining product was then diluted 100 fold with nuclease-free H_2O, and 10 µl of this diluted sample added to 0.5 U of Ex Taq proofreading polymerase, 5 picomoles of primers LTR4 & LTR6, 4 nanomoles of each dNTP and made up to 20 µl in the provided polymerase buffer. This second round of PCR used an initial 2 minute melting step, then 39 cycles of a 94°C, 20 second melting step, a 60°C, 20 second annealing step, and a 72°C, 3 minute extension step, all finished with a 5 minute final extension step. A 1% agarose gel was then run to visualise the invPCR2 product. Bands could then be cut out, amplified in *E. coli* and sequenced.

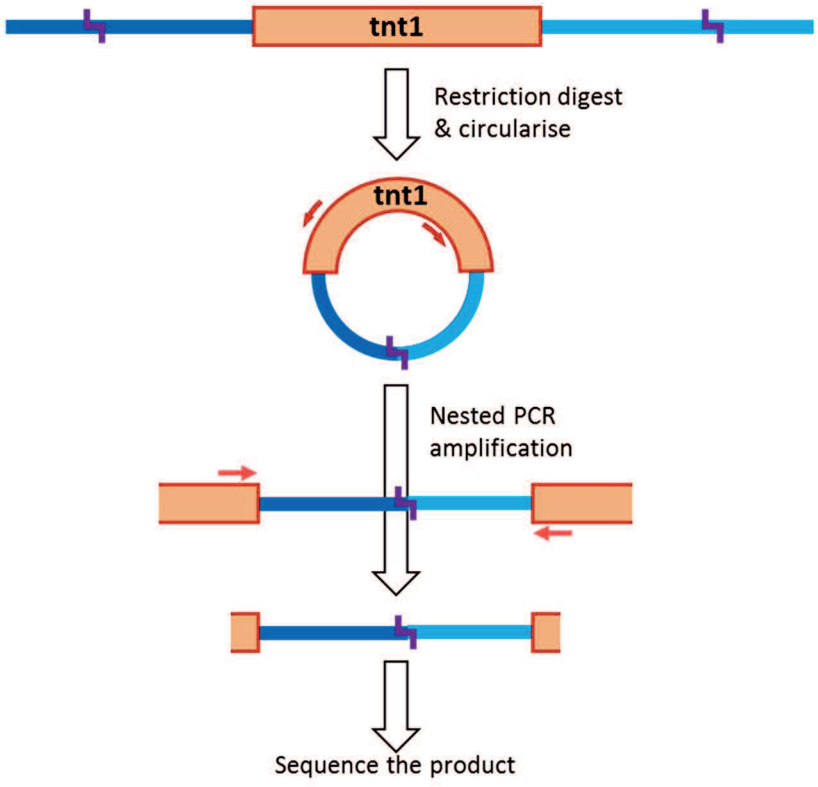


Figure 4.1 – Visualisation of the invPCR protocol
The *Medicago* gDNA is digested with a set of restriction enzymes with compatible recognition sequences (ϟ), then circularised. Two rounds of PCR with nested primers (dark red arrows followed by light red) complementary to the tnt1 flanking regions are used to amplify the FST from both sides of the tnt1 insertion, which can then be sequenced.

4.2.2 – WGS

DNA was extracted from a randomly chosen plant of each of four mutant lines, using a Qiagen DNeasy kit, according to the manufacturer's protocol. Samples with a concentration of >200 ng/µl were submitted for Illumina library preparation (500 bp fragments with 125 bp paired ends reads) and barcoding (to allow all samples to be run on a single lane) by the Technology Facility (Department of Biology, University of York, UK). Sequencing was carried out by Harvard Bauer Core Facilities (FAS Division of Science, Harvard University, Cambridge, MA, USA) using an Illumina HiSeq 2500.

We used next generation sequencing to sequence the genomes of four of our *Medicago truncatula* lines. We chose to use the Illumina dye sequencing-by-synthesis technology for this project as we judged that what the Illumina technology lacks in read length was more than made up for by the lower cost per Gb and its considerably better accuracy than the competing technologies (e.g. Ion Torrent)[296]. The lower read length would make the genome harder to assemble, but for our approach assembly was not essential, as we merely needed to find the sequence around the insertions. The *Medicago* genome was already available, so we could assemble our data against this if necessary.

Given the coverage promised by the Bauer Core Facility, and the estimated size (360 Mb) of the *M. truncatula* R108 genome[297], we estimated that we could sequence four R108-sized genomes to a depth of 10x on a single Illumina lane. This allowed us to sequence all four of the promising mutant lines (NF443, NF1436, NF3209, and NF3438). The total amount of sequence returned from the Bauer Core Facility was larger than expected, giving between 13x and 22x coverage of each line (**Table 4.1**).

Table 4.1 – Estimated coverage of the *Medicago* genomes sequenced

Line	Reads (million)	Coverage
NF443	66.05	22.3x
NF1436	63.88	21.6x
NF3209	39.26	13.3x
NF3438	56.10	19.0x

Assuming 125 bp per read and a genome size of 360Mb.

4.3 – Attempts at finding tnt1 insertions with invPCR

Previous work[298–302] had produced Sanger sequences of invPCR products amplified in *E. coli* for seven Noble Foundation lines. We performed invPCR on several lines to confirm their estimation of the total number of sequences found (**Figure 4.2** shows an example of the bands produced by invPCR). Taking the DNA sequence of the amplicons, we identified the parts of these sequences corresponding to *Medicago* DNA, the cut site from the restriction digest, and the 5 bp duplication characteristic of tnt1 insertion[275]. These sequences were then used to locate the site of insertion in the *Medicago* A17 genome[303]. With the exception of NF1436 (where the amplicon sequences could not be matched to a single genomic location), we located between 3 and 12 insertions in each of the lines tested (**Table 4.2**). NF3209 was the most intensely sampled line, and showed the most insertions. A number of the non-redundant sequences could not be located, either matching repetitive genomic regions or composed of solely tnt1 or other artefactual sequence.

Table 4.2 – Insertions found in 7 mutant lines with invPCR

Line	Total Sequences	Non-redundant Sequences	Locatable Insertions
NF443	3	3	3
NF425	12	9	6
NF472	10	5	4
NF1328	12	10	5
NF1436	6	5	0
NF3209	29	16	12
NF3438	7	7	5

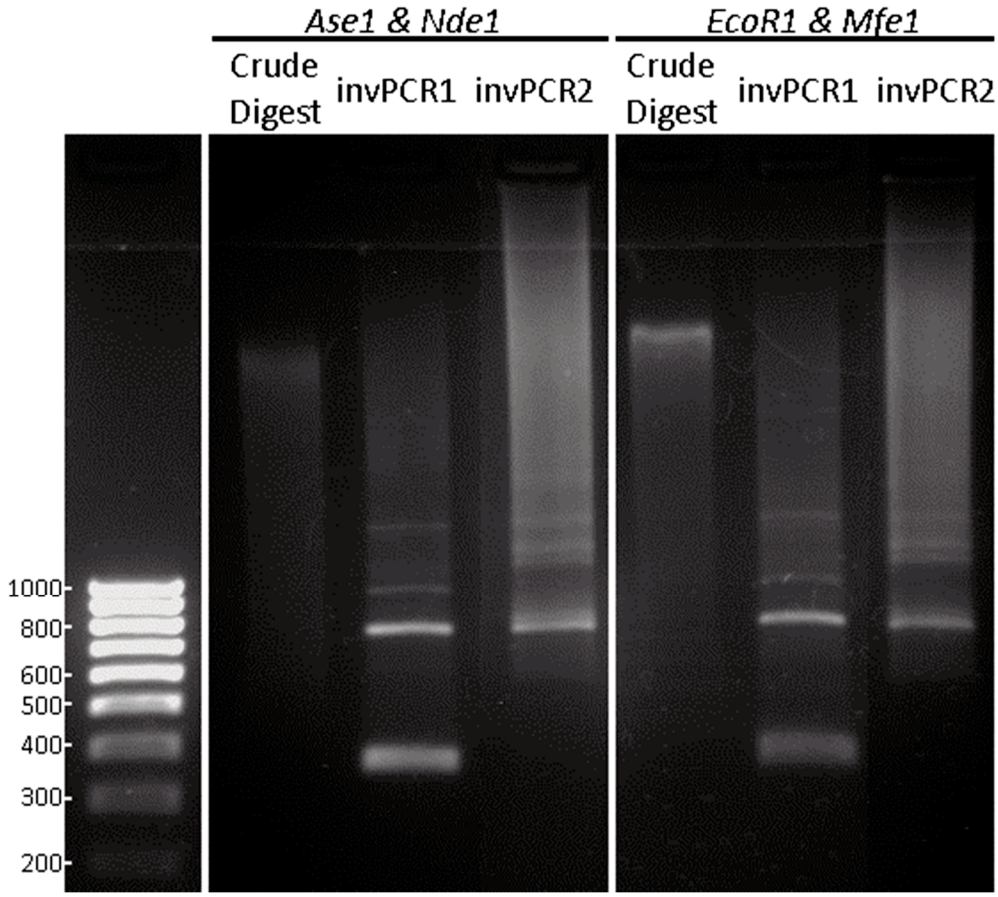


Figure 4.2 – Example of invPCR reaction
Sample invPCR reactions run on a 1.5% agarose gel. Samples produced from a DNA extraction from NF3438 leaves, digested with a pair of restriction enzymes with compatible sticky ends (Ase1 & Nde1, and EcoR1 & Mfe1) with a 100 fold dilution between invPCR reactions 1 and 2. See **Section 4.2.1** for the method.

4.4 – Identification of tnt1 insertions by WGS

4.4.1 – Finding the transposons

Tnt1 transposon insertions were located in the host genome using the BWA[304] & Samtools[305] software packages, working in a Linux environment (Fedora release 24), using the tnt1 sequence (see **Appendix 6**) to 'fish' for Illumina reads that overlapped host and tnt1 sequence (**Figure 4.3**; for scripts used, see **Appendix 5.1.1**. See Veerappan et al (2016)[306] for a similar method arrived at independently). Tablet[307] was used to visualise the alignment of Illumina reads to tnt1 sequence, and obtain the reads whose sequence contained between 20 and 100 bp of the 5' or 3' end of the tnt1 sequence. This meant that the other 25 to 105 bp should match to the *M. truncatula* genome, the Illumina read overlapping where an insertion has entered the host genome. These FST-reads were sorted to remove duplicates, and sequences from the

5’ and 3’ of the tnt1 were checked for those with the same 5 bp sequence flanking the tnt1 sequence (duplicated by the insertion scar), which indicated that the pair of sequences were likely to be the two sides of the same tnt1 insertion. The putative *Medicago* sequence was then blasted against the *Medicago* A17 and R108 genome assemblies with the MtHapMap Project’s BLAST tool, and double checked using *Medicago*-specific BLAST tools hosted by JCVI and NCBI[308–310]. While not every sequence could not be located in one or both of the published *M. truncatula* genomes, (particularly the ones that contained less than 30 bp of genomic DNA), the majority of these insertions were successfully localised. The *M. truncatula* A17 v4.1 genome browser[303] was then used to describe the local environment of each tnt1 insertion, and sort them into those likely to be impacting gene expression (those in exons, and secondarily those in the introns, promoters or terminators of genes) and those unlikely to be doing so (those more than 2 kb away from a gene). The A17 genome assembly was used for this, as the R108 version was not fully assembled or annotated.

Using the Illumina data, we discovered a large number of tnt1 insertions in each of the four mutant lines that were within a region of 2 kb upstream and 500 bp downstream of predicted or confirmed genes in the *M. truncatula* A17 v4.1 genome assembly (**Table 4.3**).

Figure 4.3 – Fishing for FST with bwa-mem
Illumina reads (red lines) were aligned to tnt1 sequence (blue bar) with the bwa-mem software. Those reads where only part of the 125 bp length overlapped the 5’ or 3’ end of tnt1 were assumed to contain host DNA indicating where a tnt1 had inserted into the host genome.

Table 4.3 – Summary of tnt1 insertions found by WGS in the four mutant lines

	NF443	NF1436	NF3209	NF3438
Exonic	17	7	13	15
Intronic	7	5	12	4
<2 kb upstream	8	6	5	6
<500 bp downstream	5	3	3	3
Intergenic	14	11	14	10
Multi-hit	3	1	6	1
Scaffold Only	7	2	4	3
Candidate Insertions	**37**	**21**	**33**	**27**
Total Insertions	**61**	**35**	**57**	**42**
No match	0	2	5	1
tnt1 fragments	7	7	1	3

Candidate insertions are all those thought to be in a region (2000 bp upstream/500 bp downstream of a gene) likely to alter gene expression, and thus have the potential to cause the mycorrhizal phenotype seen in the lines. Multi-hit insertions are those where the FST sequence matches two or three different locations in the *Medicago* genome with equal specificity (±1 SNP). Scaffold only insertions are those that match genomic scaffolds that have not been assembled into the A17 v4.1 chromosome assembly. These were excluded as candidate insertions as any nearby gene was not present in A17, making it is very unlikely that such a gene would produce a mycorrhizal phenotype. Tnt1 fragment insertions are non-redundant Illumina reads that contained only tnt1 sequence and 'no match' sequences are those that could not be matched to either tnt1 or *Medicago* DNA. The latter two are not included in the total insertions per line.

4.4.2 – Expression of target genes

We decided to further narrow down the list of insertions, to obtain initial targets for the co-segregation assay to connect candidate genes to the mutant phenotype. Taking all the candidate insertions (**Table 4.3**) we used the Noble Foundation's gene expression atlas[286] (based on published microarray data from a wide range of conditions, including data from laser capture microdissection (LCM)[100]) to assess the expression of these genes. Overall, 20-30% of the genes with nearby tnt1 insertions showed an AM-linked expression pattern (i.e. their expression was higher in at least some AM symbiotic conditions than it was in AM control conditions. Some of these genes also showed other expression peaks) (**Table 4.4**).

We would expect that the causal gene is likely (but not necessarily) induced by mycorrhizal colonisation, but that the rest of the insertions should represent a random pattern. However, the observed enrichment in AM responsive genes seems far more than could be explained by that hypothesis. RNAseq estimated that 6% of genes in *Lotus japonicus* were differentially regulated upon colonisation with *Rhizophagus*

irregularis[311]. A microarray of RNA from LCM of *Medicago* cortical cells containing arbuscules, and those adjacent to cortical cells containing arbuscules, found around 3% of genes were differentially regulated, considerably less than appeared in a whole root sample[100]. Our current understanding on gross changes to gene expression during the symbiosis is limited, as shown by Giovannetti et al (2015)[312], who found that there was little overlap in genes upregulated on *L. japonicus* roots treated with *Gigaspora margarita* germinated spore extract (GSE) and *M. truncatula* roots treated with *R. irregularis* GSE (assayed by microarray). Therefore, caution should be taken when examining gene expression patterns[313]. A possible explanation for the greater abundance of insertions near AM linked genes would be that the phenotype of the mutant line is caused by an additive effect of several genes. However, co-segregation showed that this was not the case with at least NF443 and NF3438 (see **Chapter 5**), so this also appears an incomplete explanation. Overall, we are unsure why this high a proportion of disrupted genes show AM-linked expression, and must conclude that 20-30% is probably an underestimate, as this percentage rises to 25-40% if we only look at genes for which we have expression data. We conclude that this expression data is a useful guide to pick initial targets of inquiry, but that we should avoid completely discarding a gene based on its expression in one experimental condition, as plant genetic responses may be sensitive to genotype and environmental interactions, and some known mycorrhizal genes (i.e. DMI3) are consistently expressed regardless of the presence or absence of AMF.

Table 4.4 – Expression of genes affected by tnt1 insertions

	NF443	NF1436	NF3209	NF3438
AM-linked expression	8	5	8	8
No AM-linked expression	21	11	12	12
No expression data available	8	5	13	8

Expression pattern of the Medicago A17 4.1 predicted genes disrupted by or closest to each candidate insertion (**Table 4.3**) in each of the four mutant lines.

4.4.3 – *In silico* zygosity calling for the tnt1 insertions

The zygosity of the tnt1 insertion locus could be assessed *in silico* using a similar technique to that used to locate the insertion, by aligning the Illumina data to the sequence of the predicted genes, and using Tablet to look for the distinctive 5 bp duplication (see **Appendix 5.1.2** for the appropriate script). When the tnt1 insertion

enters the DNA, it duplicates the 5 base pairs where the transposase bound, leaving a duplicated scar sequence on either side of the transposon. Illumina read data was aligned to the 1 kb region of DNA around each of the insertion sites discovered in **Section 4.4.1**. This surrounding sequence was obtained from the R108 0.95 genome assembly[297]. The resulting Tablet assembly was then visually assessed for the 5 bp duplication scar. If tnt1 was present, Illumina reads from both sides of the tnt1 would align to the genome sequence up to and including this 5 bp duplication, then stop aligning as they ran into tnt1 sequence. Thus, when the Illumina reads were assembled against the *Medicago* genome sequence of a tnt1 insertion locus with a copy of the tnt1 inserted into both chromatids, all the reads aligned around the 5 bp duplication site would be shorter than 125 bp, and stop aligning to the host sequence at the duplication. Alternatively, if the insertion was heterozygous, roughly one half of the Illumina reads would show the above pattern, while the other half would align exactly to the host sequence across the duplication site (illustrated in **Figure 4.4** and **4.5**).

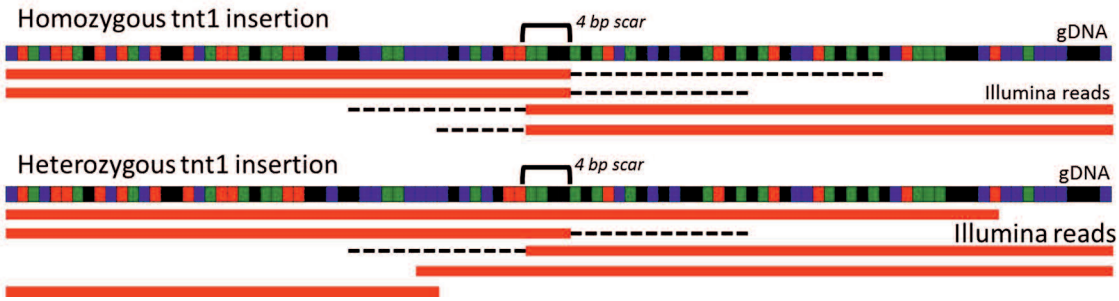


Figure 4.4 – Visual expectation of homo/heterozygous insertions in Tablet
Expected read pattern resulting from a homozygous (top) or heterozygous (bottom) tnt1 insertion.

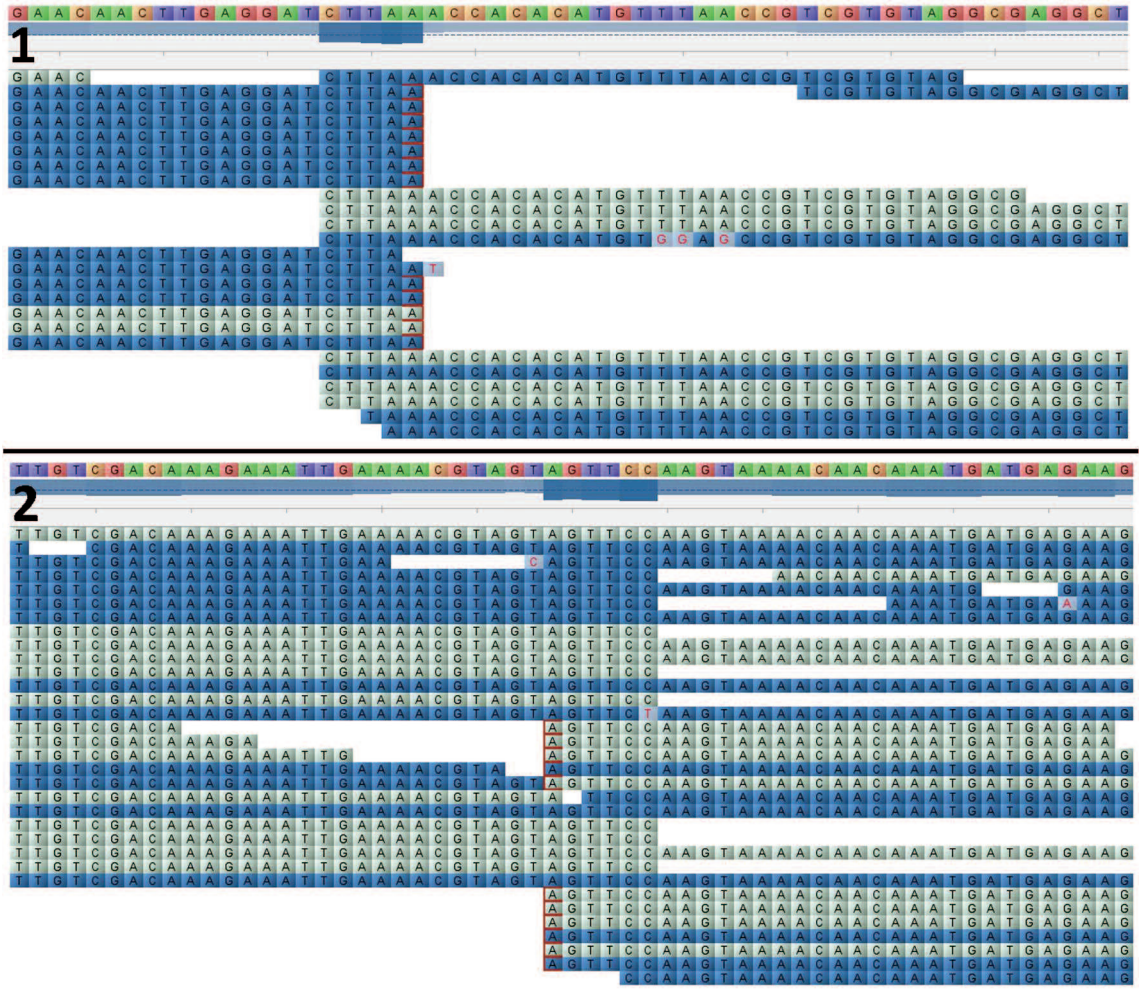


Figure 4.5 – Actual Tablet visualisation of homo/heterozygous insertions
Appearance of:

(1) A homozygous insertion (NF3438 i25 into RAM1) in Tablet[307] around a CTTAA duplication site. Illumina reads are coloured by read direction.

(2) A heterozygous insertion (NF443 i20 into SWEET14) around a GTTCC duplication site. A number of Illumina reads cross this duplication, uninterrupted, indicating that the tnt1 insertion is only into one of the pair of chromosomes.

This process was not performed on all insertions due to the time consuming nature of the visual assessment. We tested all of the genes of interest described in **Section 4.5**. In all cases that the zygosity of these insertions could be determined *in silico*, that zygosity was confirmed with PCR genotyping (see **Chapter 5**). For insertion sites in regions of repetitive DNA, or into genes that were part of large conserved families, the *in silico* zygosity calling was difficult due to closely related sequences also being aligned to the host sequence. This would have the effect of incorrectly describing a homozygous insertion as heterozygous. Therefore, in cases where most reads contained >2 SNPs, or the ratio of non-tnt1 containing reads greatly exceeded a 1:1 ratio with tnt1 containing reads, the zygosity call was considered suspect. Optimising

the parameters of the `bwa mem` function (discussed in **Appendix 5**) allowed most of these issues to be resolved. However, this appeared to be mainly on a case-by-case basis, the most optimal parameter set depending on the nature of the specific sequence. More work on this technique, and testing in other systems, is required before it can be considered truly reliable.

4.4.4 – Confirming the absence of mutation in known MYC genes

We found by co-segregation (see **Section 5.5**) that a tnt1 insertion into the first exon of RAM1[72] was causing the phenotype in NF3438. Therefore, we scanned the sequence of 19 genes involved in the AMS to ensure that none of the sequenced lines had mutations in already known mycorrhizal genes (see **Table 4.5** for a list of these genes). To limit the amount of testing, we excluded most of genes known to have strong NOD^- phenotypes (e.g. NSP1 and NSP2), as these should have been excluded by the initial screening criteria. Illumina reads for each line were aligned to these sequences (using the same script presented in **Section 4.4.3**), and the `.bam` output was visually assessed for the 5 bp read-break characteristic of a tnt1 insertion site. The alignment was performed on the gDNA, rather than mRNA sequence, as insertion in the promoter, UTRs and introns might affect expression.

As a future development for routine use of WGS screening, this method could be automated by extracting the individual reads that partially align to the known mycorrhizal genes, and comparing the read sequence to tnt1.

Table 4.5 – List of known mycorrhizal genes

Gene	Function
*Gm*AMT1.4	Ammonia gas channels believed to retrieve fungal delivered N from the periarbuscular matrix[314]
*Gm*AMT4.1	
LIN	E3 Ubiquitin ligase which degrades an unknown suppressor of extracellular AM/nodule symbiont movement[123]
*Ps*CRY	DELLA family gene degraded by gibberellic acid to negatively regulate colonisation during high nutrient conditions[166,315]
DMI1	Nuclear membrane K^+ transporter that counterbalances charge for Ca^{2+} movement during CSP nuclear calcium spiking[56]
DMI3	Calcium/Calmodulin binding protein kinase, the receptor for CSP nuclear calcium spiking[316]
HA1	Periarbuscular H^+-ATPase that provides the proton gradient from nutrient uptake from the periarbuscular matrix[138]
KPI106	Inhibitor of SCP1[124]
*Ps*LA	Redundant with *Ps*CRY
miR171h	Prevents colonisation of growing root tip by degrading NSP2[317]
miR396a/b	Control root development, repressed during the AMS[318]
PT4	Organophosphate/H^+ cotransporter that takes up Pi from the periarbuscular matrix[319]
RAM1	GRAS transcription factor that drive mycorrhizal gene expression[72]
RAM2	Produces cutin monomers that induce fungal hyphopodium formation[93] and provides metabolic and structural fatty acids to the fungus [129]
*Lj*SbtM1	Subtilases that loosen the extracellular matrix to allow hyphal spread[120]
*Lj*SbtM3	
SCP1	Carboxypeptidase that promotes hyphal spread inside the root[124]
STR	Pair of half-ABC transporters that form a transporter for an unknown substrate in the periarbuscular membrane[131]
STR2	

Known mycorrhizal genes screened for presence of tnt1 insertion in the four sequenced lines. The sequence of each gene was obtained from the cited paper or from the NCBI database, and matched to the closest gene in the *M. truncatula* A17 4.1 genome assembly[303]. Some gene sequences could not be located for *M. truncatula*. In these cases the most homologous *M. truncatula* gene was used to assemble the sequence. See **Appendix 6** for all sequences.

With the exception of NF3438 insertion 25 (into *RAM1*/Mt7g027190), none of the lines contained any tnt1 insertions in these MYC genes, supporting the assertion that their phenotype was caused by a novel MYC gene.

4.4.5 – Low confidence tnt1 insertions

A number of tnt1 FSTs in each line could not be matched to a genomic location. These fell into three categories.

First, some FSTs did not show a match with any version of the *M. truncatula* genomes[308–310]. The majority of these FSTs were found to be homologues to internal regions of the tnt1 sequence. The reason for this is unknown, but it may indicate the presence of nested tnt1 insertions, or adjacent fragmentary transposons (**Figure 4.6**). Alternately, this may simply be an error in the sequencing. Either way, these reads were not informative, thus were discarded from the analysis. The remaining FSTs which had no homology to *M. truncatula* or tnt1 sequences, were short (20-40 bp), but this still should have been long enough to accurately match them to the host genome.

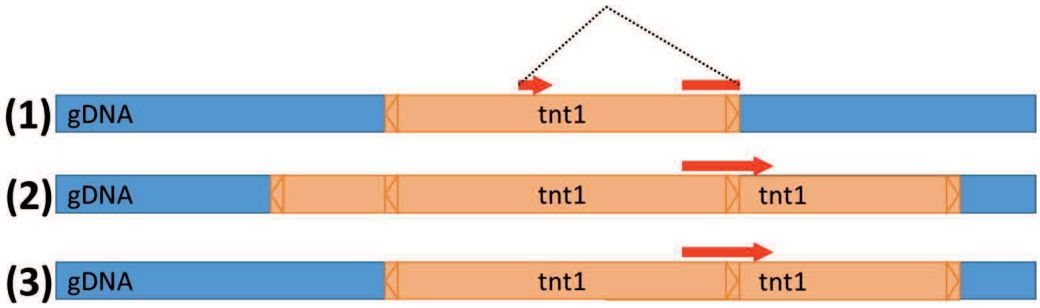


Figure 4.6 – Some Illumina reads contain multiple discrete sections of the tnt1 sequence
Visual description of the Illumina reads that contain multiple discrete sections of the tnt1 sequence (1) (shown in orange, with triangles indicating the tnt1 flanking regions), and two possible explanations outside experimental error for how this could have occurred, with the read overlapping nested (2) or fragmentary (3) tnt1 insertions.

The second category were 26 FST that had homology to partially assembled scaffolds in the R108 genome[297], but not to the A17 assembly[295] (see **Table 4.3**). Additionally, there were two FST (NF443 i49 and i50) that showed homology to non-assembled scaffolds in the A17 assembly. Annotation for these unassembled scaffolds in both accessions is poor, so the insertions corresponding to these FST were excluded from the initial analysis. For those FST that showed homology to R108 but not to A17, we took the R108 5 kb of scaffold sequence either side of the FST, and aligned that to the A17 genome. We found that 5 of the 24 FST were homologues to the inferred 'other side' of other tnt1 insertions for which we had previously had FSTs corresponding to only one of the 3' or 5' side. Another 5 FST were found to be novel insertion loci, and a sixth matched invPCR NF3438 Seq12. To confirm that the alignment to R108 is correct, and that our plant genomes differed from A17 in these regions[320], we aligned the raw Illumina data to the published R108 scaffolds, and visually confirmed in Tablet that there were not break-points. The remaining 15 insertions (see **Table 4.6**) could either not be located in the A17 chromosomal assembly, or were not syntenic with it.

The third category of FST are those which show equal homology (±1 bp mismatch) to multiple locations in the *M. truncatula* genome. Of these 13 'multi-hit' FST, five were homologous to highly repetitive regions of non-coding DNA outside our region of interest (2 kb upstream or 500 bp downstream of the nearest gene). These insertions were discounted as likely causes of the mycorrhizal phenotype. The other eight had homology to possible locations within the region of interest, but none within exons.

Table 4.6 – Locating insertions from the R108 scaffold assembly

		NF443	NF1436	NF3209	NF3438
	Initial R108 scaffold matches	12	3	6	5
Suggested Location	Matches existing WGS insertion	2	1	2	0
	Matches existing invPCR insertion	0	0	0	1
	Novel insertion locus	3	1	0	1
	Still unknown	7	1	4	3

24 tnt1 insertions could not be located to the A17 genome in the first pass. This table describes the distribution of these by mutant line, and the A17 genome locations assigned on extension of the sequence using the R108 scaffold. Those that match a WGS insertion are the other side (either 5' or 3') for an insertion for which only one side had been found previously. Those that are unknown either had no match between the R108 scaffold and the A17 genome, or showed substantial lack of synteny.

4.4.6 – Distribution of the tnt1 insertions

Combining the invPCR and WGS data, the four lines have between 35 and 61 high-confidence tnt1 insertions, with between 32 and 50 insertions per line that can be located to a single point in the A17 chromosome assembly (**Table 4.3**)[303]. These insertions were carried forward for the further analysis (described in **Section 4.5**). The distribution of the tnt1 transposon in our lines confirm the hypothesis of preference for exon integration suggested by Tadege et al (2008)[272], who used FST sequencing to locate 964 tnt1 insertions in the *Medicago* genome. They found 34.1% of insertions were exon localised, with 23.2% in introns and 42.6% in other regions. Our dataset matches this exonic proportion (31.6%), with 15.4% in introns and 52.9% in other regions. Tadege et al (2008)[272] also found a range of 6 to 59 total tnt1 insertions per line, with a mean of 25. Our findings were within this range, albeit towards the high end of the distribution. This may be due to chance selection of four lines from the collection, or may be because WGS discovers insertions that were not found in other approaches.

The important question then is; did the WGS approach find all the tnt1 insertions present in the tested mutant lines, and if not, can we estimate the effort that would be required to find them all? The first possibility is that we did not find all of the insertions present in the mutant lines. Comparing the WGS dataset to the invPCR data, there were FST for six insertion loci found with invPCR but not with WGS across the three line. We have no good invPCR data for NF1436, but the Noble Foundation tnt1 insertion database[277] lists nine tnt1 insertions found in NF1436 by tail-PCR, of which three were not found in our WGS dataset.

PCR genotyping of the NF1436 individual that had been sequenced failed to find any tnt1 insertion at these loci (**Figure 4.7**). The likely explanation for this discrepancy is that the Noble Foundation seedstock is segregating for heterozygous insertions, and that these three insertions had been lost from the sequenced individual. This inability of PCR based methods of insertion discovery to identify zygosity without secondary PCR genotyping is a clear weakness relative to WGS. For reverse genetic screens (as seen in **Section 6.3.1**) we would advise that they obtain multiple lines believed to contain their mutation of interest to limit the impact of this segregation. Despite this problem, the Noble Foundation collection does remain the most saturated *M. truncatula* mutant population currently available, so should still be consulted for reverse genetic work.

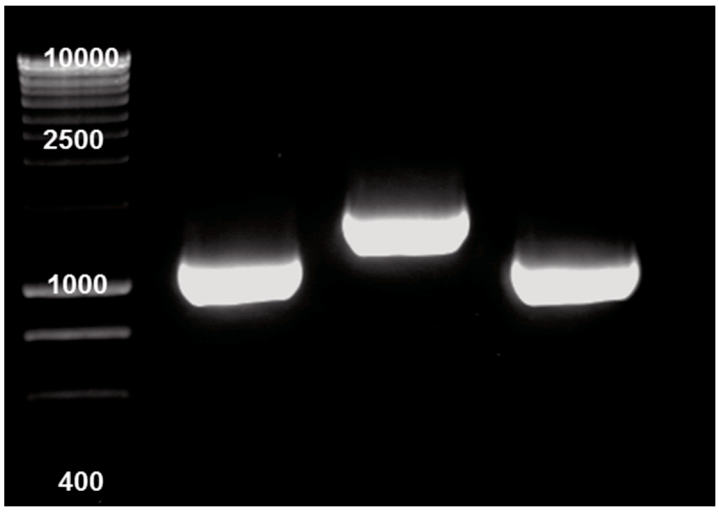


Figure 4.7 – Three FST found by the Noble Foundation are not present in our NF1436 seed stock
PCR for three FSTs given in the *Medicago* tnt1 database entry for NF1436 that were not recapitulated in the Illumina WGS dataset. The three bands were the size predicted for the wild-type chromosome (lacking any tnt1 insertion) and not the larger fragment (6-7 kb) that would be produced if the chromosome contained tnt1.

While we can discount at least some of the PCR discovered insertions as being lost from the sequenced individual due to segregation, we can attempt to estimate the sequencing coverage required to be confident we have found all the tnt1 insertions in any given *M. truncatula* mutant. Sampson et al, 2011[321] calculated that a read depth of 42x would be required to find 99% of SNP variation in a genome. However, this is looking for heterozygous SNPs (requiring 2-8 reads at any given genomic location for accurate diagnosis of a SNP). A transposon insertion mutation, where a given WGS read would contain 10s to 100s of nucleotides difference to a non-mutated individual, should be comparatively less vulnerable to sequencing error, and thus easier to find. In theory, single read coverage should be enough to locate all homozygous insertions, but the depth of coverage required to obtain essentially complete is hard to estimate, varying with genome size, genome composition and methodological biases.

Indeed, we are unsure of our estimate coverage values (**Table 4.1**), which were based on the Mudge et al's 360 Mb estimate[297]. The MegaHit assembly[322] of our own pooled R108 genome data was significantly less complete that the one of Mudge et al, but showed a similar estimated total scaffold length (376 Mb vs 388 Mb). This is also consistent with Tang et al's A17 genome[295], supporting the rough estimate of a sub-400 Mb genome size for R108. However, we observed that many of our insertion loci had less than 10x coverage, with gross inflation of the read depth into the tens of thousands over microsatellites and other repetitive regions. We tested this with the tnt1 coverage data, assuming that the average coverage of the tnt1 sequence in each line should be relative to the abundance of tnt1 insertions in that line. However, this does not seem to be the case, as NF443 showed substantially lower coverage per tnt1 insertion than for the other lines, and no line showed as much coverage as the whole genome estimate (**Table 4.7**). One explanation for this is that many of the tnt1 insertions in these parental lines would be heterozygous, thus we would expect to see less than 100% genome coverage. This appears a good explanation for NF1436 and NF3438, with ~75% relative tnt1 coverage suggesting an insertion population that is half homozygous and half heterozygous. NF443, however, has a relative coverage of 46%, lower than would be seen even if all insertions were heterozygous (50%).

Table 4.7 – Did WGS discover all tnt1 insertions present in the mutant lines?

Line	Reads (Million)	Whole genome coverage	Estimate number of tnt1 insertions	Average coverage of tnt1	Coverage of tnt1 relative to whole genome
NF443	66.05	22.3x	61	626x	46%
NF1436	63.88	21.6x	35	572x	76%
NF3209	39.26	13.3x	57	468x	62%
NF3438	56.10	19.0x	41	618x	79%

Estimated coverage of the tnt1 sequence in *Medicago* lines sequenced on a single Illumina lane for this project, based on an estimated genome size of 360Mb and the number of tnt1 insertions discovered (**Table 4.3**).

Another way to examine this question is to look at the number of reads found for the boundaries of each insertion in a line. It would be expected that read coverage of these points should form a normal distribution around a value equal to the average genome coverage. Plotting a histogram of read coverage of tnt1 insertion sites in our sequenced genomes (**Figure 4.8**), we see that only NF1436 falls into a largely normal distribution, with the other lines, especially NF3438, being over-represented in tnt1 insertions with 3 or fewer associated reads. According to this method, the ability to locate all insertions appears strongly influenced by the number of tnt1 in the line, as NF1436, which has the fewest estimated insertions, shows the cleanest normal distribution. The log-normal distribution observed could again be caused by extensive heterozygosity of the insertions in the sequenced lines, and this method makes it harder to estimate the effects of that.

Overall, we suggest these methods mainly give similar data, although we are unable to explain the differences seen in NF3438, where insertion boundary coverage (**Figure 4.8**) does not appear to match with the overall coverage of tnt1 (**Table 4.7**). We believe the first method (**Table 4.7**) gives a more useful single figure output. From this, we can plausibly claim to have found all the tnt1 insertions in all lines apart from NF443, although NF3209 may also be rather low (62%, suggesting a large majority of insertions are heterozygous). For future work that requires confidence that all tnt1 insertions in a Noble Foundation line have been discovered should sequence to at least 20x. The 42x coverage suggested by Sampson et al (2011) is probably unnecessary, but around 30x coverage should be sufficient, and is achievable with 2-3 plants per lane on contemporary Illumina machines, and will doubtless improve in future.

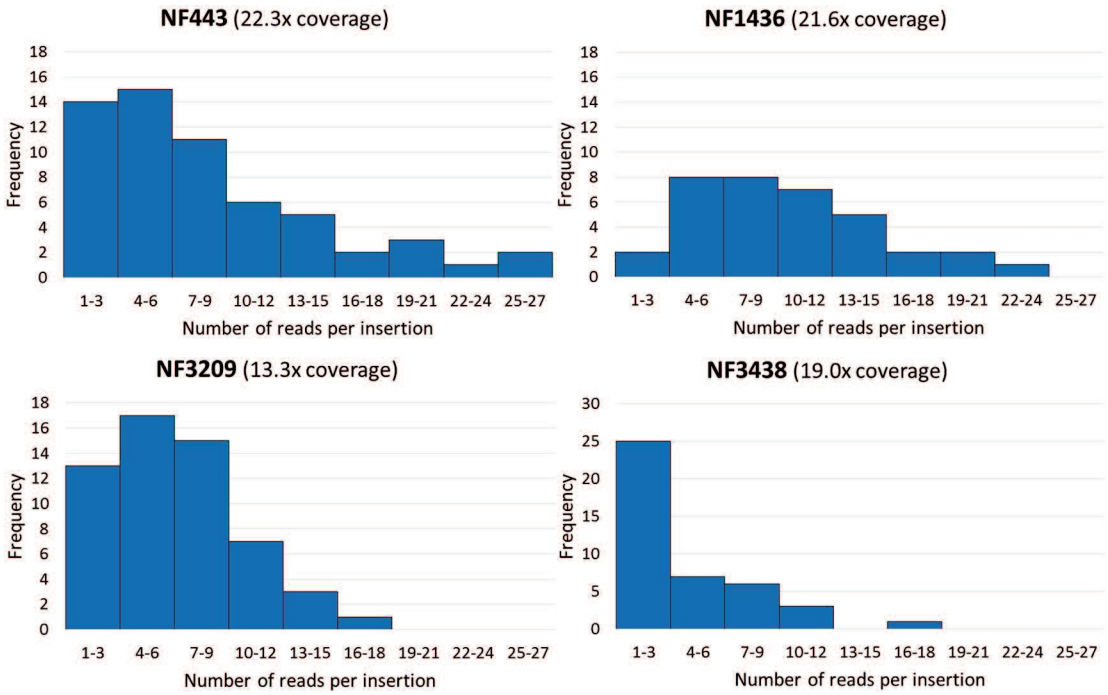


Figure 4.8 – Did WGS discover all tnt1 insertions present in the mutant lines?
Read coverage of each insertion/genome boundary, with binned read coverage on the x axis, and number of tnt1 insertions with that number of reads on the y axis.

4.4.7 – Genetic differences between *M. truncatula* accessions

A complicating factor in locating insertions is the large differences we observed between the A17 and R108 genomes. Attempts at assembling our genomic data using the published A17 chromosomes as a template, as well as comparisons between the two published *M. truncatula* genomes[295,297], suggest extensive rearrangement has occurred. We found evidence of the exchange of chromosome fragments (**Figure 4.9**), as well as large amounts of smaller-scale changes, including inversions (**Figure 4.10**), insertions and deletions. While we have not attempted to quantify the scale of these changes, the data to do so is publically available. These observations are supported by older reports of changes in linkage mapping and semi-sterility between accessions of *Medicago*[320], including a major exchange event between chromosomes 4 and 8. The reason for these changes are unclear. We advise that those working with R108 plants take the time to find the R108 sequence for areas they wish to PCR amplify, as a significant proportion (approximately 1/4) of primers designed using A17 sequence failed to amplify the expected sequence in R108. We hope Mudge et al[297] will produce a 1.0 release of the R108 chromosomal genome soon. We have attempted to assemble our own R108 data with the SPAdes[323] and MegaHit[322] programs, without

much success (N50 7.7 kb). Using the A17 genome to link R108 contigs is probably unwise due to the extensive rearrangement (**Figure 4.9**), so we suggest that obtaining long read genome sequence (e.g. from MinION or similar chemistries) would be the easiest way to obtain a full R108 assembly.

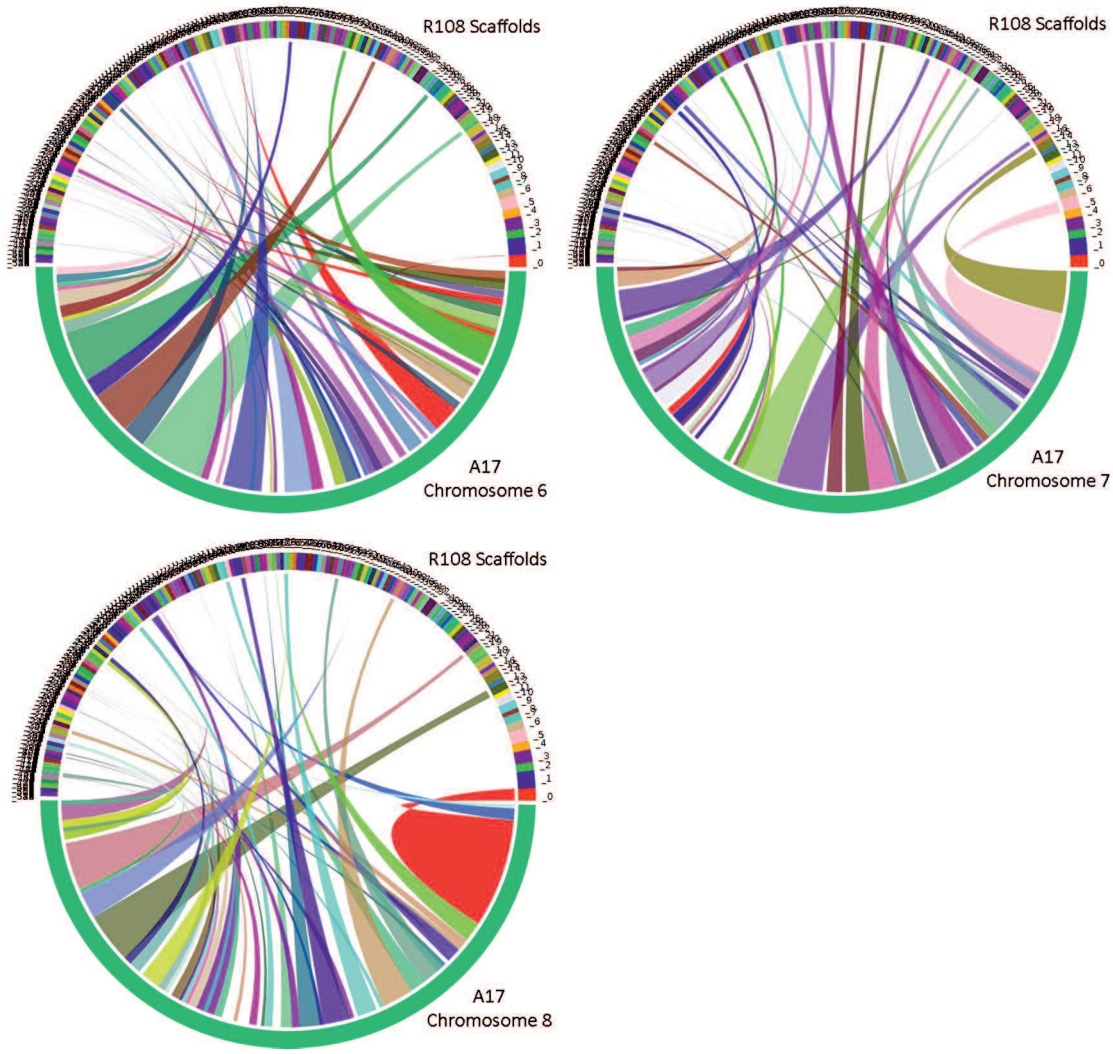


Figure 4.9 – R108 and A17 genomes show evidence of chromosome rearrangement
Graphical depictions (produced in BBmap) of synteny between chromosomes 6, 7 and 8 of the 4.0 A17 genome assembly and the scaffolds of the R108 0.95 assembly. Note the alignment of the main part of scaffold 0 to chromosome 8, and a lesser part to chromosome 6, while significant parts of chromosome 7 has no synteny in the R108 assembly.

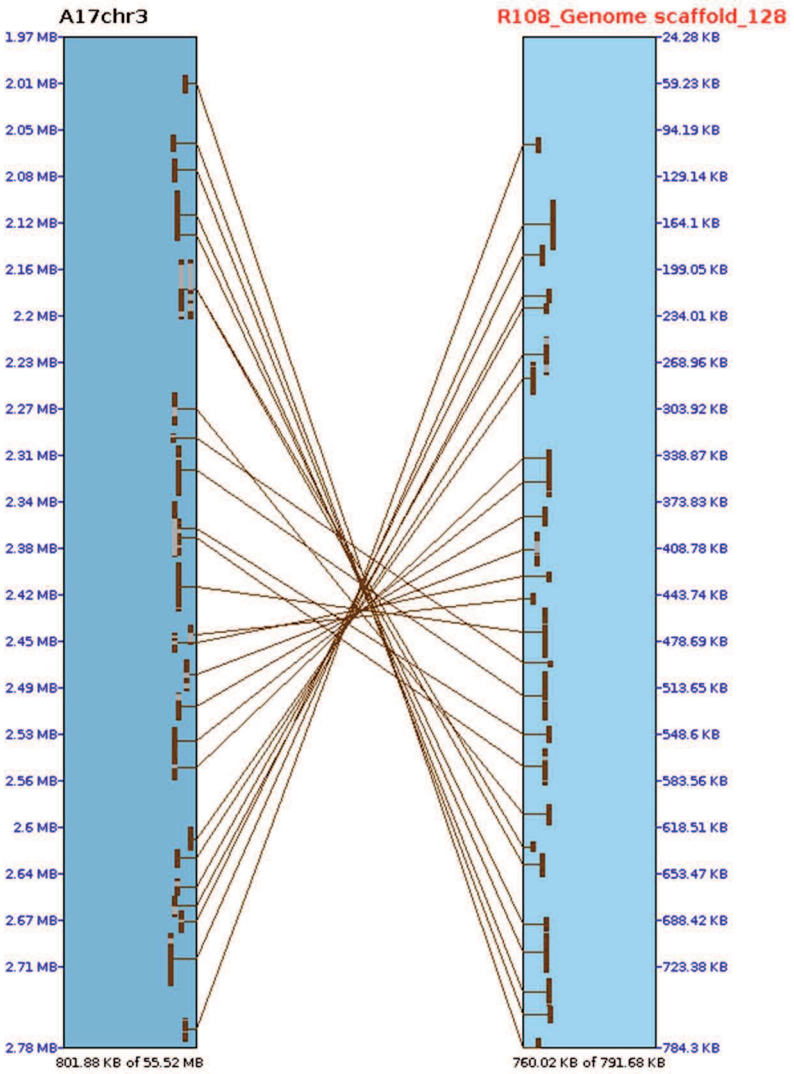


Figure 4.10 – R108 and A17 genomes show evidence of inversion
Graphical map of fragment synteny (produced in BBmap) between a segment of chromosome 3 [A17 4.0 assembly] and Scaffold 128 [R108 0.95 assembly] showing an inversion of the approximate region 2.2 to 2.4Mb on chromosome 3.

4.5 – Selecting likely candidate genes

4.5.1 – What makes a gene a good candidate?

Each of the four sequenced lines has a large number of insertions, making PCR genotyping of large F_2 populations time consuming. However, working with certain assumptions, we can narrow down the candidate insertions to those likely to be causing the reduced mycorrhizal colonisation phenotype, and assess those first. In this work, we made three main assumptions. First, that a tnt1 insertion into a gene will cause a recessive loss of function mutation. The tnt1 ORF product is only produced during callus culture or strong biotic or abiotic stress[273,324]. We do not believe that the plants have been exposed to such stressors during our experiments, so the ORF

promoter should not be driving ectopic expression of nearby native genes, which would lead to possible dominant phenotypes. The exception to this assumption would be a dominant or semi-dominant loss of function mutant caused by disruption of a negative regulator. However, a simple χ^2 test on the ratio of mutant to wild-type phenotypes in the F_2 should allow us to detect the form of inheritance of the mutation.

Our second assumption is that an insertion outside of 2 kb upstream and 500 bp downstream of the predicted mRNA transcript is sufficiently unlikely to cause a change in expression that these insertions can be discounted. Within this region of interest, we assume that an insertion into an exon is most likely to disrupt a gene product, followed by insertion into introns or promoter regions. An insertion downstream of the coding region is least likely to cause significant expression changes. Finally, between insertions within the exons, those nearest the start codon should be most likely to block protein production. We expect the same pattern to hold true in the promoter, with regard to control of gene expression.

Our third assumption is that a gene with strongly mycorrhizal-associated expression is more likely to be causing a mycorrhizal phenotype than one without significant expression during the symbiosis. The main problem with this approach is that it does not take into account regulation by post-translation modification, so genes with an equal mRNA abundance in AM and NM roots could have different protein levels. But, genome wide data for post-translation modification is not available, so we cannot screen potential insertions in this manner. Another potential exception to this rule is genes such as some ligand receptors, which could cause major cell signalling responses even when present at low abundance via signal amplification. However, the Nod factor receptor NFP is highly upregulated on contact with *Sinorhizobium meliloti*[313], so this does not apply to all receptor kinases.

Analysis of the WGS dataset for this third assumption was made more complicated by the large number of insertions which show mycorrhizal expression (**Table 4.4**), as each line carries multiple (between 6 and 11) genes with a tnt1 insertion in our region of interest that are upregulated during the AMS. Some may be unaffected in expression by the tnt1 insertion, or do not produce a mycorrhizal phenotype due to redundancy

of function or by playing a minor role in the AMS. The AM phenotype of one or more lines may be polygenic, and this should be predictable if the ratio of phenotypes in the F_2 population significantly differs from 3:1 wild type and mutant (see **Chapter 5**).

With this in mind, we present the location of each high confidence tnt1 insertion in the four lines sequenced, and highlight the insertions we believed (based on the assumptions described above) to be the most likely to cause the reduced AM colonisation phenotype of the mutant line.

4.5.2 – NF443: likely candidates

There are 61 insertions in NF443 that can be located to single point in the A17 4.1 genome assembly (**Figure 4.11**). While there is at least one insertion on each chromosome, the majority are located in chromosome 3 (12 insertions) and chromosome 7 (13 insertions). Based on the criteria in **Section 4.5.1**, we considered the following insertions (**Table 4.8**) the most likely to cause the mycorrhizal phenotype. The expression pattern of the genes disrupted by these insertions are given in **Figure 4.12**.

Table 4.8 – Promising candidate insertions for causing the mycorrhizal phenotype in NF443

Insertion	Nearest gene	Location relative to the nearest gene	Predicted gene function
16	3g072300	Intron 1	NPF family transporter
36	7g446190	Exon 1	Leucine rich repeat receptor-like kinase
42	7g116650	Exon 6	Tyrosine kinase
43	8g012350	Exon 2	FKBP family protein isomerase

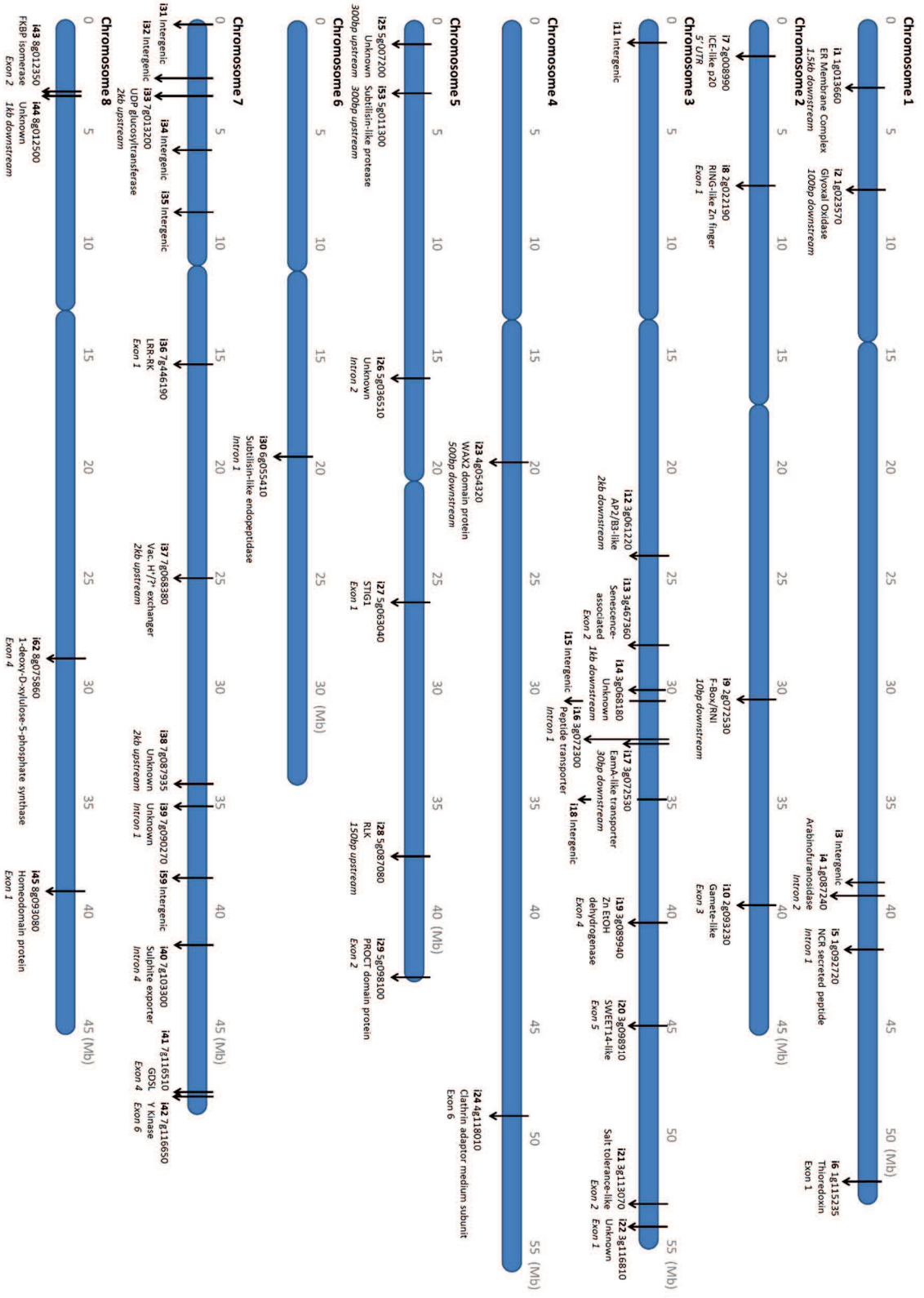


Figure 4.11 – Chromosome map of NF443 showing location of all tnt1 insertions
Map of the chromosomes of *M. truncatula* (based on the A17 v4.1 genome assembly). Each arrow marks the location of a tnt1 insertion found in the WGS dataset from NF443, and lists the location relative to the nearest gene, and the numeric code and predicted function of said gene.

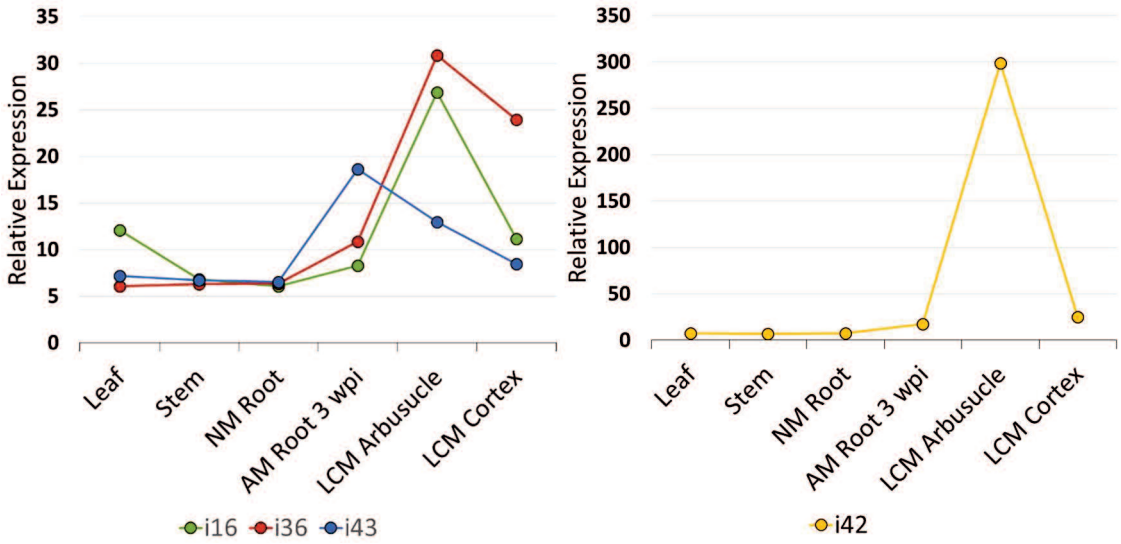


Figure 4.12 – Expression of genes affect by candidate insertions in NF443
Expression of four genes at tnt1 insertion loci in NF443 considered candidates for causing the reduced colonisation phenotype of that line. Expression data is derived from the Noble Foundation's *Medicago* Expression Atlas[313].

4.5.3 – NF1436: likely candidates

35 of the detected insertions into line NF1436 could be located to a single genomic locus (**Figure 4.13**). They are fairly evenly distributed across the genome, with no more than 7 in any one chromosome. We considered there to be three good candidates for causing the deformed arbuscule phenotype, which are listed in **Table 4.9**. The expression of genes disrupted by these insertions is described in **Figure 4.14**.

Table 4.9 – Promising candidate insertions for causing the mycorrhizal phenotype in NF1436

Insertion	Nearest gene	Location relative to the nearest gene	Predicted gene function
15	4g125060	Exon 5	GDSL acyl-esterase
17	5g087080	200 bp upstream	Receptor-like protein
28	8g069625	Exon 5	TIR-NBS-LRR

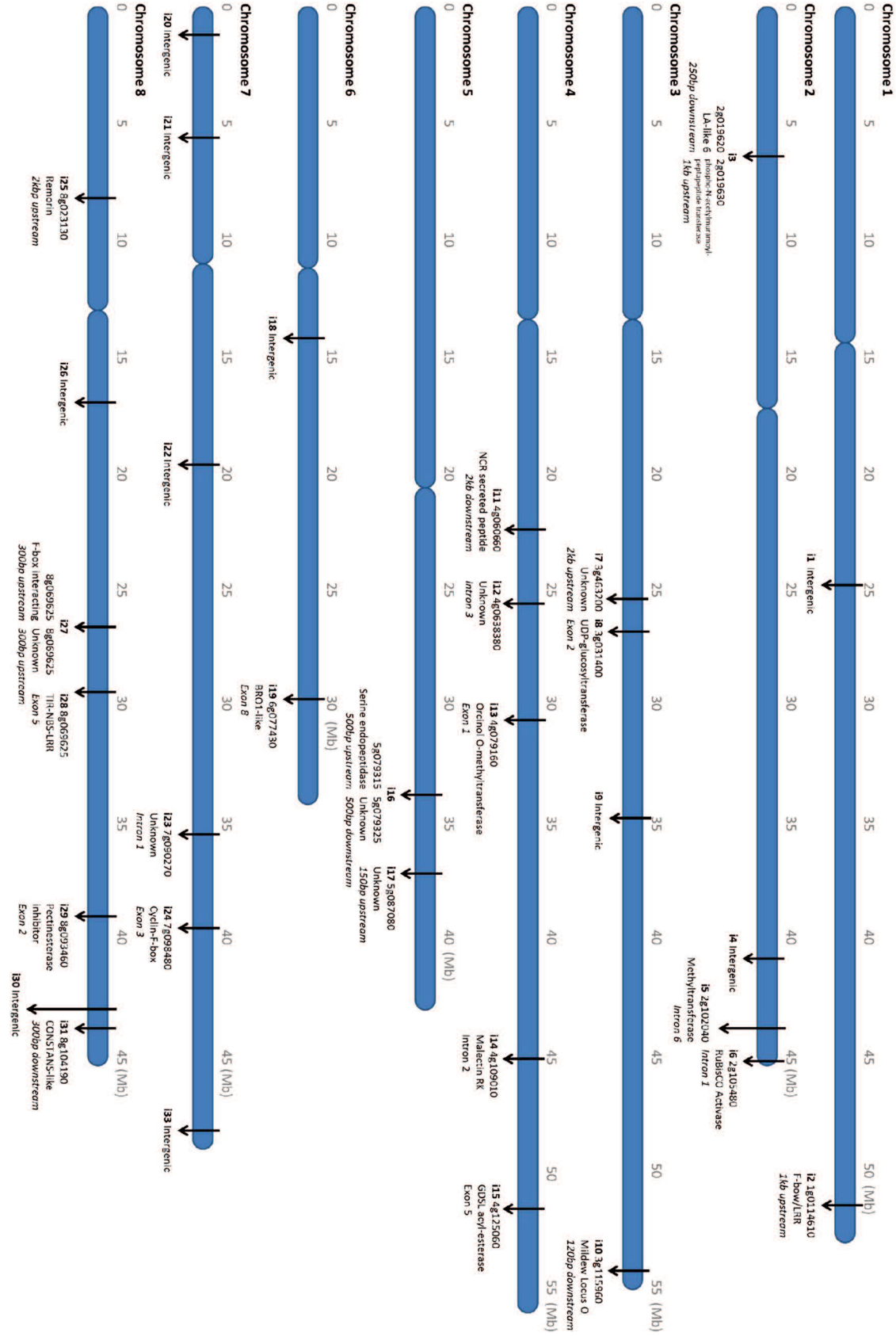


Figure 4.13 – Chromosome map of NF1436 showing location of all tnt1 insertions
Map of the chromosomes of *M. truncatula* (based on the A17 v4.1 genome assembly). Each arrow marks the location of a tnt1 insertion found in the WGS dataset from NF1436, and lists the location relative to the nearest gene, and the numeric code and predicted function of said gene.

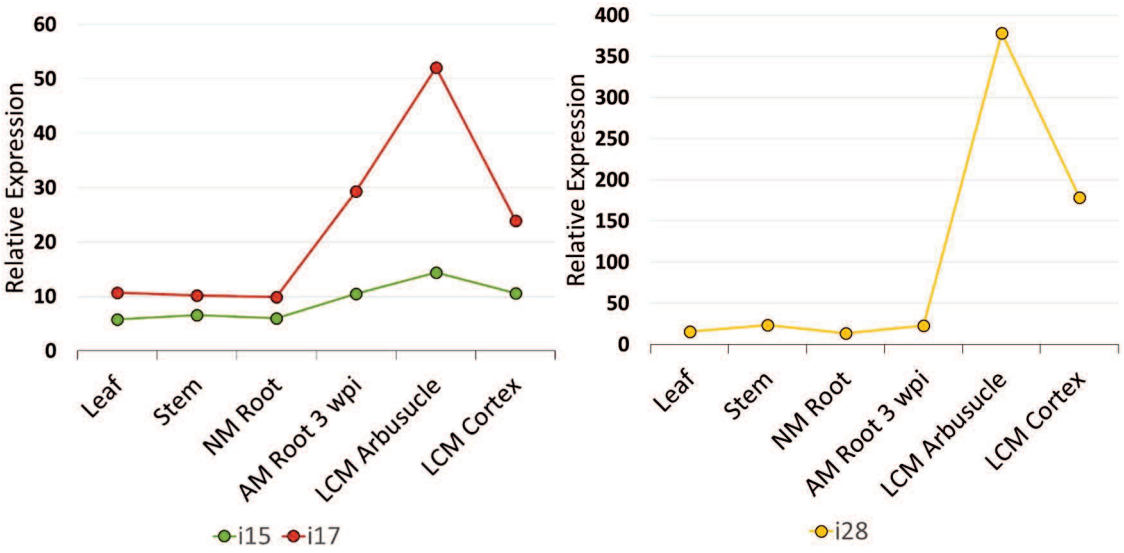


Figure 4.14 – Expression of genes affect by candidate insertions in NF1436
Expression of four genes at tnt1 insertion loci in NF1436 considered candidates for causing the deformed arbuscule phenotype of that line. Expression data is derived from the Noble Foundation's *Medicago* Expression Atlas[313].

4.5.4 – NF3209: likely candidates

NF3209 contains 57 tnt1 insertions (**Figure 4.15**). Five appear good initial candidates to cause the delay in colonisation phenotype (**Table 4.10**). Expression data for the disrupted gene was only available for three of these insertions (1, 31 and 38; **Figure 4.16**). The other two insertions were considered interesting candidates given their exonic location (giving a high chance of preventing protein accumulation of the gene), and the type of gene disrupted. i12 affects a LRR-RLK, a broad class of receptor kinases that includes the LysM receptors, many of which are known to function in the AMS[325]. Insertion 32 affects a gene encoding a transcription factor, suggesting a role in control of other genes. No zinc finger TF is currently known to be directly involved in the AMS (which mainly uses GRAS family TFs) but many zinc finger TFs are involved in plant defence signalling[326], which needs to be controlled to allow the AMS, and at least one TF from the WRKY subfamily has been found upregulated in mycorrhizal roots[100,158].

Table 4.10 – Promising candidate insertions for causing the mycorrhizal phenotype in NF3209

Insertion	Nearest gene	Location relative to the nearest gene	Predicted gene function
1	1g022380	Only Exon	RGA4
12	3g070220	Exon 1	Leucine rich repeat receptor-like kinase
31	7g011490	Intron 1	emp24/gp25L/p24 family
32	7g023010	Only Exon	Zinc finger transcription factor
38	7g090530	Exon 2	2C family protein phosphatase

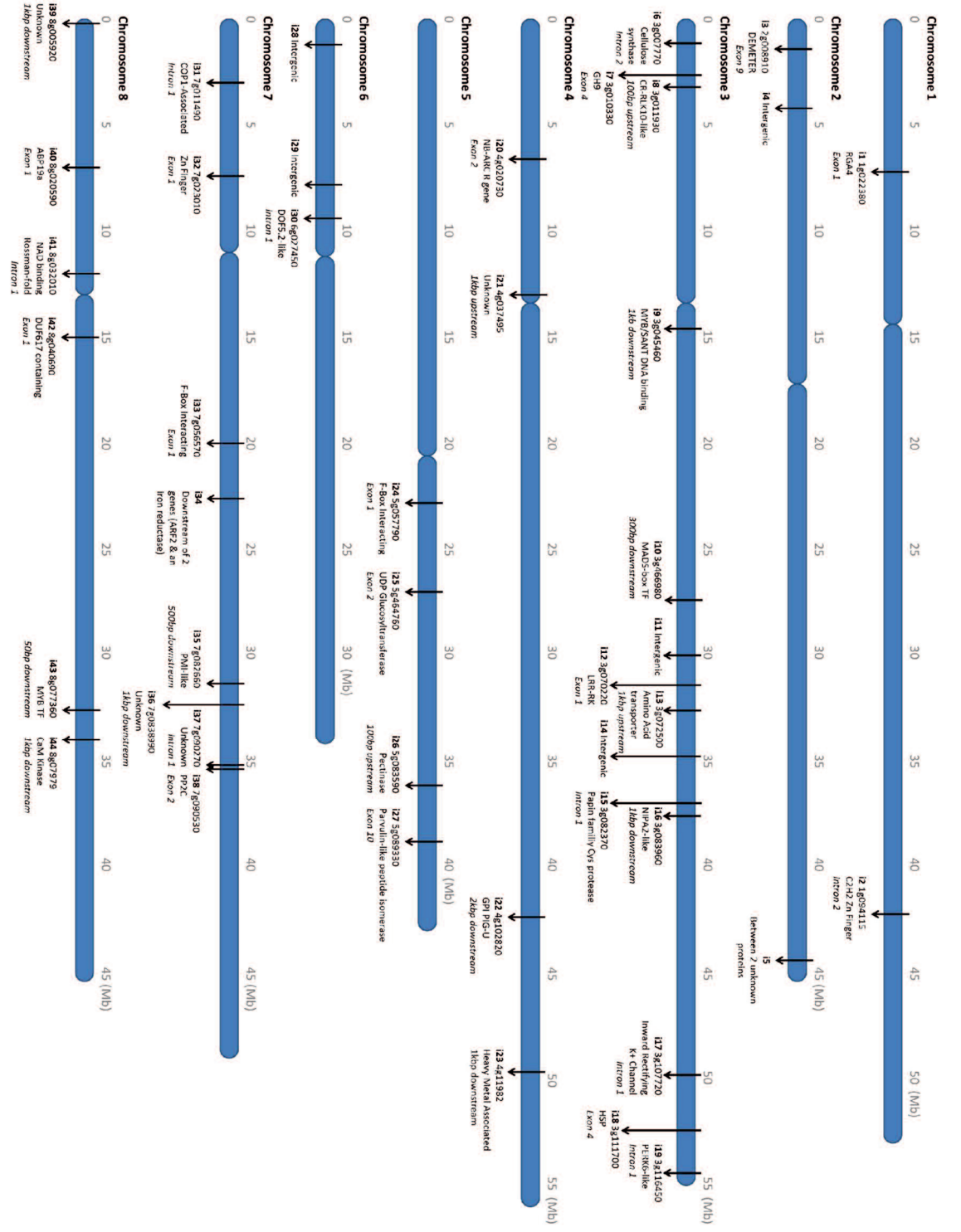


Figure 4.15 – Chromosome map of NF13209 showing location of all tnt1 insertions
Map of the chromosomes of *M. truncatula* (based on the A17 v4.1 genome assembly). Each arrow marks the location of a tnt1 insertion found in the WGS dataset from NF3209, and lists the location relative to the nearest gene, and the numeric code and predicted function of said gene.

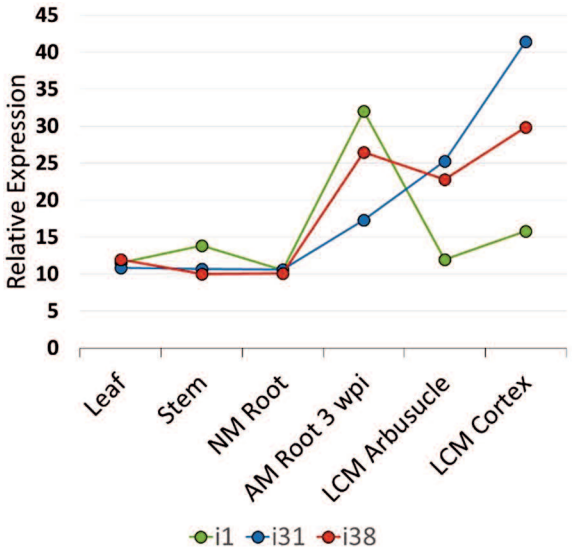


Figure 4.16 – Expression of genes affect by candidate insertions in NF3209
Expression of three genes at tnt1 insertion loci in NF3209 considered candidates for causing the delayed colonisation phenotype of that line. Expression data is derived from the Noble Foundation's *Medicago* Expression Atlas[313].

4.5.5 – NF3438: likely candidates

NF3438 has 42 locatable insertions (**Figure 4.17**), with a slight bias in distribution towards chromosomes 3, 5 and 8. There are 4 insertions which fit our criteria for likely candidates to cause the reduced colonisation phenotype (**Table 4.11**). The expression pattern for the gene disrupted by three of these insertions was available, and is presented in **Figure 4.18**. The fourth insertion, i25, affects a GRAS transcription factor, many of which are known to be involved in the AMS[158].

Table 4.11 – Promising candidate insertions for causing the mycorrhizal phenotype in N3438

Insertion	Nearest gene	Location relative to the nearest gene	Predicted gene function
15	4g108210	Exon 2	ABC-A family transporter
25	7g027190	Exon 1	GRAS family transcription factor
27	8g005740	Exon 1	Phospholipid transporting ATPase
34	8g103227	Exon 4	LIN-like E3 Ubiquitin ligase

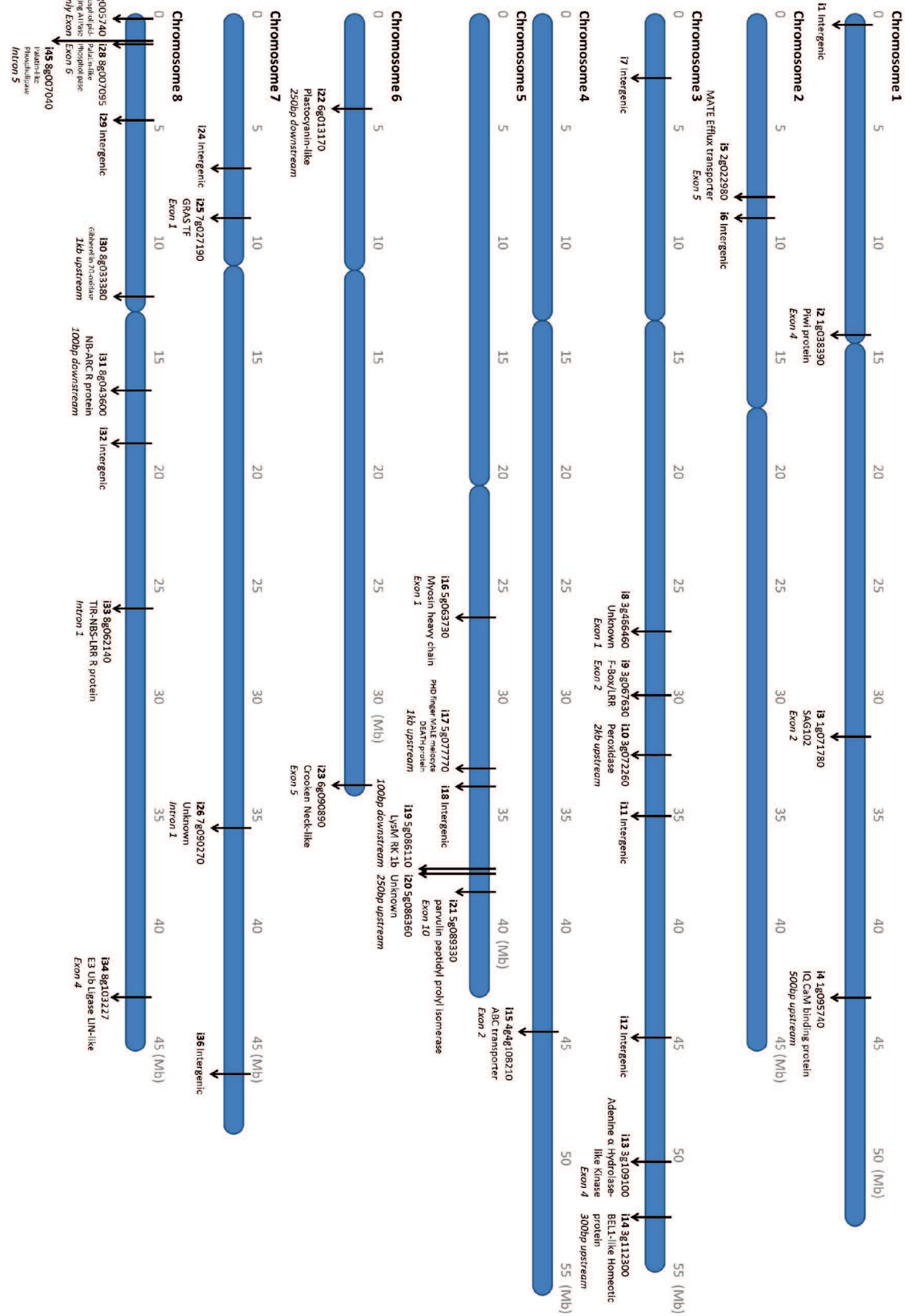


Figure 4.17 – Chromosome map of NF3438 showing location of all tnt1 insertions
Map of the chromosomes of M. truncatula (based on the A17 v4.1 genome assembly). Each arrow marks the location of a tnt1 insertion found in the WGS dataset from NF3438, and lists the location relative to the nearest gene, and the numeric code and predicted function of said gene.

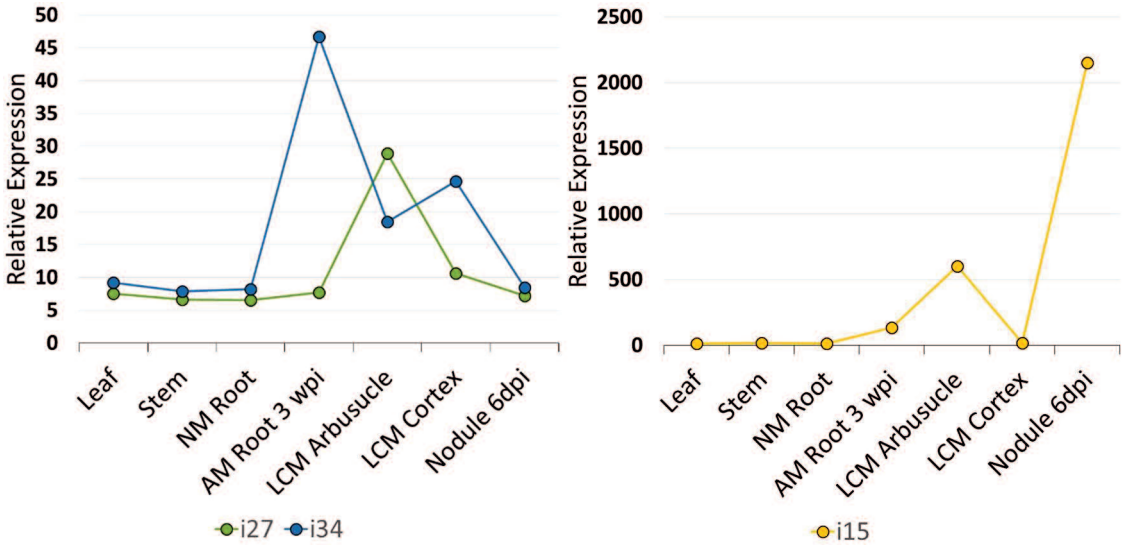


Figure 4.18 – Expression of genes affect by candidate insertions in NF3438
Expression of three genes disrupted by insertions in NF3438. Expression data is derived from the Noble Foundation's *Medicago* Expression Atlas[313]. Nodule expression data is included to show the duel symbiotic expression of 4g108210, disrupted by i15.

4.6 – Discussion

4.6.1 – Missing insertions from the invPCR dataset

As described in **Chapter 5**, backcrossing and co-segregation of three mutant lines (NF443, NF3209 and NF3438) showed that none of the insertions found by invPCR were causal. Further work on the invPCR was considered, but digestion of NF3209 DNA with the enzyme pairs EcoR1 & Mfe1 and Ase1 & Nde1 suggested that we had already obtained sequences corresponding to each of the bands seen on an electrophoresis gel after the second round of amplification. But, with WGS we found 57 tnt1 insertions (that could each be assigned to a single genomic location; see **Table 4.3**), and only 12 with invPCR (**Table 4.2**). invPCR is limited by the need for a pair of restriction sites either side of the target of interest, giving a product within a size range of 2-3 kb after circularisation and amplification with tnt1 flanking primers (see **Figure 4.1**). Best results require Sanger sequencing of the whole length from tnt1 flank to flank, a practical limitation of ~1.2 kb. Even with common restriction sites like EcoR1, their presence at this frequency is far from guaranteed.

Using our WGS data, and the R108 genome data available[297], we reconstructed the region around each of the 35 insertions not found in the invPCR dataset, and looked for paired compatible (e.g two EcoR1 sites or an EcoR1 site and a Mfe1 site) restriction

sites around the location of tnt1 insertion, giving an amplicon of no more than 2 kb. We found that only 19 of these insertions met this requirement with one or both of the pairs of restriction enzymes used in this study (EcoR1/Mfe1 or Ase1/Nde1; see **Figure 4.19**). Most of these insertions could be found by Ase1/Nde1, and relatively few with EcoR1/Mfe1. This left 16 insertions that did not have paired restriction sites within 2 kb. It is possible some of these could have been found as FST with long (~3 kb) amplicons, or by using other sets of enzymes. However, each additional enzyme set used adds a substantial time/money cost, as most amplicons will recapitulate already known insertions.

Figure 4.19 – 40% of insertions in NF3209 missing from the invPCR dataset could not have been found with the restriction enzymes used
35 tnt1 insertions were found by WGS, but not with invPCR, in NF3209. These are sorted to show those with a <2 kb invPCR amplicon with restriction enzyme sets Ase1/Nde1 or EcoR1/Mfe1.

4.6.2 – A17 and R108: why so different?

The A17 and R108 ecotypes of *Medicago truncatula* are visually distinct, with A17 exhibiting a more sprawling growth character and truncate leaves. R108 is slightly more compact, with orbicular leaves bearing a smaller chevron. These differences extend into changes in metabolism, including effects on nutrient deficiency[327], salt stress[328] and defence[329]. But in comparison to other model plants like *Arabidopsis*, these changes seem within the normal in-species variation. However, A17 and R108 exhibit significant hybrid sterility, and have undergone significant chromosome rearrangement events (see **Section 4.4.7** and Kamphuis et al (2007)[320]) since their last common ancestor. We had initially supposed that these differences had been caused

by the extensive callus culture process that R108 had been taken through to improve its transformability[330]. Callus culture is known to induce many genetic and epigenetic changes, up to and including chromosomal exchange[331].

However, Kamphuis et al (2007)[320] suggests that A17, not R108, is the outlier, showing similar sterility issues with other accessions (including A20 and J7). This immediately raises questions about the value of using the A17 genome as the primary source for genetics in *M. truncatula*, if it shows such genetic distance from the majority of ecotypes. Clearly, producing an independent assembly of the R108 genome is needed for the continued growth in use of R108-based genetic resources. This includes not just the Noble Foundation's tnt1 collection, but with techniques like RNAi and CRISPR to quickly produce KO lines, the transformability of R108 will push more work into that ecotype.

Ideally, we would see more *Medicago* ecotypes, especially those with distinctive phenotypes, sequenced. Looking at the genetic differences between ecotypes to explain functional differences has been the basis of the 1001 *Arabidopsis* genome project, and smaller scale approaches are paying off in *Medicago* as well[332]. A very important application of this technique would to establish a genetic basis for the enhanced transformability/culture tolerance seen in R108. If this could be replicated in other species currently recalcitrant to genetic modification, it would be a great boon to plant genetics. However, with such extensive differences between R108 and A17, such an approach is unlikely to find a causal locus. We propose that such an experiment should instead aim to examine variation between R108 and other ecotypes that lack such large genetic differences (perhaps after an initial screen for ecotypes with no evidence of hybrid incapability). With current sequencing costs, the down-costs on such an experiment would not be large[ii]. Given that collections of recombinant inbred lines between *Medicago* ecotypes already exist[333], screening such for differences in the dependence on mycorrhizal colonisation for P nutrition and MGR

[ii] Three lines could be comfortably sequenced on a single Illumina Hi-Seq lane (cost <£3000), perhaps with secondary sequencing on a high processivity platform to allow full assembly of each. The total cost is unlikely to exceed £10,000, and looking for genetic differences between R108 and the three lines should be enough to narrow down this hypothetical locus controlling transformability.

would appear to be a useful endeavour. Looking for pre-existing natural variation can show us which genes are adapted by evolution to moderate the AMS, and thus which genes are good targets for breeding to improve crop mycorrhization.

4.6.3 – The role of WGS in future forward genetic screens

The WGS pipeline allowed us to find a much greater number of insertions in our NF *Medicago* lines, with less lab worker time, and a comparable cost, to previous invPCR approaches. Is this approach more widely applicable? For insertion mutants with small numbers of insertions (e.g. T-DNA mutants) invPCR remains the more efficient first approach. The small number of insertions means that saturating the invPCR amplicon population will not include a lot of inefficient replication. However, these insertions are no more likely than TE insertions to be flanked by suitable restriction sites. Like the Noble Foundation's tnt1 mutants, they will have gone through callus culture, so potentially have somaclonal mutations causing the observed phenotype. If co-segregation of a T-DNA mutation located by invPCR (or tailPCR) shows that mutation is not causing the phenotype, WGS to look for other mutations is probably the best approach. A SNP-based mutant platform is well served by taking a WGS approach, although obviously our bioinformatics pipeline would need to be replaced with appropriate SNP-calling software. For forward genetics screens using the Noble Foundation collection, or similar mutants generated by random insertion of high-copy known sequences, then we suggest that WGS using our pipeline, or a similar one[306], is the most efficient approach, and will only get more so as sequencing cost decreases. Other groups have begun using WGS as their standard approach[306]. Our pipeline can be further streamlined to reduce operator input requirements, although to approach a graphical user interface-based system to minimise the need for specific knowledge of backend programming by the user would require a large amount of professional coding work and to bring the system to a Windows operating system. Given the large amount of work involved for a reasonably specific need, this is probably not worth it.

New molecular genetics techniques like CRISPR/Cas9[334] and 'genetic coding' approaches (e.g. GoldenGate[335]) are likely to make producing highly targeted mutations far more straightforward in the near future. This will clearly be a major boon for reverse genetic approaches. Simultaneously, the cost of RNAseq is dropping

precipitously, enabling target genes to be quickly identified. Does this mean that forward genetics will inevitably be pushed into the past?

Forward genetic approaches offer an objectivity of analysis that reverse genetics lacks (as reverse genetics requires pre-selection of targets), and is able to discover changes related to translational or post-translational control of gene products that RNAseq would not pick up (although proteomic approaches also address this problem, and forward genetic struggles with redundant genes). Forward genetics is most powerful when used to examine poorly understood genetic processes, as they will inevitably recapitulate known genes in an already studied system (as **Chapter 5** shows). This can make finding new genes in a well-studied system very expensive. Forward genetic screens also produce of a large mutant population, a very expensive process. But, once such a population has been established, it can be used to answer essentially any question for the given plant species. Similarly, these mutant populations are also excellent resources for reverse genetics, alleviating the need for a lab to produce their own KO lines. When producing such a population, a careful balance is required. We want enough mutations per plant to not require tens of thousands of plants to be screened to find our phenotypes, but too many mutations make post-initial screening work more difficult. Large number of mutations also makes genotyping the segregating population more time consuming, and raises the chances of two or more phenotype producing genes being disrupted, which makes co-segregation analysis difficult. Obviously, the ideal number of mutations per line in a population differs depending on the number of genes involved in the process being tested (and thus the chance of hitting more than one of them).

Thus, do we think the Noble Foundation tnt1 collection is the ideal for forward genetics[iii]? It is better than a T-DNA insertion population, as these tend to have too few mutations (1-2 per plant)[336] for efficient screening. Tnt1 mutant libraries are also possible in species beyond *Medicago*[337], although the possibility of defence related

[iii] T-DNA is a better candidate for reverse genetic approaches if large libraries are created[415], as the lower event number relative to retrotransposon or chemical mutagenesis means there are fewer alterations to overall plant function that could cause spurious correlations in later experiments (e.g. other genes being knocked out and showing up as highly regulated in RNAseq).

activation in some species may be problematic for maintaining a stable collection[338]. But, with sequencing becoming routine, we think that the future probably lies in mutant populations produced by EMS, or by similar chemical mutagens[339]. With a good quality genome available, calling EMS-induced SNPs should be relatively straightforward, and the procedure produces a lot of mutations (>1000 per treated seed[336]), which reduces the number of lines that need to be screened. EMS also removes the need for regeneration of stable transformants (as tnt1 or T-DNA insertion requires in *Medicago*), which reduces the capital cost of establishing a mutant population. Once locating EMS mutations is routine, the only weakness of such a population would be that it only produces G:C→A:T transitions. Ideally, other chemicals or irradiants[339] can be established that will overcome this problem. The Noble Foundations tnt1 *Medicago* collection has already cleared the hurdle of capital investment, so will present an excellent resource for years to come. However, for screening projects in less established species we would recommend chemical induction of SNP or INDELS as the ideal strategy for forward genetics in the age of genomics.

In the work presented above, we generated WGS data from DNA extracted from a single plant per mutant line, each a sibling from a selfed population that stably retained the MYC phenotype, but was segregating for other tnt1 insertions. Ideally, we would have used DNA from the mutant parent that produced our F_2 crosses, but this had not been collected (and the plants were no longer available) when we decided to sequence the lines. But, in future work, should we WGS a single plant, or a number of plants that retain the phenotype of interest? The latter approach should, in theory, allow us to computationally reduce the number of insertions that require PCR genotyping in the F_2, even more so than zygosity calling (**Section 4.4.3**). This technique would work best on a population that is as diverse as possible, but where all individuals retain a causal insertion. If the initial seed stock was not diverse enough (i.e. containing many stable insertions), then backcrossing and selection for plants retaining the phenotype could be used to re-introduce that diversity, and reduce the number of stable insertions per plant. Then, a pooled DNA sample from a panel of these plants could be sequenced, and the read frequency of the insertion and non-insertion alleles of each insertion locus examined. Only those insertions which are present in all sequencing reads across their locus can be candidates for causing the

phenotype. This approach has the possibility to drastically reduce the amount of PCR genotyping required, but may require a second round of backcrossing to be carried out if more than one insertion is found in all sequenced plants. While multiple rounds of backcrossing takes longer, backcrossing requires less researcher-hours than large amounts of genotyping. While this approach may also be used for SNP populations, it fails to work with dominant mutations[iv]. Pooling plant samples also requires that the researcher be confident that the phenotype is controlled by a single locus, and that they can confidently distinguish it from the natural variation seen in wild type plant/AMF interactions. Adding a non-mutant sample to the pool would not completely invalidate the technique, as the sequencing data could still be used to find insertions to genotype for, but it would potentially waste a lot of time before this mistake was noticed.

4.6.4 – The role of *Medicago truncatula* in future work

In this study, we used *M. truncatula* due to the power of the Noble Foundation tnt1 collection, and prior lab experience with the species. However, it is far from the only model system for the AMS. The go-to model, *Arabidopsis thaliana*, does not interact with any mycorrhizal symbiont, but legumes *M. truncatula* and *Lotus japonicus* have both been used extensively, and *Zea mays, Oryza sativa, Solanum tuberosum, S. lycopersicum* and *Petunia hybrida* all have their advocates. The use of the legumes can be both an advantage (as we already know a lot about the CSP from the RNS) and a disadvantage (as we must always ask how much the addition of the RNS has altered the AMS in the legumes, and thus how applicable our understanding of such a system is to non-legume species). Both *M. truncatula* and *L. japonicus* are easier to work with in the lab than most crop plants (e.g. smaller, shorter generation time), but then require findings in these systems be translated to agriculturally applicable organisms. Of the crops, *O. sativa* has the smallest genome, of equivalent size to the model

[iv] Allele frequency measures can be confused by a mix of homozygous and heterozygous individuals. A recessive mutation must be homozygous to cause the phenotype, so we can exclude any insertion with less 100% allele frequency in the population pool. But a dominant mutation gives a phenotype when heterozygous or homozygous (giving an allele frequency of between 50 and 100%). Thus, we can conceive of two hypothetical populations which each have a 75% population allele frequency. In one, a dominant mutation is present in all plants (in half as a homozygote, in the other half as a heterozygote). In the other, the insertion is a spurious match, present as a homozygote in 75% of the plants in the population, and absent in the remainder. Looking at the allele frequency, both these hypothetical populations are indistinguishable.

legumes and, like the legumes, has excellent genetic resources available and a history of transformation. It does however have a longer generation time, making crossing schemes more complex. The *Solanum* species are probably outdated as models, with long generation times, large genomes and (in the case of *S. tuberosum*) higher ploidy. *S. tuberosum* retains some importance, being one of the easiest models to asexually propagate, and one of the most important crops (although lagging behind *Z. mays* and *O. sativa*). *Petunia hybrida* has also been promoted as a model organism in the last few years[340], and seems to offer a compelling alternative to the legume models. The ease of crossing and the much larger number of seeds obtained per cross make it a good alternative for screens requiring large amounts of backcrossing, and asexual propagation is another useful tool. The large genome size and more complex genome (being hybrid between two landraces)[341] are drawbacks, however. We agree with Bombarely et al (2016)[340] that *Petunia* will likely rise in prominence as the genetic resources are established. We also expect the trend of working directly with crops (*O. sativa* and *Manihot esculenta* in particular) to increase, as improved experimental tools and increasing availably of genome makes the extra hassle of working directly with the end crop competitive to the task of establishing theory in model organisms and then translating it to crops.

Chapter 5 – Identification of candidate genes

5.1 – Introduction

In **Chapter 3**, we have shown that our mutant lines have reproducible impaired mycorrhizal colonisation phenotypes, and in **Chapter 4** we discovered the location of the large number of tnt1 transposons in each line. The next stage, then, was to work out which of the tnt1 insertions caused the phenotype. This was done in two stages. First, we produced a segregating population of plants from each line, by backcrossing plants from our stable mutant population (the F_0; the siblings of the sequenced plants) to the R108 wild type (**Figure 5.1**). This produced offspring which were heterozygous for all the tnt1 insertions that were fixed in the stable mutant population, and either lost or retained heterozygosity for all those insertions which were heterozygous in the F_0. These offspring (hereafter called the F_1) were selfed. Their offspring (the F_2) were expected to show a Mendelian ratio for the genotype of each of the tnt1 insertion fixed in the F_0 (25% having two copies, 25% without any, and 50% with a single copy of the transposon). The F_2 plants were grown with AM inoculum, then phenotyped to look for reduced or wild-type levels of AM colonisation, and genotyped for the candidate insertions. A tnt1 insertion that was causing the reduced AM colonisation phenotype would be expected to segregate with it. Once we found a tnt1 insertion where we had a co-segregation of phenotype and genotype, we would then prove the involvement in the AMS of the gene disrupted by the tnt1 insertion. This could be done in a number of ways, principally by obtaining an independent knock-out line for the gene and finding the same phenotype, or complementing the mutant line with a wild-type copy of the gene and showing this leads to phenotypic rescue.

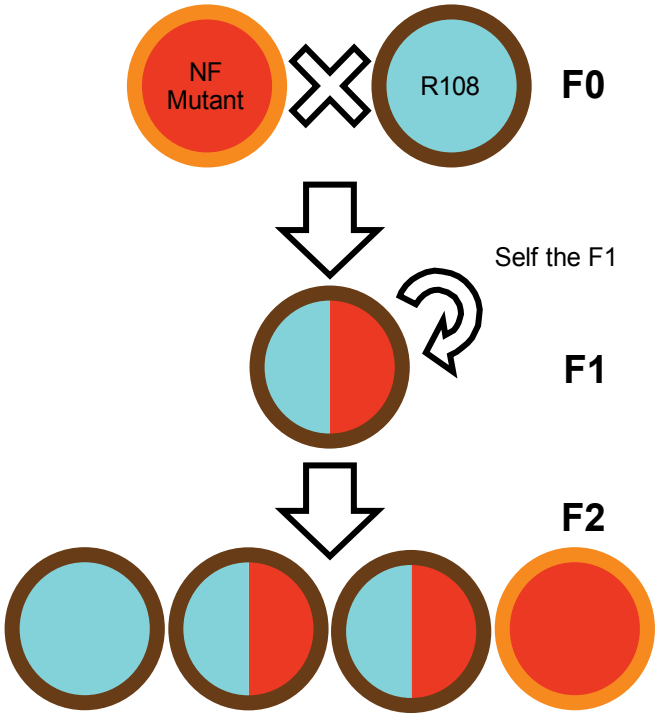


Figure 5.1 – Backcrossing strategy
Visual description of the backcrossing process for a recessive loss of function mutation caused by a tnt1 insertion. Red and cyan represent the presence or absence of the tnt1 insertion causing the reduced AM colonisation phenotype, which is itself shown by the orange outline (whereas the brown outline describes a wild type AM colonisation phenotype).

A priori, we would assume that a tnt1 insertion should be a loss of function mutation, either disrupting a gene transcript or reducing its expression. The tnt1 reverse transcriptase promoter is only active in the germline tissue during cell culture. It is active in the somatic tissue, but any changes in gene expression due to new transposition events should be restricted to single cells or small root sectors, and thus not significantly alter the whole root expression of other genes[273]. It is also assumed that the mutation is likely to be recessive, although a dominant or semi-dominant mutation that reduced colonisation is also a possibility (e.g. if a tnt1 insertion disrupted the 3' end of a coding region sufficient to remove a C terminal regulatory domain, leading to auto-activation of a repressor of AMS signalling). This should be distinguishable by the ratio of mutant to wild-type phenotypes in the F_2 population. A recessive loss of function mutation should have a 1:3 mutant to wild-type phenotype distribution, as only those individuals with two copies of the causal tnt1 insertion should have the mutant phenotype, with heterozygous individuals being phenotypically indistinguishable from the wild-type control. A dominant loss of function mutation would show the opposite pattern of phenotypes, with 3 reduced

AM colonisation individuals for every 1 wild type phenotype in the F_2, which would be only present in those plants lacking tnt1 insertions on either chromatid at the causal locus. A semi-dominant mutation would be intermediary, with heterozygous F_2 plants showing reduced AM colonisation, but not as strongly as the homozygous F_2 or the mutant control plants. In actuality, both the causal loci we were able to locate with co-segregation during this project fit our base assumption, that of a recessive loss of function mutation, with heterozygous F_2 plants showing no difference in colonisation from the R108 wild-type control.

One note of caution must be made with the tnt1 collection. All the plant lines have been taken through callus culture to activate the transposase. It is well known that callus culture leads to changes in the plant epigenome, as well as the formation of new SNPs and INDELs[342]. While most of the epigenetic changes are not stable in the offspring, SNPs can easily become fixed, and both present a possible non-tnt1-linked cause of mycorrhizal phenotypes. These changes would also be more likely to produce dominant mutations than tnt1 insertion would. With the Illumina sequencing we can computationally scan for SNPs in a line that does not show a tnt1-linked phenotype, so these problems are not insurmountable.

5.2 – Methods

5.2.1 – Backcrossing the mutant lines

Medicago plants were grown in 9x9x10cm pots of *Medicago* mix (see **Appendix 4**), watered daily with tap water and fertilised weekly by watering to soil capacity with 2 ml/litre Phostrogen All Purpose Plant Food (Bayer, Leverkusen, Germany). The plants were staked as required, and crossed once in flower, normally 6-8 weeks after planting.

Crossing was performed under a dissecting microscope to the method laid out in Veerappan et al (2014)[343]. Sufficient sepals and petals were removed from un-opened flowers to allow access to the genitalia. Those flowers whose anthers had dehisced were used as pollen donor, and the pollen from these flowers was dabbed onto the

stigma surface of a female donor flower after its locules had been removed. The leaves behind the select female flower were removed, the bud marked with cotton thread, and the branch inserted into a 15 ml polypropylene tube containing 2-3 ml of diH_2O, and sealed with cotton wool to prevent the stigma from drying out. After 2-3 days the tubes were removed and the newly formed pods were wrapped in a ~3 cm section of mesh netting to prevent the seedpod falling from its labelled branch. Pods were collected once they dried and fell. The non-crossed flowers and pods were removed from the plants during this time to maintain flower production.

F_1 seed from the crossed pods were planted into larger 10x10x11cm pots of *Medicago* mix to ensure greater seed yield, grown in the same conditions and watering regimen, and allowed to flower and self-fertilise. Once pod devolvement began the plants were covered in mesh netting to collect the pods as the plant dried.

To produce the F_2 seeds for the co-segregation analysis, pollen from each of the four mutant lines was used to fertilise flowers on R108 wild type plants. This allowed us to control for any pollen being accidently left in the flower from the anthers, as the resulting F_1 would not carry any tnt1 insertions. A total of four NF443-3 (1-4), four NF1436-1 (1-4), two NF3209-4 (2-3) and three NF3438-7 (1-3) F_1 plants were produced, with an additional line (NF3209-4-1) produced from the reciprocal cross (male R108 and female NF3209-4).

5.2.2 – Phenotyping and genotyping the F_2

Assessment of the backcrossed populations was carried out in P24 pots containing 80 ml of 1:4 sand:Terragreen substrate, supplemented with 0.1 g/litre of bonemeal as a non-plant accessible P source to encourage mycorrhizal colonisation. Plants were inoculated with AMF either by addition of 5 ml of PlantWorks RootGrow Professional to the media or by taking between 20% and 50% of the substrate volume from stock pots, along with finely chopped root sections (<2 cm) (see **Table 5.1** for details on inoculation and growth period for each screen, **Section 2.3.3** for details on stock pot inoculum production and **Appendix Table 9.16** for the species of AMF in the different inocula). Panels of F_2 plants were grown for 5 to 7 weeks alongside R108 wild type and the appropriate F_0 plants acting as controls. The greenhouse conditions (see **Appendix**

Table 9.19) were quite variable, with visible changes to size and growth stage of plants grown in different seasons.

Phenotyping of plants was carried out by the ink/acetic acid staining (detailed in **Section 2.5.1**). All data is given as percentages of root length colonised, with the different fungal structures assessed in a hierarchical order (thus, root sections noted as containing vesicles also contain hyphae, and those described as arbuscules contain hyphae and may contain vesicles). Genotyping was performed with triplex PCR (**Section 2.4.3**).

Table 5.1 – Description of the seven F_2 screen performed

Line screened	Date	Growth period	Inoculation source
NF443	July 2015	5 weeks	1/3 volume stock pot (Leek with *R. irregularis)*
NF443	February 2016	5 or 7 weeks	1/3 volume stock pot (Leek with *R. irregularis)*
NF3209	July 2014	6 weeks	5 ml of RootGrow Professional
NF3209	November 2015	6 weeks	1/2 volume stock pot (Leek with *R. irregularis)*
NF3209	June 2016	5 weeks	1/5 volume stock pot (Leek with *R. irregularis)*
NF3438	August 2014	5 weeks	5 ml of RootGrow Professional
NF3438	December 2014	6 weeks	1/3 volume stock pot (Chive with four species mix)

5.3 – NF443 co-segregation analysis

5.3.1 – F_2 Screen (July 2015)

18 NF443-3-1 F_2 plants and 3 plants of each R108 and NF443 F_0 controls were grown for 5 weeks. This initial screen was intended to weed out candidate insertions clearly not linked to the phenotype. Thus, to reduce the time spent phenotyping the population, the stained roots were qualitatively assessed by comparison to the phenotype of the controls. Those that exhibited well-formed, cell filling arbuscules in multiple cell files along the cortex, and AM structures across at least half of the root (as the R108 roots did) were considered 'High' colonisation, and those that had few or no mature arbuscules and low colonisation (as the NF443-3 roots) were considered 'Low' colonisation. 7 F_2 plants were assigned as high colonisation, 6 as low

colonisation, and 5 with intermediate phenotypes were discarded from this preliminary analysis.

The 13 selected F_2 plants and one of each control plant genotype selected at random were then genotyped for 28 insertions discovered by WGS of the parental NF443-3 line. In this High-Low colonisation screen, we would expect that a causal insertion should have homozygous tnt1 insertions in the low colonisation samples and either heterozygous or wild-type genotype in the high colonisation samples if the insertion produced a recessive mutation. Alternately, if the mutation was dominant or semi-dominant, should the plant be homozygous or heterozygous for the tnt1 insertion in the low colonisation samples, and have the wild-type genotype with high colonisation. Due to the preliminary nature of this screen, and the qualitative rather than quantitative assigning of phenotypes leading to potential experimental error, it was not expected to see a perfect distribution of genotypes even in a causal insertion, but a good candidate should show a distribution of genotypes that largely fit with the expected genotype for the phenotype category.

No insertion showed the exact genotype distribution expected of a causal insertion, but that several showed patterns close enough to be considered good candidates for further study (**Figure 5.2**). These were i41 and i42, which both showed up in all but one of the low colonisation plants as homozygous tnt1 insertions, and lacked any homozygous insertions in the high colonisation plants. They did, however, appear as heterozygous in the NF443-3 mutant control. This could be explained by a dominant mutation, but they were also heterozygous in many high colonisation plants, a pattern incompatible with that hypothesis.

Another insertion locus, i39, presented a secondary candidate for future screening. All but one low colonisation plant had the homozygous tnt1 insertion. However, two of the seven of the high colonisation plants also had the insertion on both chromosomes, though the majority had the wild type genotype. Insertion 39 is also into an intron (unlike i41 and i42, located in exons), thus the insertion is less likely to be affecting the protein product. The gene (Mt7g090270) with i39 in its first intron is an unknown protein with a consistent expression level throughout the plant. It is conserved in

many other dicots. The closest *Arabidopsis thaliana* homolog is a histidine-tRNA ligase, which shares a 56 amino acid α helical domain of unknown function at the N terminus of the protein, but has no similarity to the C terminal half of Mt7g090270. It is hard to envision a role for a tRNA loading cofactor in the AMS. Nevertheless the association between phenotype and genotype means it bears further investigation.

Insertion 27 presents the opposite pattern of what was expected, with the high colonisation plants showing a homozygous insertion genotype, and the low colonisation plants mainly heterozygous. This would be the pattern expected of a knockout of a gene that repressed the AM symbiosis. While such a mutant should not have been found with the screening method used, it could potentially have been hidden by a stronger negatively impacting mutation in the NF443 parent. Alternatively, this association could simply be down to chance. Given the large number of samples tested, we must be cautious of false positives.

Two of the candidates highlighted in **Section 4.5.1**; i16 and i43, no longer appear to be good candidates given this dataset. Multiple low colonisation plants show non-homozygous insertion genotypes at these loci, and NF443-3-1 F_2 plant A2, which has a high colonisation phenotype, has two copies of tnt1 at i43. Insertion 36, the LRR-RK suggested as the most interesting candidate gene in **Section 4.5.1**, was tested for in this screen, but the primer set did not produce any bands, as did primers for i29 and i40. Insertions 6, 13, 20, 22, 24, 26 and 37 were also tested, and the entire F_2 population found to be wild type, suggesting that these insertions were heterozygous in the NF443-3 parent sequenced, and have subsequently been lost from the NF443-3-1 F_2. These insertions are therefore considered non-causal, since plants in the F_2 still show the reduced colonisation phenotype. Zygosity of the sequenced line was assessed *in silico* (see **Section 4.4.3** for methodology; data not shown), and suggested that most of the above insertions were indeed heterozygous in the parental line, with the remainder being unable to be confidently called due to repetitive sequence leading to potential off-target alignment in the region of interest.

		Number of tnt1 present at insertion locus								
		Exon Insertions								
		i4	i10	i16	i21	i33	i41	i42	i43	i45
High Colonisation Plants	R108	0	0	0	0	0	0	0	0	0
	A1	NA	0	NA	0	NA	1	0	1	1
	A2	NA	NA	NA	NT	2	1	1	2	1
	A5	NA	NA	0	1	2	0	0	0	0
	A6	1	2	0	0	0	0	0	0	1
	B6	1	NA	1	NT	NA	1	1	0	1
	C1	1	0	0	NT	0	0	0	1	2
	C4	2	NA	1	0	1	1	0	NA	1
Low Colonisation Plants	NF443	1	2	1	1	2	1	1	2	2
	A3	2	2	1	NT	0	2	2	2	1
	B2	2	NA	2	NT	1	2	2	0	0
	B5	2	1	0	0	2	2	2	2	0
	C2	0	1	NA	0	0	NA	2	1	2
	C3	1	0	2	0	2	2	2	1	0
	C6	2	0	0	0	1	1	0	2	2

		Promoter Insertions					Intron Insertions			
		i2	i7	i25	i28	i38	i19	i27	i30	i39
High Colonisation Plants	R108	NA	0	0	0	0	0	0	NA	0
	A1	1	0	0	1	0	0	2	1	2
	A2	1	1	1	1	1	0	2	NA	1
	A5	2	NA	0	2	0	0	2	0	0
	A6	0	2	1	1	2	0	2	NA	0
	B6	1	2	1	2	1	0	2	1	0
	C1	0	1	0	NA	2	0	0	1	0
	C4	1	1	2	1	1	0	2	NA	2
Low Colonisation Plants	NF443	NA	2	2	2	2	0	2	1	2
	A3	0	1	1	0	1	0	1	0	2
	B2	1	1	0	1	1	1	1	1	2
	B5	2	0	1	2	0	0	1	0	2
	C2	0	1	0	0	2	0	0	1	2
	C3	2	2	0	2	2	NA	1	1	2
	C6	1	2	2	0	2	0	1	1	1

Figure 5.2 – NF443 high/low colonisation screen
Number of chromatids carrying the tnt1 transposons at selected NF443 insertion loci of the F_2 individuals with wild type root colonisation (High) or reduced root colonisation (Low) with wild type (R108) and mutant parent (NF443) controls. Insertion loci are subdivided into groups to indicate where in the predicted *Medicago* gene they inserted (promoters defined at <2 kb upstream of the coding region) and colour coded to show a homozygous insertion genotype (red), heterozygous (pink) or homozygous no insertion/wild type (white). 'NA' indicates a PCR result that could not be definitely assigned, and 'NT' indicates that individual was not genotyped for that insertion locus. Not reported are i6, i13, i20, i22, i24, i26, i29, i37 and i40.

One potential question about the dataset is the apparent skew towards a greater number of insertions in the low colonisation plants. Looking at only those insertion loci that contain a mix of genotypes across the F_2 population, the homozygous insertion genotype accounts for 41% of successfully genotyped insertions in the low colonisation plants, compared to 23% in the high colonisation samples. The high colonisation samples come reasonably close to the expected Mendelian distribution of 1:2:1 (35% wild type, 42% heterozygous and 23% homozygous tnt1), while the low colonisation samples have 27% wild type, 32% heterozygous and 41% homozygous for the tnt1 insertion. We do not consider the cause of this to be a methodological issue with the genotyping, as such an issue would have manifested in the high colonisation samples as well. This could be the result of linkage drag from the causal insertion, or the effect of several insertions causing small reductions in mycorrhizal colonisation, but as the observed pattern was not repeated (**Section 5.3.2**), we consider it a chance occurrence.

5.3.2 – F_2 Screen (February 2016)

The qualitative screen had narrowed down the candidate insertions, but did not have sufficient F_2 plants to quantitatively show a significant co-segregation of phenotype and genotype. To correct this, we performed a larger screen, using two time points to examine if NF443 had a reduced colonisation phenotype, or a delayed colonisation phenotype like NF3209. 72 NF443-3-1 F_2 plants and 12 R108 and NF443 F_0 controls were planted, with half grown for 5 weeks, and half for seven weeks. Only one F_2 plant (from the 5 week set) was lost, and three control plants were randomly selected from each time point to have their colonisation assessed.

Assessment of the stained roots showed that the 7 week old plants exhibited a step change in total colonisation, with 10 F_2 plants showing less than 30% total root colonisation, and the rest showing at least 60%, with no intermediate phenotypes. This is what we would expect of a recessive loss of function mutant, with a Mendelian ratio of 10 F_2 plants with a colonisation phenotype that clearly groups with the mutant parent controls, and 26 F_2 plants that group with the wild type controls (**Figure 5.3**). However, this pattern was not seen in the 5 week old plants. Instead we observed a smooth curve of total colonisation, rather than the sudden 30% jump seen in the 7

week plants. This does correlate with an increased variability in colonisation of the control lines at 5 weeks, so this pattern may be due to chance variability (see **Section 5.6.1** for further discussion).

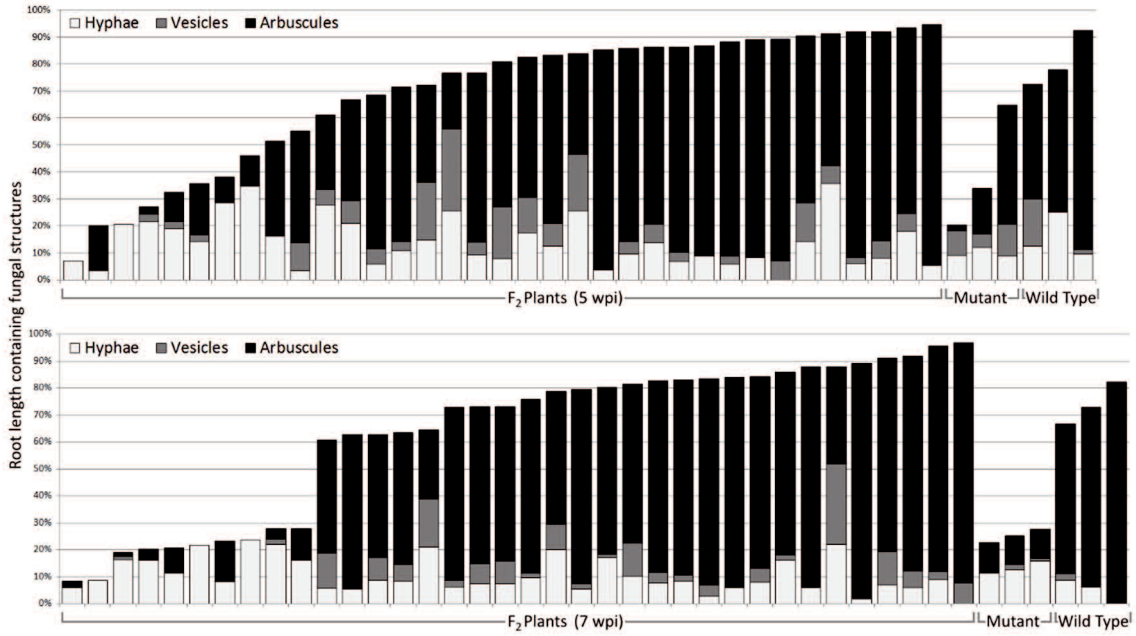


Figure 5.3 – Colonisation of two NF443 F_2 populations
Percentage of root length containing different AM fungal structures of 71 NF443-3-1 F_2 plants at 5 wpi (upper) or 7 wpi (lower). Bars correspond to individual plants ranked by total fungal colonisation, with wild type (R108) and mutant (NF443 F_0) controls displayed on the right.

The 5 week old NF443 F_2 population was genotyped for the three insertion loci (i39, i41 and i42) identified by the initial NF443 F_2 screen as having a strong association between genotype and phenotype. The homozygous tnt1 genotype was significantly correlated with the low colonisation phenotype in each of the three insertion loci (**Figure 5.4**).

Examination of these loci showed that i41 and i42 are near the telomere of chromosome 7q, and are only 83 kb apart, separated by 10 other predicted proteins. Given the average map distance in *Medicago truncatula* chromosome 7 is estimated to be 1.4 cM per Mb[344], if one of these insertions is causal, we will not be able to use co-segregation analysis to distinguish between them, as this would require approximately 860 plants F_2 per segregant which was clearly unfeasible.

Insertion 39 is 12.7 Mb (or approximately 16.8 cM) away from i41. Thus, close enough that a causal insertion would likely cause the other two insertions to be significantly associated by linkage drag. It is sufficiently far removed, however, that around 1 in 6

F_2 plants will segregate these loci. The 7 wpi F_2 plants provide more data, and a more significant distinction between the wild type and parental mutant phenotypes (**Figure 5.5**). In this dataset, it is clear that the variation in colonisation of the wild type has homogenised, moving to 40-80% of root length containing arbuscules (compared to 15-80% at 5 wpi), while the colonisation of the mutant plants had not increased. This is coupled by a pair of segregants between the i39 and i41/42 loci appearing (**Figure 5.5**; two outliers of less than 5% root length containing arbuscules, one in each of the homozygous wild type and heterozygous genotypes). While i39 locus is still significantly associated with the phenotype, the i41/i42 pair is at least two orders of magnitude more associated (p=0.0281 compared to 4.66×10^{-4}; Kruskal-Wallis).

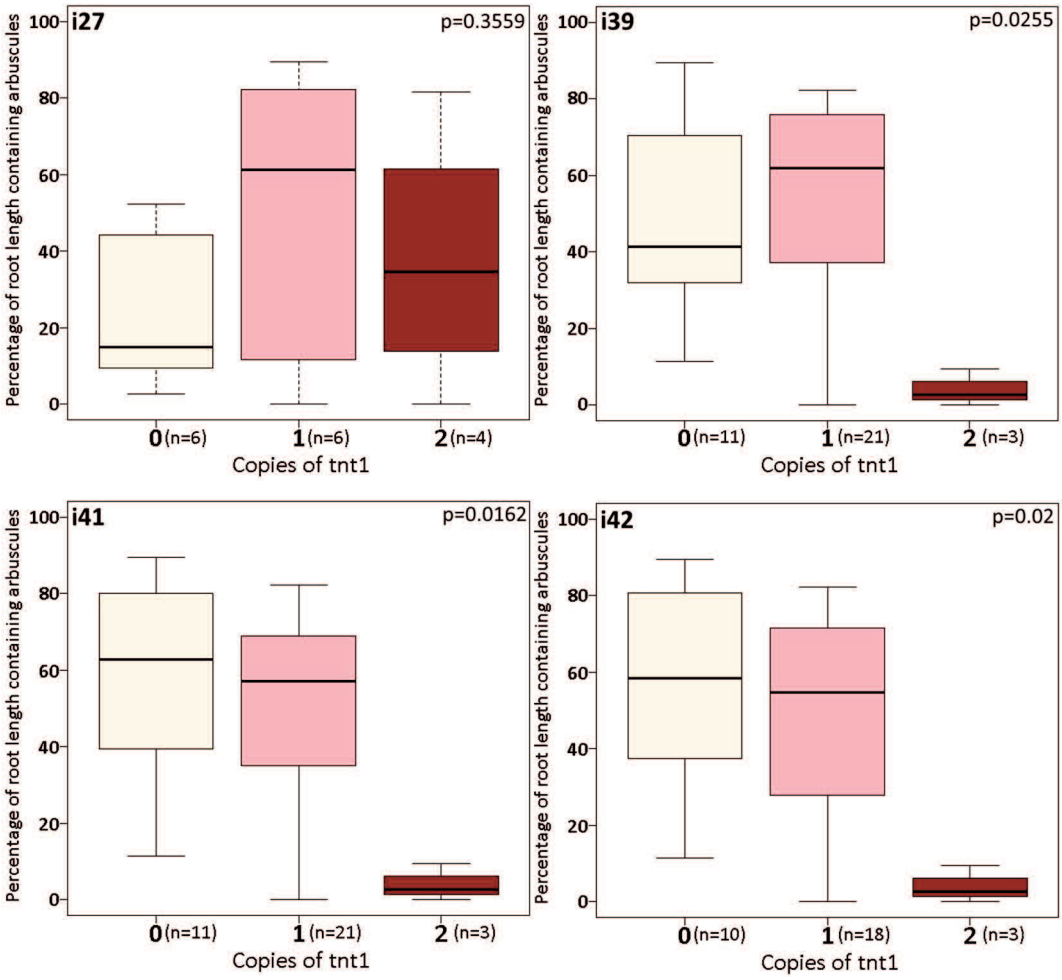


Figure 5.4 – Co-segregation analysis for NF443 at 5 wpi
Correlation of tnt1 insertion genotype with the mycorrhizal colonisation phenotype in the NF443-3-1 F_2 plants 5 wpi. The homozygous insertion genotype is significantly associated with the percentage of root length containing arbuscules present in the F_2 plant roots. P values by Kruskal-Wallis test. Number of samples per genotype given on the x axis (n=#).

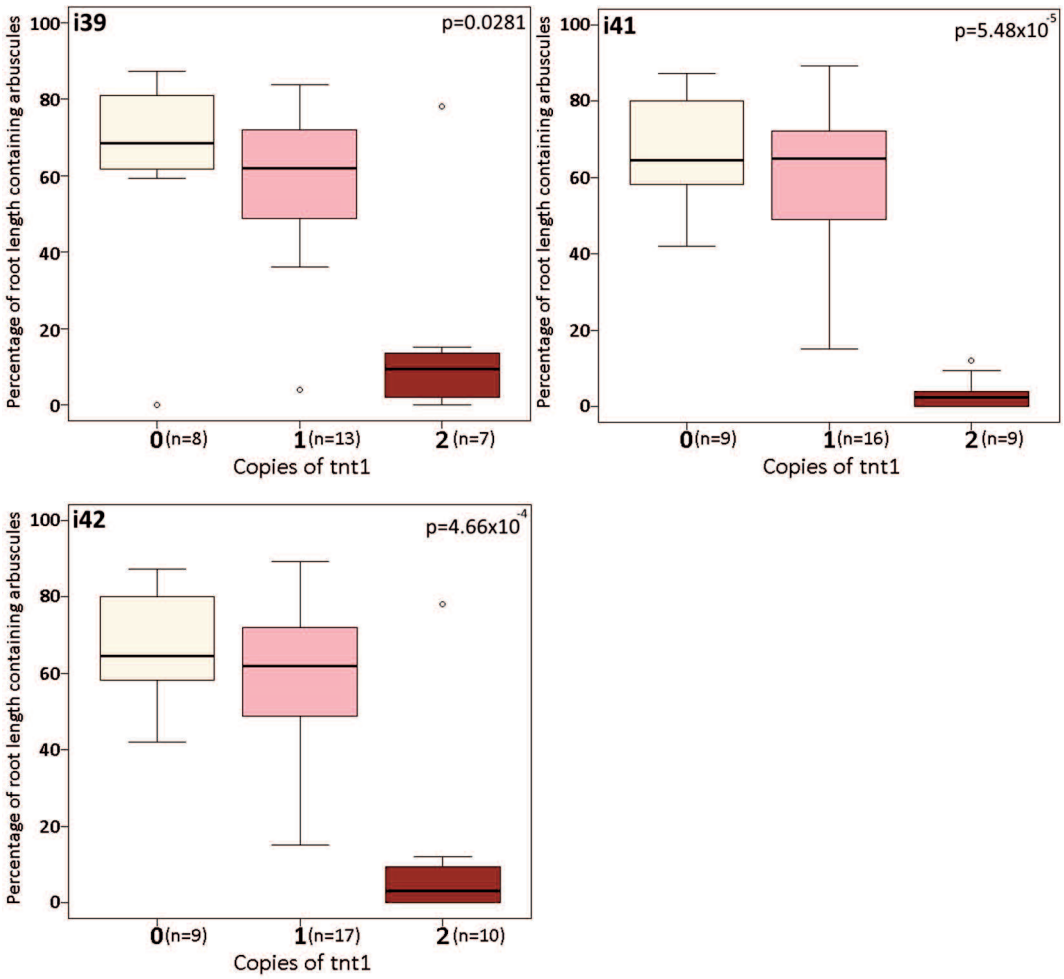


Figure 5.5 – Co-segregation analysis for NF443 at 7 wpi
Correlation of tnt1 insertion genotype with the mycorrhizal colonisation phenotype in the NF443-3-1 F_2 plants 7 wpi. The homozygous insertion genotype is significantly associated with the percentage of root length containing arbuscules present in the F_2 plant roots. P value by Kruskal-Wallis test. Number of samples per genotype given on the x axis (n=#).

It is important to note that while the Kruskal-Wallis test delivers slightly different p values for the phenotype/genotype co-segregation in i41 and i42, this is not due an individual segregant, but instead because not all F_2 individuals were successfully genotyped for each insertion. Plants where a genotype could not be conclusively confirmed were left out of the analysis for that insertion locus. With the 5 wpi samples, the changes in the p value are due to four F_2 plants that were successfully genotyped for i41 (three heterozygotes and one wild type, with root occupation by arbuscules between 35% and 79%) but not for i42. In the 7 wpi F_2, the change in p value and outliers seen in **Figure 5.5** between i41 and i42 are due to i41 lacking genotype data on two individuals which are homozygous for the tnt1 insertion in i42 (78% of the root length occupied by arbuscules) and heterozygous (45.7% arbuscules). Given the map distance between i41 and i42, it is likely that these 6 individuals with

incomplete genotyping have the same zygosity at each locus, but even if a segregant was found, it would be unwise to assign phenotypic causality on the basis of a single data point, given how variable the AM colonisation has proven to be in this work.

We also assessed the whole population genotype at insertion locus 27, which had shown the opposite pattern of colonisation in the high-low screen (**Figure 5.2**). In the 5 wpi F_2 screen there was no difference (p=0.3559, Kruskal-Wallis test) between genotypes, and the heterozygote had the highest average root length colonisation (**Figure 5.4**). Thus, we can conclude that i27 does not produce an AM phenotype, and the original result was likely down to chance.

In summary, we have found a genomic locus that co-segregates with the low colonisation phenotype of NF443. The locus contains two tnt1 insertions (i41 and i42) which cannot be separated by co-segregation, both of which have inserted into gene coding regions. 7g116650 (which has i42 in its sixth of seven exons) is predicted to be a tyrosine kinase expressed in arbuscule containing cells, while 7g116510 (with i41 inserted into the fourth of five exons) is a GDSL lipase that is expressed mainly in the stem. This makes i42 the obvious candidate for causing the phenotype, but further work was needed to prove this.

5.4 – NF3209 co-segregation analysis

5.4.1 – F_2 Screen (July 2014)

30 seeds of each of NF3209-4-1 and NF3209-4-2 were grown for 6 weeks. As with the other lines, the F_2 phenotypes formed a smooth distribution, rather than grouping into more obvious wild type and reduced colonisation phenotypes (**Figure 5.6**). This is less surprising with NF3209, as it showed a weaker phenotype than NF3438 or NF443. R108 had been the pollen donor in the backcross to create NF3209-4-1, and NF3209 the pollen donor to produce NF3209-4-2. The phenotype of the two populations was not significantly different, suggesting that the causal mutation does not show sex specific inheritance.

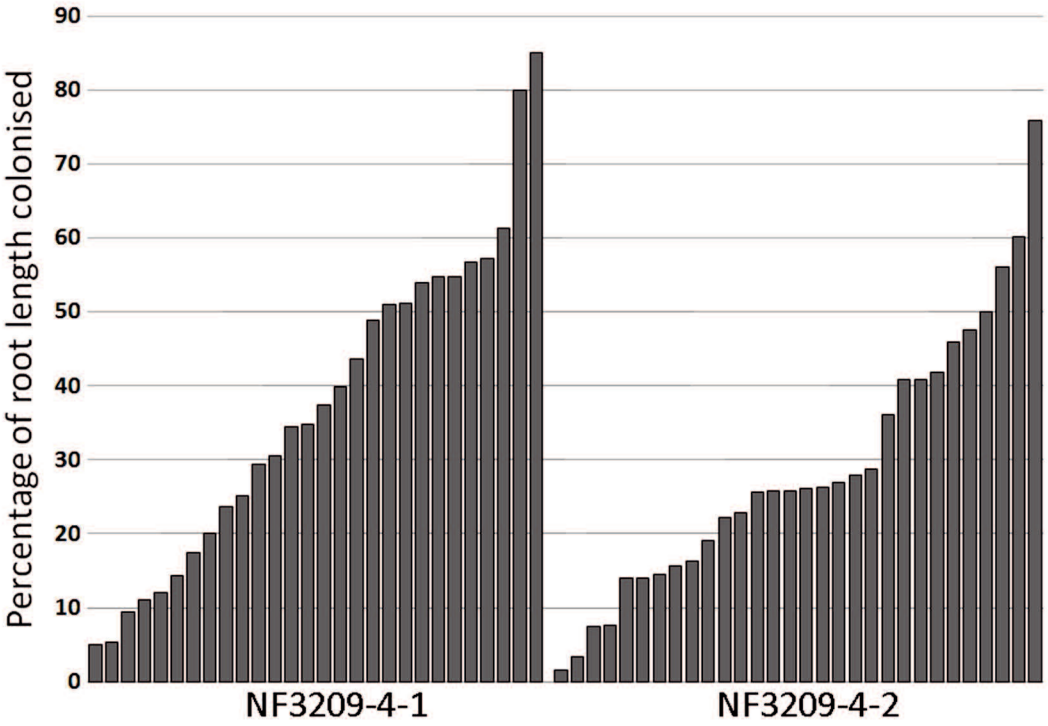


Figure 5.6 – Colonisation of two NF3209 F_2 populations
Percentage of root length containing any AMF structure of 28 NF3209-4-1 F_2 ($\bar{x}$ root length colonisation ± SD of 37.5±21.3%) and 30 NF3209-4-2 F_2 plants (28.9±17.3%), both at 6 wpi. Bars correspond to individual plants ranked by total fungal colonisation, and these populations were not significantly different (p=0.107, student's T test).

A high-low colonisation assay was performed, using 4 individuals with >55% colonisation and 4 with <15%, from each of the two populations, as a bellwether for the overall population. These 15 plants (only 3 from NF3209-4-2 F_2 met the 'high colonisation' criteria) were genotyped for the eight insertion loci within 2 kb of a gene that had been identified by invPCR (**Table 5.2**).

Table 5.2 – NF3209 insertions located by invPCR

Insertion	Location relative to nearest gene	Nearest Gene
A (Seq 1)	Intron 1	3g107720; AKT1-like monovalent cation transporter with stem and seed expression
B (Seq 3)	350 bp downstream	3g466980; putative MADS-box TF
C (Seq 5)	60 bp upstream	5g083590; putative pectinase
D (Seq 12)	130 bp upstream	3g011930; putative cysteine rich receptor kinase expressed in root hairs
E (Seq 14)	Intron 1	3g007920; WEB1-like coiled-coil protein with seed coat and droughted aerial tissue expression
F (Seq 16)	Exon 9	2g008920; *Medicago* orthog of *At*DEMETER
G (Seq24)	Exon 4	3g010330; putative glycoside hydrolase 9, expressed in the root tip and in droughted roots
H (Seq27)	Intron 4	1g025700; putative GDSL lipase, expressed in the petals

Figure 5.7 displays the results of the high-low colonisation screen, and demonstrates that none of the insertion loci correlated with the phenotype. All loci other than C have homozygous tnt1 insertion individuals in the high colonisation category, and insertion C has individuals in the low colonisation category with a wild-type genotype, ruling it even a dominate mutation. Insertion loci B and E are not included, as the primers failed to produce wild-type amplicons.

	Number of tnt1 present at insertion locus											
	A		**C**		**D**		**F**		**G**		**H**	
	4-1	4-2	4-1	4-2	4-1	4-2	4-1	4-2	4-1	4-2	4-1	4-2
High Colonisation	**1**	**1**	**0**	**1**	**2**	**1**	**0**	**0**	**2**	**1**	**2**	**0**
	NA	**2**	*NA*	**0**	**1**	**1**	**1**	**2**	**0**	**1**	**2**	**0**
	0	**1**	**1**	**1**	**1**	**1**	**2**	**1**	**1**	**1**	*NA*	**2**
	1		**1**		**1**		**1**		**1**		**1**	
Low Colonisation	**1**	**1**	**0**	**1**	**2**	**0**	**0**	**0**	**2**	**0**	**2**	**0**
	NA	**2**	**0**	**0**	**1**	**0**	**2**	**2**	**0**	*NA*	**0**	*NA*
	NA	**2**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**
	NA	**0**	**1**	**1**	*NA*	**0**	**0**	**1**	**2**	**0**	**2**	**0**

Figure 5.7 – NF3209 high/low colonisation screen
Results of the high-low colonisation F_2 genotyping. 4-1 and 4-2 refer to the two F_2 populations, derived from plants NF3209-4-1 and NF3209-4-2. Describes the number of transposons at the diploid insertion locus, so '0' implies a wild type genotype lacking any copies tnt1 at the insertion locus, '1' a heterozygote, and '2' a plant homozygous for the tnt1 insertion. NA indicates a failed PCR reaction.

To confirm the efficacy of this high-low screen, the whole NF3209-4-2 F_2 population was genotyped for two randomly chosen insertions, C & G. Neither of these appeared to be a good candidate in the high-low colonisation test, and screening the entire population confirms this (see **Appendix Figure 9.7** for data), as the homozygous insertion plants do not show a significant difference in root colonisation from the homozygous no-insertion plants. A new primer set was established for insertion E, and the NF3209-4-2 population genotyped. Again, it showed no co-segregation of phenotype and genotype.

5.4.2 – F_2 Screen (November 2015)

Completion of the WGS at this point in the project greatly increased our estimate of the number of tnt1 insertions in line NF3209, so we conducted another F_2 screen to test the co-segregation of candidate insertions discovered by WGS. 36 NF3209-4-1 F_2 individuals, along with 6 control plants of the wild type R108 and mutant parent

NF3209, were grown for 6 weeks (October through November 2015). We increased the strength of the inoculum, as the R108 controls in the previous July 2014 experiment showed relatively low colonisation, and boosting the inoculum strength had allowed better differentiation between mutant and wild type with NF3438 (**Section 5.5.2**). Stained roots were assessed and seen to be heavily colonised, in both sets of controls (**Figure 5.8**). The mutant control roots contained an average of 52.5% by length of arbuscule colonisation, and the wild type roots 64.1%. This difference was too close ($p=0.1441$; Kruskal-Wallis) to accurately ascribe a phenotype to the F_2 population, so F_2 root colonisation was not assessed.

This screen, with the strongest inoculum used in the NF3209 screen to date, suggests that a sufficient abundance of AMF can overcome the phenotype of the mutant, rendering it essentially indistinguishable from the wild type. This confirms that the NF3209 mycorrhizal phenotype could be more accurately described as a delayed colonisation phenotype than a reduced colonisation phenotype (see **Section 3.3.3**).

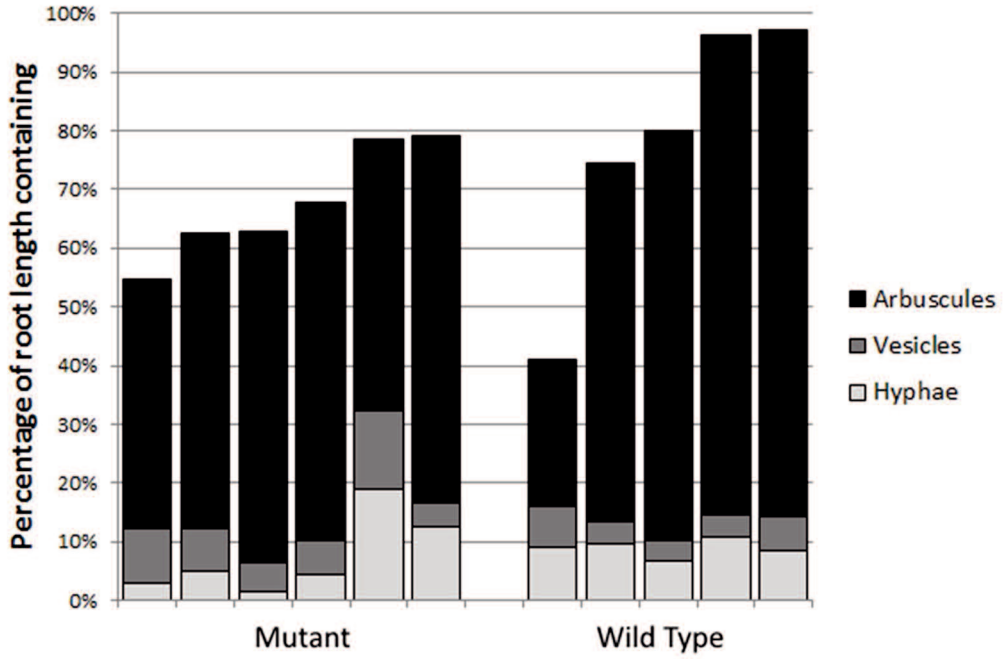


Figure 5.8 – Colonisation of the November 2015 NF3209 screen control plants
Percentage of root length containing different AM fungal structures of 6 wpi control plants. Wild type (R108) and mutant (NF3209 F_0), with bars correspond to individual plants ranked by total fungal colonisation.

5.4.3 – F_2 Screen (June 2016)

Another F_2 screen of NF3209 was carried out, using a shorter growth period and weaker inoculum to improve the distinction between phenotypes in the F_2. 36 NF3209-4-3 F_2 seed and 6 of each R108 and NF3209 F_0 controls were grown for 5 weeks.

As expected, overall colonisation was reduced compared to the November 2015 screen, with an average combined (mature and immature) arbuscule count of 30.6% in the wild type and 5.9% in the mutant controls. While there was a lot of variation within the mutant and wild-type controls, the difference was significant (p=0.0254; Kruskal-Wallis for root length occupation by combined arbuscules), and power analysis suggested that we had a sufficient number of replicates to discover a causal insertion by co-segregation[v]. The colonisation of the F_2 population was assessed (**Figure 5.9**), and showed the same smooth curve of colonisation as in the July 2014 experiment, ranging from ~3% to ~55% of root length occupied by fungal structures.

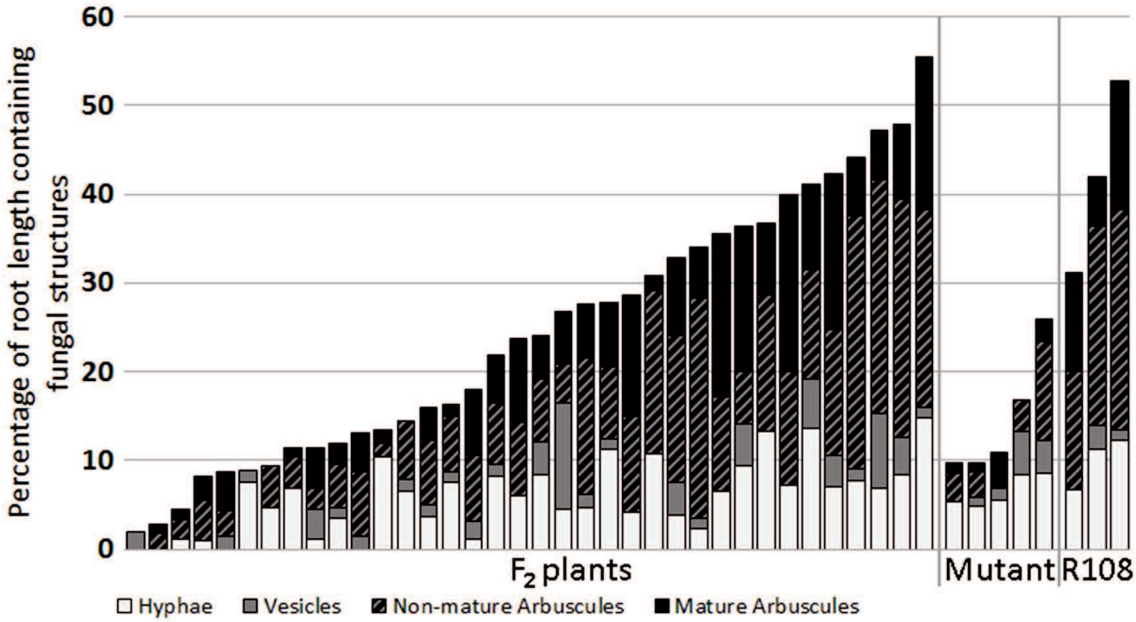


Figure 5.9 – Colonisation of the June 2016 NF3209 F_2 population
Percentage of root length containing different AM fungal structures of 36 NF3209-4-3 F_2 plants at 5 wpi. Mature arbuscules are cell filling and aseptate, whereas non-mature arbuscules may either be partially formed or degenerating. Bars correspond to individual plants ranked by total fungal colonisation, with R108 (wild type) and mutant (NF3209 F_0) controls displayed on the right.

[v] Continuous independent power analysis with false positive and false negative rates of 0.05 gave a minimum sample size of 16 (4 of which would be homozygous for the given tnt1 insertion) to obtain significance. Our 36 plant F_2 panel more than doubles this.

The entire F_2 population was screened for 23 insertions identified by WGS that were within our region of interest (between 2 kb upstream and 500 bp downstream of a gene's coding region). Given the time involved in primer design and triplex PCR genotyping, it was not feasible to test all insertions in NF3209, so we concentrated our efforts on this region, excluding 14 insertions. While there are known mutations that affect gene expression from further away[345], we considered the chance of an insertion outside this region affecting the gene's expression to be highly unlikely. A further 5 insertions had been previously identified by invPCR and screened in **Section 5.4.2**, so these were not tested again. None of the F_2 plants had any copies of the tnt1 insertion at four of the tested insertion loci (i1, i20, i24 and i41). As in NF443, we believe that these insertions were heterozygous in the sequenced plant, and have been lost from the population by segregation. At another locus (i30), despite multiple attempts with different primer sets, none produced an amplicon corresponding to the wild-type sequence. While amplicons for the tnt1-containing sequence were reliably and repeatedly produced, we could not distinguish homozygotes and heterozygotes, so the insertion was excluded from the final analysis. This left 18 insertion loci that could be analysed (see **Appendix Figure 9.8** for co-segregation analysis of each insertion), but none of them was significantly, or even suggestively, associated with the phenotype. For insertion 27, the heterozygous genotype was significantly more colonised than the wild type and the homozygous insertion genotype (Dunn test), but this does not fit with any reasonable model of how a mutation would affect a mycorrhizal gene, so is likely a false positive. Given we tested 23 insertions, finding a false positive ($p<0.05$) would be expected.

The analysis of these screens was inconclusive as to the cause of the delayed colonisation phenotype of NF3209. We consider it highly unlikely that any of the remaining tnt1 insertions could be causal. Therefore, no candidate gene has been identified.

5.5 – NF3438 co-segregation analysis

5.5.1 – F_2 Screen (August 2014)

28 NF3438-7-1 F_2 and 8 R108 wild-type seeds were planted, of which 23 F_2 and 5 wild type germinated. The AM colonisation was assessed, and the wild-type control showed only a low level of colonisation (17±10% of root length containing AMF) with a large spread, one individual almost uncolonised (**Figure 5.10**). This suggests a low strength of the inoculum, leading to inconsistent colonisation as low infectious particle count meant a lot of pot to pot variety in inoculum strength. This makes it hard to draw any accurate conclusion about the phenotypes displayed in the F_2 population. The F_2 colonisation is again distributed in a smooth curve.

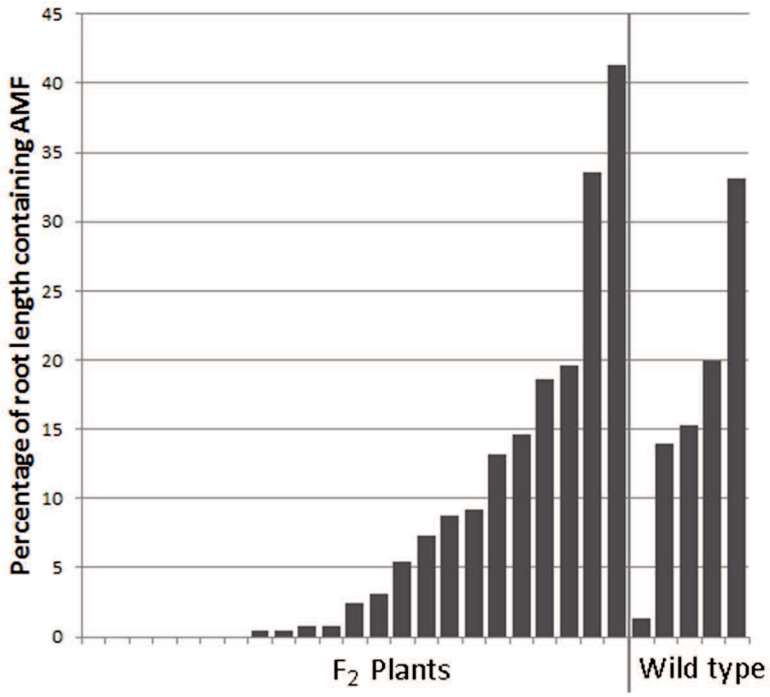


Figure 5.10 – Colonisation of the August 2014 NF3438 F_2 population
Percentage of root length containing AMF structures of 23 NF3438-7-1 F_2 plants at 6 wpi. Bars correspond to individual plants ranked by total fungal colonisation, with 5 wild type (R108) control plants displayed on the right.

Assigning a definite phenotype to each F_2 individual from the data would be unwise given the variation in the wild-type, but association should point us in the direction of an insertion segregating with the phenotype. We genotyped the population for the three tnt1 insertions (A-C; see **Table 5.3**) identified by invPCR. A fourth (D) was not initially tested (it was included in the later December 2014 screen, see **Section 5.5.2**), given it was outside our defined area of interest. Insertion B appeared the best

candidate, as the insertion was likely inside the promoter region; and thus possibly reducing expression. However, when the F_2 population was genotyped, the three insertion loci did not have sufficient number of individuals homozygous for the tnt1 insertion to perform significance testing.

Table 5.3 – NF3438 insertions located by invPCR

Insertion	Location relative to nearest gene	Nearest Gene
A (Seq 9)	25 bp downstreamm	8g028460; putative methyltransferase
B (Seq 10)	500 bp upstream	1g095740; putative IQ CaM binding protein
C (Seq 12)	In Intron 3	2g036040; DTD (protein that detoxifies D-tyrosine by removing it from a loaded tRNA)
D (Seq 11)	2500 bp upsteam	2g025710; putative MBOAT domain protein

5.5.2 – F_2 Screen (December 2014)

Due to the difficulty interpreting the data from the August 2014 screen, another F_2 screen was performed. The stock pots had been growing for six months, so provided a much stronger and more consistent inoculum than the commercial RootGrow inoculum. To further ensure a higher level of colonisation, the growth period was extended to six weeks, with a larger number of plants grown to account for the variability. 48 NF3438-7-2 F_2 seeds were planted alongside 12 R108 wild type seeds. 41 F_2 and 11 wild type individuals germinated and were assessed for their AM colonisation phenotype at 6 wpi. While the wild type showed a greater level of colonisation than before, it was still highly variable (34±16% of root length containing AMF), and the same smooth curve trend was seen in F_2 colonisation, rather than forming obvious wild type and mutant clusters (**Figure 5.11**). In all screens after this, F_0 plants were used as a mutant phenotype control to better assign F_2 individuals to a mutant or wild type phenotype. However, the larger F_2 population ensured that enough data points had been gathered to statistically correlate phenotype and genotype.

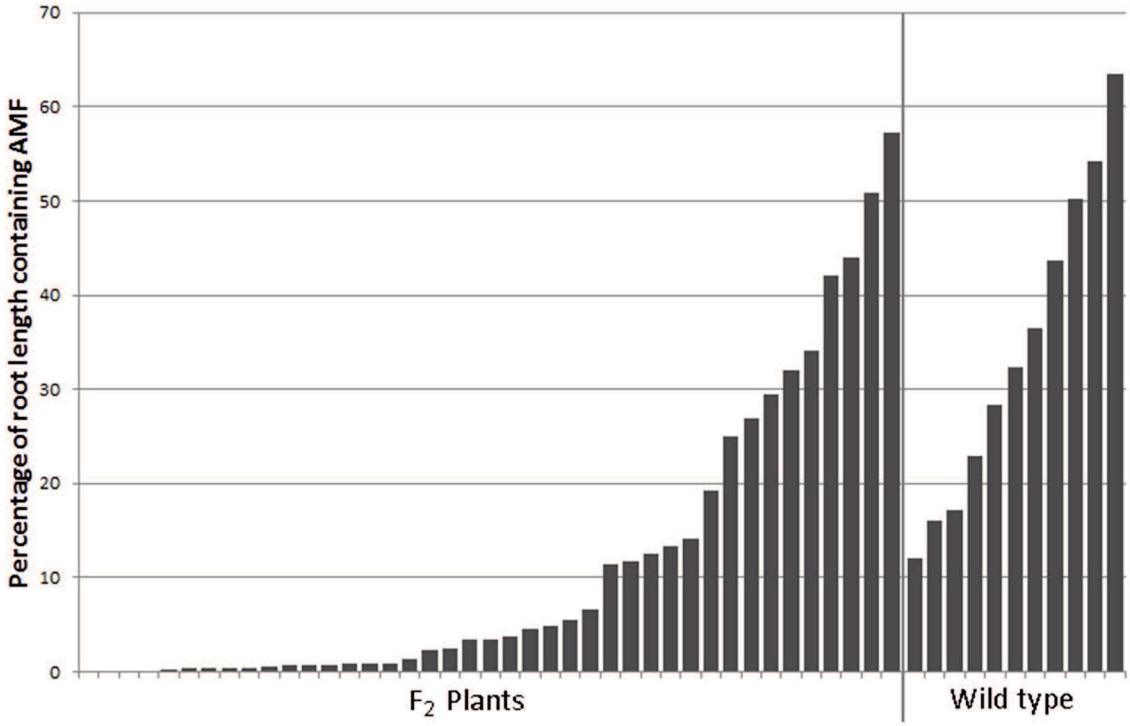


Figure 5.11 – Colonisation of the December 2014 NF3438 F_2 population
Percentage of root length containing AMF structures of 41 NF3438-7-2 F_2 plants at 6 wpi. Bars correspond to individual plants ranked by total fungal colonisation, with 11 wild type (R108) control plants displayed on the right.

An initial high-low colonisation screen was carried out to rapidly rule out a number of insertions, so the entire 41 individual F_2 population did not have to be genotyped for all insertions. The four F_2 plants with the highest levels of root length colonisation were used as the high colonisation exemplars, and four randomly chosen plants with less than 2% root length colonisation were selected as the low colonisation samples. These eight lines were then genotyped for the four insertions described in **Table 5.3**. From this (**Figure 5.12**) we can discount Insertions B, C & D, as all show plants with high colonisation phenotype to be homozygous for the tnt1 insertion. While insertion A lacked any homozygous individuals in the high colonisation sample, the low colonisation plants were primarily heterozygous. This could indicate a dominant or semi-dominant mutation (which would not be out of the question given the phenotype distribution observed in **Figure 5.11**), but it was uncertain how much effect on gene expression an insertion in the terminator region would have. We considered this only worth following up if the WGS data failed to produce any better candidates.

	Number of tnt1 present at insertion locus			
	A	B	C	D
High Colonisation	1	2	0	2
	1	2	0	1
	1	2	*NA*	2
	0	0	2	1
Low Colonisation	*NA*	2	*NA*	1
	1	*NA*	0	1
	2	*NA*	0	0
	1	2	0	0

Figure 5.12 – NF3438 high/low colonisation screen (part 1)
Results of the high-low colonisation F_2 genotyping for insertion loci from the invPCR dataset. Describes the number of transposons at the diploid insertion locus, so '0' implies a wild type genotype lacking any copies tnt1 at the insertion locus, '1' a heterozygote, and '2' a plant homozygous for the tnt1 insertion. NA indicates a failed PCR reaction.

After completing the genotyping shown in **Figure 5.12**, we received and processed the Illumina sequencing data for NF3438. The high-low colonisation panel was genotyped for the most likely candidate insertions from **Section 4.5.5** (**Figure 5.13**). Insertions 16 and 27 could be ruled out as they had high colonisation plants with two copies of the tnt1 insertion. All high colonisation individuals had either one or zero copies of the tnt1 insertion at i13 and i34, but these are not compelling candidates as neither has many homozygous individuals in the low colonisation category. Insertion 25 presented an obvious candidate, with both successfully genotyped low colonisation plants homozygous for the tnt1 insertion, and none of the high colonisation plants having more than one copy.

Therefore, the entire F_2 population was genotyped for i25, and those plants with two copies of the insertion were significantly less colonised than the heterozygotes or those without any tnt1 insertions at the i25 locus (**Figure 5.14**; p=0.0049 and 0.0073 respectively; Dunn test). While there were non-homozygous individuals which had less than less than 5% of their root length containing arbuscules, this was likely symptomatic of overall highly variable nature of the colonisation in this screen (as was seen in the wild type control plants; **Figure 5.11**). Critically, no plant homozygous for the insertion had more than 4.6% of its root length occupied by arbuscules, clearly demonstrating co-segregation between the genotype of i25 and the low AM colonisation phenotype of the NF3438 line. We conducted a BLAST search with the

sequence of Mt7g027190, and discovered that the GRAS TF it coded for was RAM1, part of the transcription factor complex at the end of the CSP that regulates pro-AMS gene expression (se **Section 1.5.3**). RAM1 was identified as a gene essential to the AMS in 2012[72].

	Number of tnt1 present at Insertion locus				
	i13	i16	i25	i27	i34
High Colonisation	**0**	**0**	**0**	**0**	**1**
	1	**2**	**1**	**1**	**1**
	0	**0**	**1**	**2**	**0**
	1	**0**	**0**	**2**	**1**
Low Colonisation	**2**	**0**	*NA*	**1**	*NA*
	NA	**0**	*NA*	*NA*	*NA*
	1	**2**	**2**	**0**	**0**
	1	**0**	**2**	**2**	**1**

Figure 5.13 – NF3438 high/low colonisation screen (part 2)
Results of the high-low colonisation F_2 genotyping for insertion loci from the WGS dataset. Describes the number of transposons at the diploid insertion locus, so '0' implies a wild type genotype lacking any copies tnt1 at the insertion locus, '1' a heterozygote, and '2' a plant homozygous for the tnt1 insertion. NA indicates a failed PCR reaction.

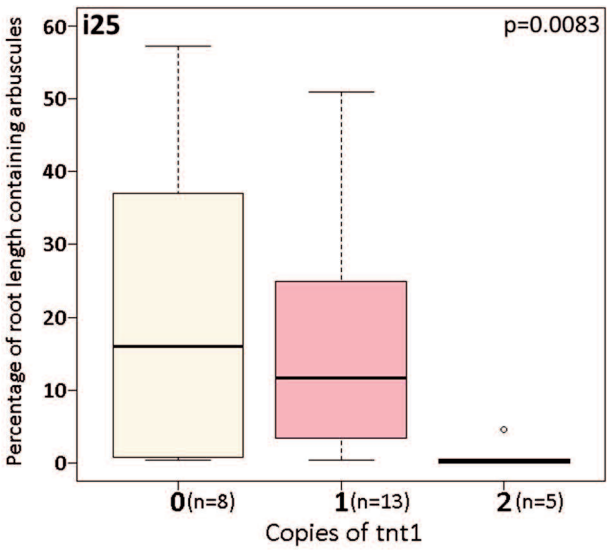


Figure 5.14 – Co-segregation analysis for NF3438
Correlation of number of tnt1 transposons at the i25 locus with the mycorrhizal colonisation phenotype in the NF3438-7-2 F_2 plants. P values derived from Kruskal-Wallis test. Number of samples per genotype given on the x axis (n=#).

5.6 – Discussion

5.6.1 – Patterns of colonisation in the F_2

A priori, we had expected the phenotype of F_2 populations to fall into two distinct groupings, phenocopying the wild type and mutant F_0. In NF3209, we might expect to see some overlap (as seen in the parental phenotype; **Section 3.3.1**), but for NF443 and NF3438, there had been no overlap in percentage colonisation between the mutant and wild-type controls. However, this phenomenon was only seen in the 7 wpi samples from the February 2016 NF443 screen (**Figure 5.3**), with all other screens showing a smooth trend of colonisation. This is the phenotype segregation pattern we would have expected to see from a group of multiple loci that additively contribute to the phenotype. There are a large number of tnt1 insertions in each of our lines (21-37 candidate insertions), making it likely that one or more of these lines have multiple insertions that would disrupt genes important for the mycorrhizal symbiosis. RNA sequencing estimates suggest some 3-6% of the legume genome is differentially regulated during the AMS[100,311]. There are genes involved in the symbiosis that are not regulated on the RNA level, but the 3-6% estimate also includes many genes downregulated during the AMS which would not have been isolated by our screen, as they would lack reduced colonisation phenotypes in the knock-out. Additionally, many of these differentially regulated genes will also not be required, but merely symptomatic of the AMS. Thus, there are no firm numbers to estimate the total genome involvement in the AMS, but assuming that 3% of genes would reduce AM colonisation if disrupted, then it is near certain that even NF1436 (the line with the least insertions) would have insertions within the 2 kb region of interest of at least two of these genes[vi], and NF443 is likely to have an insertion into the exons of at least two of these genes. The 3% figure is likely an overestimation, with many of those genes being redundant or unimportant enough that they would produce no phenotype if

[vi] This probability was calculated by modelling of the question as a Birthday Problem in Wolfram Alpha[416]. We use a theoretical genome with 33 genes, one of which would produce a mycorrhizal phenotype (i.e. 3% of the genome). We assume that the chance of insertion into each gene is equal, and thus find the probability of multiple insertions into our 'single mycorrhizal gene'. Taking the number of insertions into the region of interest in each of our lines (between 2 kb upstream and 500 bp downstream; amounting to approximately 2/3 of the total number of insertions in each line; see **Table 4.7**), we find a probability of 99.98% for NF1436, and a 61.3% chance of three mycorrhizal genes with nearby insertions. Looking only an insertions into exons (those most likely to prevent production of functional protein), we see a 99.36% chance of affecting at least two mycorrhizal genes for NF443, and a 49.48% chance for NF1436.

knocked out. Indeed, the co-segregation data does suggest that the phenotypes of both NF443 and NF3438 are single locus traits (**Figure 5.5** and **5.14**), despite F_2 populations for both lines showing this smooth curve of phenotypes (**Figure 5.3** (5 wpi) and **Figure 5.11**).

Another explanation for this phenomena may be the high variability of the mycorrhizal colonisation phenotype. This has been a constant complication throughout the study, seen in the parental lines (see **Chapter 3**) and one which we were unable to reduce with alterations to experimental design. It may be that this biological variability is the inevitable consequence of performing a species-species interaction experiment, as both species produce compounding variability. We see this well illustrated in the NF443 February 2016 5 wpi screen (**Figure 5.3**), where the mutant F_0 control plants show a far greater variation than they do in the 7 wpi screen, with only ~5% root length colonisation separating the most colonised mutant and least colonised wild type control, compared to a nearly 40% difference in the 7 wpi controls. The wild-type control plants in the 5 and 7 wpi samples are similarly colonised with a slight increase in the amount of arbuscules (59±16% at 5 wpi to 68±11% by 7 wpi). Interestingly, the mutant controls show a non-significant reduction in colonisation from 5 wpi (40±19% of root length containing AMF) to 7 wpi (25±2%). While we lack sufficient sample size to rule out chance, this may indicate that the NF443 phenotype is a colonisation maintenance mutant. This is the opposite of the delayed colonisation phenotype that NF3209 displays. Colonisation is initially established at a close to wild type rate, and then falls, perhaps through host rejection leading to starvation of the AMF or even the initiation of a defence response.

5.6.2 – A pair of insertions correlate with the phenotype in NF443

The line NF443 has a phenotype of reduced AM colonisation, with a greater reduction in the amount of arbuscules than in the amount of internal hyphae, and with arbuscule development arresting in the birds foot stage (**Section 3.3**). We found that a locus containing two tnt1 insertions into exons of a pair of genes in close proximity (83 kb) in the telomeric end of chromosome arm 7q. The homozygous genotype at these insertions significantly segregated with the phenotype in the 5 and 7 wpi F_2 populations (**Figure 5.4** and **5.5**). The heterozygous plants have phenotypes not

significantly different from those without any tnt1 insertions, suggesting the causal mutation is recessive. Additionally, the nearby insertion 39 in 7q also showed an association with the phenotype at a lower significance value, a pattern we would expect to see from linkage drag.

Given the tight linkage between the NF443 insertions 41 and 42, it is not feasible to distinguish them with further co-segregation screens, as this would require in excess of 4000 F_2 plants to obtain 5 segregants. We would also need to prove the segregating gene was causing the phenotype, as correlation is not causation.

A priori, the assumption must be that it is i42 (disrupting a predicted tyrosine kinase Mt7g116650) that is the likely candidate. Mt7g116650 has a strong mycorrhizal expression profile (around a 30x peak above background expression in arbuscule containing cells; **Figure 4.15**), whereas the GDSL acetylcholinesterase (7g116510) disrupted by i41 is expressed primarily in the stem, with little expression in the root. A kinase is also easier to place as a signalling protein, but it is possible that a GDSL might be involved in the synthesis of messengers[vii]. At least one GDSL gene (*Enod8*) has been shown to be important in the nodule symbiosis, but it is expressed in the nodules themselves[346]. While the kinase is the obvious candidate, we must test both genes directly for their mycorrhizal phenotype to be certain. There are several different ways to approach this, and we decided to complement the genes of interest in composite plant culture (see **Chapter 6**).

5.6.3 – No loci significantly co-segregate in NF3209

Work with the NF3209 line was plagued by problems of inconsistent phenotypes. This appears to have been caused by the changes to the mycorrhizal inocula as well as the changing strength of the phenotype over the growth period of the plant. The data supports the characterisation of the NF3209 phenotype as a delay in colonisation. The WGS approach had given us a large number of candidate insertions (see **Figure 4.7**), so we had plenty of targets to genotype in the F_2 screens. Now expecting a delayed colonisation phenotype, the F_2 screen of June 2016 was harvested at the right time to

[vii] The stem location means any involvement in the AMS the GDSL may have must be trans-acting, rather than directly facilitating fungal growth, as seen with fungal CAZymes or carboxypeptidase SCP1.

obtain substantial differences in arbuscule count between the wild-type and mutant control plants (**Figure 5.9**). This difference (30.6±6.3% vs 5.9±3.9% of root length containing arbuscules) was significant, and should have been strong enough to distinguish a co-segregating locus. After initial attempts at high-low screening of the likely candidates (**Section 4.5.4**) proved fruitless, the entire F_2 population was genotyped. Eventually, all insertions that were within our pre-defined region likely to affect transcription (within 2 kb upstream and 500 bp downstream of a gene's transcript) were genotyped for this population, and found not to co-segregate with the phenotype. While it is not impossible that an insertion further away from a gene than this could be having a sufficient effect on gene regulation to cause the delayed colonisation phenotype, it is unlikely. Other mutations due to the callus culture process are thus the likely cause of the NF3209 phenotype. This could include somaclonal mutations (i.e. SNPs, small INDELS or epimutations[342,347]) known to occur during the callus culture[272]. Another possibility is the activity of another copia-like retrotransposon, MERE1, native to *Medicago* and also activated in callus culture[348]. Using the same method of transposon localisation (as described in **Chapter 4**), we observed tree MERE1 insertions in NF3209 not present in R108. These insertions were into the second exon of Mt5g095430, an unknown protein with a chromosome segregation ATPase domain, into the third exon of Mt1g031450, a wall associated receptor kinase with an extracellular galacturonan-binding domain, and into an intergenic section of DNA found only as a R108 scaffold (see **Section 4.4.5**). Neither affected gene is expressed during the AMS[313].

We would expect a non-causal insertion to exhibit linkage to a nearby causal insertion (as is seen in NF443 insertion 39). None of the NF3209 insertions tested show a segregation pattern consistent with this. Thus, we assume that the causal mutation(s) are located in the parts of the genome not near previously tested tnt1 insertions. **Figure 5.15** highlights all genomic regions outside of a rough window of effect of 15 Mb around a causal insertion[viii]. Both MERE1 insertion loci are near to tnt1 insertion loci that were not significantly associated with the phenotypes[ix], so can probably be

[viii] Based on the significant association between NF443 i39 and i41/42, approximately 12.7 Mb or 16.8 cM apart.

[ix] Mt5g095430 is 2 Mb away from NF3209 i27, and Mt1g031450 is 3 Mb from Seq27 (see **Figure 5.15**)

ruled out as candidate mutations. Thus, we would first target future efforts on finding somaclonal mutations in chromosome 6, 2q and 4q. The tnt1 insertions within these chromosomal regions that were previously excluded as candidates, due to their distance from predicted genes, should also be genotyped. It is possible the annotation of the genome is incomplete. Genes encoding none-protein products such as miRNA can have significant effects on the symbiosis[10,349], and are more likely to be missed by prediction software. Like a SNP, we would expect to have observed a linkage signature if a causal tnt1 insertion was not the untested regions (highlighted in **Figure 5.15**).

To locate a somaclonal mutation we would use two approaches. We can likely rule out significant chromosomal instability[331] as a cause of the mutation in NF3209, as we see no evidence of this in our genomic data. Thus, we would first look for a hypothetical causal SNP or INDEL by aligning our genomic data for NF3209 with the R108 v0.95 genome assembly[297] with SNP calling software[350]. SNPs found within the untested regions (see **Figure 5.15**) would be tested for association with the mycorrhizal phenotype in an F_2 population using a variation of the triplex PCR genotyping technique used in this chapter, focusing on non- synonymous changes to known genes. If this did not locate a causal mutation, we would then look for epigenetic mutations. This would require the lines to be pyrosequenced. We have made initial attempts to call SNPs in N3209 with the bwa[304], samtools[305] and bcftools[351] packages to estimate the number of SNP between NF3209 and R108. We found that while the majority of the NF3209 genome lacked any SNPs (relative to R108), there were rare patches of high density SNPs (up to 20 per kb). This relative segmentation of apparent somaclonal mutations should make it easier to locate the causal genomic region, as the highly polymorphic regions are small enough that they would not be expected to segregate during production of an F2 population. Thus, each region could be treated as a single mutation for the co-segregation screen.

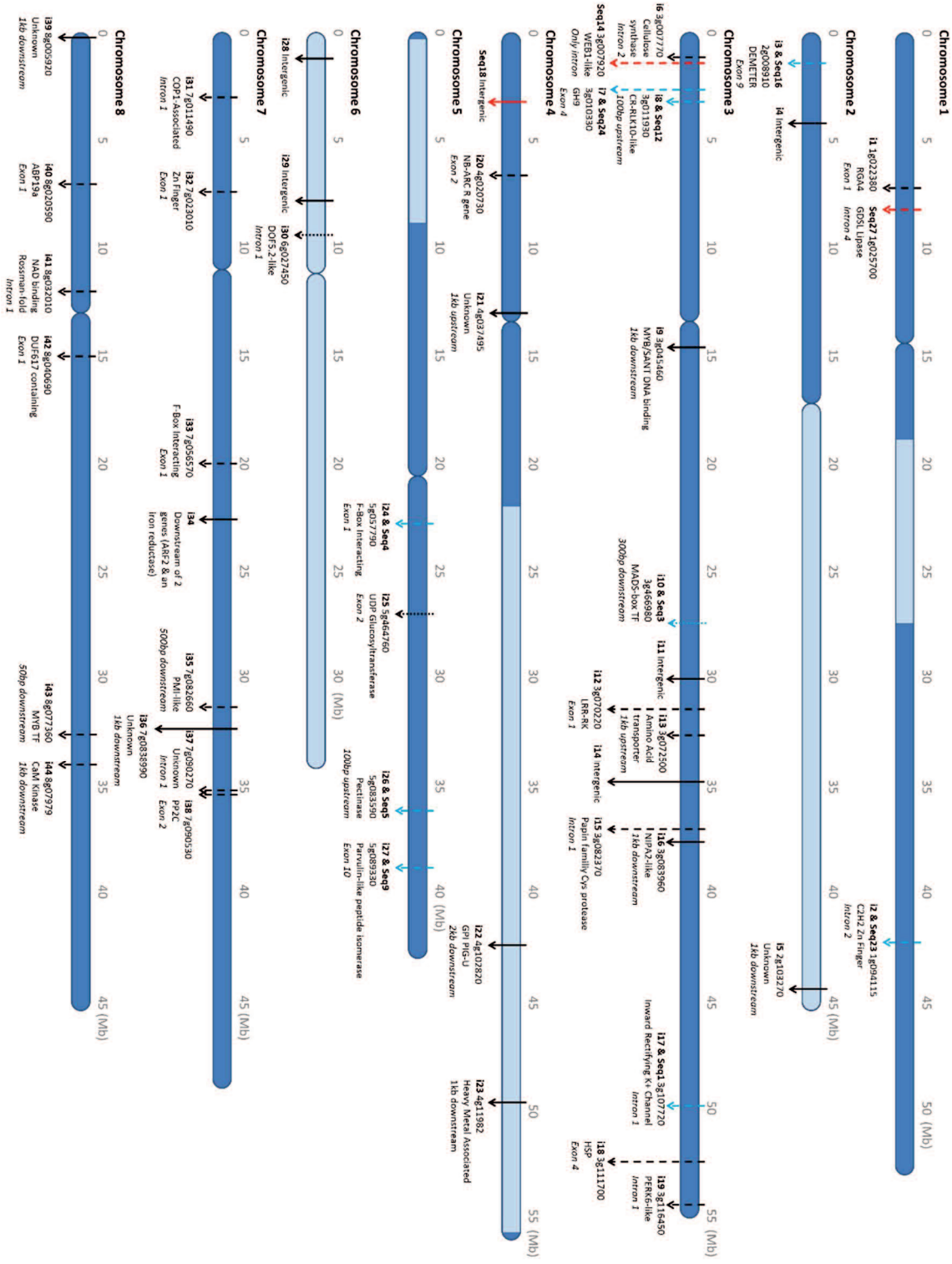


Figure 5.15 – NF3209 chromosome map, showing locations of insertions and future regions of interest
Map of the chromosomes of the NF3209 *Medicago truncatula* line showing the location of all insertions located by Illumina WGS (black arrows), invPCR (red arrows) or both methods (blue arrows) along with the predicted function of the nearest gene. Dashed arrows indicate an insertion that is not significantly associated with the phenotype (see **Appendix 3**). Solid arrows denote insertions that has not been tested, and dotted arrows these where primer sets failed to amplify one or both expected amplicons. Regions at least 15 Mb from a non-significantly associated insertion are highlighted in pale blue.

Another idea worthy of consideration is that the parental NF3209 line has two (or more) mutations that additively contribute to the reduced colonisation phenotype. That at least one of our lines would show this is highly likely (see **Section 5.6.1**). Still, if the parental phenotype is driven by mutations at multiple loci, these loci are unlikely to be linked to the tnt1 insertion we screened (**Figure 5.15**). Few of these tnt1 insertion loci show a trend towards a phenotype pattern that would be associated with a mild but causal mutation (see **Appendix 3**; **Figure 9.8**). It is unlikely that the phenotype requires two mutations to be present (e.g. both mutations disrupted genes with redundant functionality), as this would produce a 1:15 ratio of mutant to wild type phenotypes in the F_2, which is not supported by our observations (**Figure 5.9**).

5.6.4 – A novel allele of a core MYC gene

NF3438 has a clear phenotype and a tnt1 insertion (i25) that co-segregates with the phenotype (**Figure 5.14**). This insertion disrupts the first exon of the GRAS family transcription factor *RAM1* (Mt7g027190). *ram1* remains one of the strongest known mycorrhizal mutants, characterised by a near-complete abolition of colonisation[72]. Since i25 likely prevents functional expression of a known mycorrhizal gene, and correlates with the low colonisation phenotype, we feel it safe to conclude that this is the causal mutation, without demonstrating it with complementation. Thus, we will not work further on this line.

While *RAM1* is not a novel component of the AM genetic network, recapitulating this important finding is a confirmation that our new pipeline using Illumina WGS for tnt1 discovery is capable of finding mycorrhizal genes. This gene was not seen in the invPCR dataset, although it does have restriction sites used in invPCR within 1 kb of the insertion (**Figure 5.16**), which suggests that it should have been possible to find this insertion by invPCR in this case. NF3438 also represents a novel *ram1* allele, disrupted far earlier in the transcript than NF807, the tnt1 mutated line described in Gobbato et al (2012)[72].

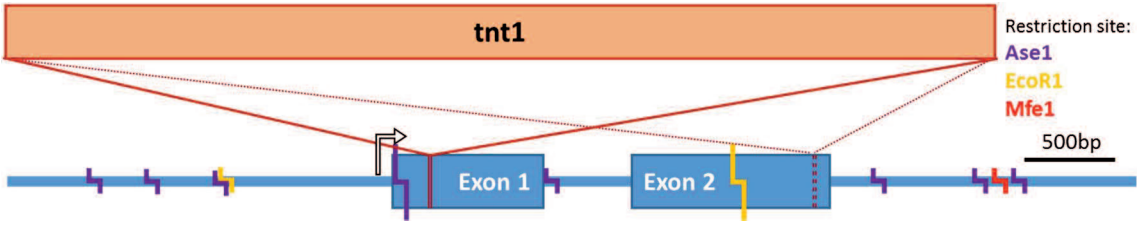


Figure 5.16 – NF3438 is a novel allele of *ram1*
To scale location of various genomic features around the gene Mt7g027190, located 9 Mb into Chr7p. Three of the four enzymes used during invPCR (**Section 4.2**) have at least one restriction site within 2 kb of the gene (ㄣ). The location of the tnt1 insertion is shown in NF3438 (solid line) and NF807 (dotted line).

Chapter 6 – A symbiotic kinase required for the AMS in *Medicago truncatula*

6.1 – Introduction

6.1.1 – Methods to prove gene/phenotype association

A co-segregation experiment can show a correlation between a genetic locus and phenotype, but to say that the phenotype is caused by disruption of a specific gene required for the AM symbiosis requires further evidence. There are a number of ways to do this, that fall into two categories; complementation and independent generation. For the line NF443, where a pair of linked transposons segregate with the phenotype, these techniques will double as a way to show which of the two genes disrupted (by i41 and i42) is causing the phenotype.

For complementation, a wild-type copy of the gene believed to be causal is transformed back into the mutant plant, where it should (if the gene is causal) replace the function of the damaged gene and rescue the mutant phenotype. Independent generation of a new mutant that recapitulates the phenotype can be done in a number of ways, either obtaining another randomly mutagenised line that disrupts the same gene, targeted disruption of the gene with a technique like TILLING or CRISPR-based mutagenesis, or by silencing transcription and/or blocking translation of the protein product by RNA interference[334,352]. The easiest method is obtaining an independent random mutant, but this requires extensive mutant collections. In *Medicago truncatula*, we have access to a number of these collections, including the Noble Foundation's tnt1 transposon insertion population and a number of smaller T-DNA insertion[353] and TILLING[354] populations. Obtaining lines from these collections is reasonably inexpensive, although secondary screening to obtain lines homozygous for the mutation of interest may be required depending on the collection. After this, RNAi is the most established technique, but normally just reduces the expression of a gene. This can be problematic with some genes, where low levels of protein can maintain normal or somewhat reduced function. This means that many independent lines must be generated and tested for the expression of the gene of interest, and can risk wrongful dismissal of relevant genes if silencing is ineffective. Additionally, if the

artificial siRNA generated shared homology to other genes (as may occur when attempting to target a gene from a large, conserved family), then the expression of off-target genes may also be affected, presenting spurious phenotypes.

Targeted gene knock outs can be performed with a variety of methods, with the recently published CRIPSR being the preferred option (having displaced TALEN and meganuclease mediated mutagenesis, as it has roughly the same outcomes with less investment and expertise required). The CRISPR/Cas9 system uses a bacterial anti-viral system to induce double stranded DNA breaks at regions with homology to a provided RNA template[355]. *In planta*, this causes the non-homologous end joining (NHEJ) repair pathway to fix the break, leading to small INDELs forming, with a targeting accuracy of a couple of base pairs. It is somewhat constrained by the need to ensure enough Cas9 protein and guide RNA is produced, and by the specific PAM sequence required near break points. Producing two break-points within 50-300 bp can lead to the creation of a large deletion, reducing the risk of an INDEL not causing a total loss of function in the target gene (e.g. by causing a non-frameshift mutation).

Complementation of the mutant is the third option. Here, a wild-type copy of the gene (including the native promoter) is introduced to the mutant plant via T-DNA integration. This should ensure that normal production of the gene is resumed, so if the gene was causal, the mutant phenotype will be rescued (revert to wild type). The two issues with this methodology are that genomic context can change gene expression (although rarely to a large extent, so the majority of transformants would be expected to produce wild-type levels of the protein), and the need to incorporate all cis-regulatory elements of the gene. The latter issue is more significant, but in most cases including the introns and 2-3 kb upstream of the gene should be sufficient.

6.1.2 – Obtaining independent mutants

For *Medicago*, the best current resource is the Noble Foundation's tnt1 collection. Each line in the collection contains a large number of insertions (see Tadege et al (2008)[272], and **Chapter 4**), but this database does not include the location of the transposon insertions in most of the lines. However, even within this smaller pool, we found at least one other line that had a tnt1 transposon in the region. As well as new

mutants of Mt7g116510 and Mt7g116650 (the two potential causative genes in NF443), we also ordered another line with tnt1 disrupting 7g446190 (which was also disrupted in NF443, by i36). This line was added because of its strong mycorrhizal expression, and because we had not been able to get several genotyping primer sets to work for the gene, so did not know its segregation pattern. It was still on the same chromosome arm, but unlikely to be causing linkage drag on the significantly associated insertions (i39, i41 and i42), given its estimated distance map distance of 46 cM, and the fact that i38, located between the two sets, does not appear to be linked in the high-low colonisation assay (**Section 5.2.2**). Nevertheless, its strong expression and protein function made it an interesting target. Therefore, three lines were ordered:

- Mt7g116510 – six NF lines could be found that had transposons in the coding sequence of this gene, including two that had been screened in Tadege et al (2008)[272] and not found to have a MYC phenotype. NF19713 (predicted to have 30 insertions) was chosen.
- Mt7g116650 – two NF lines were found, and NF15212 (predicted to have 9 insertions) was chosen, as it had fewer insertions than the alternative line.
- Mt7g446190 – only a single NF line was found, which had already been screened negatively by Tadege et al (2008)[272]. This line, NF2901 (predicted to have 7 insertions) was ordered.

These three lines were screened for AM colonisation and genotyped. However, all of them (**Section 6.3.1**) lacked any evidence of tnt1 transposons in the gene of interest.

6.1.3 – Production of complementation lines

Given the time involved in obtaining and screening additional Noble Foundation tnt1 lines, we decided to produce complementation lines rather than look at further tnt1 lines. We took what we considered the most straightforward approach, to complement the two genes disrupted by NF443 insertions 41 and 42 (Mt7g116510 and Mt7g116650). To do this, we produced T-DNA constructs using the GoldenGate cloning system that contained the R108 genomic sequence, with 2000 bp upstream and 500 bp downstream of the predicted gene transcript. While we could not control where the construct inserted, and the effect this genomic location would have on transcription, the window around the gene of interest would hopefully provide as

close to native expression levels as possible. We then used *Agrobacterium rhizogenes* to transform seedling radicals, which would grow into composite plants, with only the roots transformed[271]. While this has the potential to miss hypothetical parts of the mycorrhizal signalling pathway that require the whole plant (e.g. modification of hormone transport), most mycorrhizal signalling takes place exclusively within the roots. The benefit of this approach was its simplicity and speed, relative to producing fully transformed plants, which would have required callus culturing. Callus culture had the potential to reactivate tnt1, disrupting new genes which could have altered the results.

The AM colonisation of the two sets of transformed NF443 composite plants can be assessed, relative to a R108 control transformed with same constructs. If a gene was involved in the mycorrhizal symbiosis, the intact copy should restore function and cause the phenotype to revert back to the wild type.

Additionally, we produced marker and marker-overexpression lines with a carboxyl-terminal yellow fluorescent protein (C_t-YFP) fusion. The native promoter and terminator regions were replaced with those from the *Agrobacterium tumefaciens* mannopine synthase gene in the overexpressor construct. The YFP tag allows us to assess the subcellular localisation of the protein, as the distribution can give information about the function of the protein in the symbiosis. Overexpression would likely give more intense fluorescence, but could potentially lead to protein accumulation in other locations. Overexpression could also lead to an AM over-colonisation phenotype, which would also give clues as to the function of the gene in the symbiosis.

6.1.4 – Two pronged approach: independent mutation by CRISPR/Cas9

At the same time as the complementation, we had a 3rd year undergraduate project student take a complementary approach, producing independent gene knock outs for our genes of interest in the R108 wild type using the CRISPR/Cas9 system[356]. We used a pair of guide RNAs in each construct, targeted several hundred base pairs apart in the genes of interest. This potential for simultaneous cuts leads to the deletion of the sequence between these two points in some cells. This increases the chance of

obtaining a complete loss of gene function compared to the short INDELs that characterise the NHEJ scar produced by a single cut site[355].

6.2 – Methods

6.2.1 – New Noble Foundation Lines

We grew all available seeds from the three lines in sand:Terragreen with 25% leek stock pot inoculum (*Rhizophagus irregularis*) for 6 weeks. After this, the plants were washed to remove the substrate, a leaf sample was taken to extract DNA and half the roots taken for analysis. The rest of the plant was replanted in 9x9 cm pots of *Medicago* Mix compost, and allowed to self and set seed.

Roots were stained with ink/acetic acid and analysed in the standard way, and leaf DNA extracted with the CTAB/chloroform method and genotyped with triplex PCR.

6.2.2 – GoldenGate Cloning

The T-DNA cassettes for transformation were produced using the GoldenGate cloning system, which uses the programmable overhangs produced by the type 2S restriction enzymes Bpi1 and Bsa1 to enable the use of a standardised library of parts and a one pot digestion ligation reaction[335]. The final constructs contained two standard selectable markers under strong constitutive promoters. The first allowed for chemical selection, using *E. coli* nptII to provide resistance to the antibiotics kanamycin and geneticin. This was under the control of an *A. tumefaciens* opine synthesis promoter, so would not be expressed in the *A. rhizogenes*, just the host plant. This meant the bacteria could be removed from the transformed roots by plating onto kanamycin containing media. The second marker expressed red fluorescent protein (DsRed), to allow visual selection of transformed roots. Between these was the gene of interest, either in its native state, or overexpressed or fused to YFP. This allowed for selection against partial T-DNA integration, by taking only those roots that had both selectable systems present. The six final constructs, and the different GoldenGate components used in their construction, are listed in **Appendix 1**.

The gDNA sequence for the genes of interest was amplified from R108 leaf DNA extracted per manufacturer's instructions with a Qiagen DNeasy Plant kit. Amplicons had to be produced in sections to remove Bpi1 or Bsa1 cut sites during adaptor ligation, as the presence of these sites would cause unwanted digestion of fragments during digestion-ligation reactions. A preliminary round of pre-amplification with Q5 high-fidelity DNA polymerase (New England Biolabs, MA, USA) was carried out to bulk up the region of interest.

0.25 µl of Q5 polymerase and 0.5 µl of R108 leaf DNA extract were added to a solution of the Q5 buffer, 200 µM of each dNTP and 0.5 µM of each primer, made up to a total volume of 25 µl with diH_2O. PCR with long extension times was used, due to the length of the amplicons desired (allowing 2 min per kb, as per manufacturer's instructions). After an initial melting step of 98°C for 5 min, 35 cycles of amplification were performed, with a 20 second, 98°C melting step, a 30 second annealing step and a 6 minute 72°C extension step, with the final extension step extended to 10 minutes. The annealing temperature depended on the primers used, as per the calculator provided by the manufacturer[357]. See **Appendix 2** for primers used, and **Appendix 1** for the DNA fragments and plasmids produced.

Adaptors were added to the fragments by PCR to allow assembly into the GoldenGate universal acceptor plasmid, and to remove the Bpi1 and Bsa1 binding sites. This adaptor ligation reaction used the same reaction mix as the pre-amplification reaction, using that reaction mix as the DNA source. A similar PCR cycle was then performed, with the first 5 cycles using a low annealing temperature (47-52°C depending on the specific primer, based on the calculated T_m of the sections of the primer that would initially match the DNA template before any addition/editing occurred), then finished with 25 cycles with a higher annealing temperature (63-72°C depending on the specific primer, based on the calculated T_m for the whole length of the primer).

The resulting PCR amplicons were purified using a High Pure PCR Product Purification Kit (Roche) or MinElute PCR Purification kit (Qiagen), following manufacturer's instructions. The digestion-ligation reaction was carried out according to the long protocol presented by Patron[358]). The plasmid product was then bulked up in *E. coli*

and combined with various standard GoldenGate components from the MoClo Kit[359] and MoClo Plant Parts Kit[360] in subsequent steps using the same digestion ligation protocol, with Bpi1 replaced with Bsa1 where appropriate.

6.2.3 – Bacterial transformation

E. coli DH5α cells were transformed by a $CaCl_2$/heat shock method adapted from Sambrook & Russell's Molecular Cloning 3rd Edition[361]. *A. rhizogenes* cells were transformed by a heat shock method adapted from Hofgen and Willmitzer et al, 1988[362,363]. See **Appendix 4; Table 9.18** for concentrations of antibiotics used in the LB plates.

Competent ARqua1 cells were produced by streaking out cells from a glycerol stock on LB plates containing streptomycin. After 2 days of growth at 28°C, a colony was picked and grown up overnight in 2 ml of LB broth containing streptomycin. This culture was then added to 200 ml LB and shaken at 28°C until it reached an OD_{600} of 0.5. It was then transferred to four ice-cold 50 ml polypropylene tubes and centrifuged at 2700 g for 10 minutes. The supernatant was discarded, and the tube inverted for 1 minute to drain. The pellet was washed in ice-cold 0.15 M NaCl solution, the re-pelleted and resuspended in 3 ml ice-cold 20 mM $CaCl_2$ solution. This mixture was then aliquoted and flash-frozen in liquid nitrogen.

Plasmid DNA was obtained for further cloning or *A. rhizogenes* transformation by transferring 10 µl of the LB broth used above (or a LB broth prepared in the same manner from 16 hour old colonies obtained by streaking out a glycerol culture) to 5 ml of LB containing the appropriate antibiotic, and gently shaking at 37°C for 16 hours. Plasmid DNA was extracted using a NucleoSpin Plasmid kit (Macherey-Nagel), according to manufacturer's instructions and following all optional purifications steps, eluting into 50 µl of the appropriate elution buffer. DNA concentration was quantified with a NanoDrop 1000 or NanoDrop 8000 spectrophotometer.

The T-DNA plasmids were then transformed into ARqua1 cells. 200 µl of cells were thawed and 1 µg (in no more than 5 µl) of plasmid DNA was added to 200 µl of cells in an ice-cold 15 ml polypropylene tube. They were incubated on ice for 30 minutes,

then in liquid nitrogen for 1 minute, followed by a 37°C waterbath for 3 minutes. 1 ml of LB broth was then added, and the cells shaken at 28°C for 2 hours, before 100 µl was spread over a LB plate containing spectinomycin and streptomycin. Colonies could then be grown up in liquid LB culture to produce glycerol stocks for storage or the bacterial lawns for transformation.

6.2.4 – Production and assessment of transgenic composite plants

Composite plants of *Medicago truncatula* were produced using methods adapted from the Noble Foundation's *Medicago* Handbook[271].

Seeds were sterilised (as described in **Section 2.2.1**) and germinated on 1% water agar plates for 24-48 hours in the growth room (see **Appendix 4; Table 9.19**). Simultaneously, a lawn of *Agrobacterium rhizogenes* expressing the construct of interest was produced by applying 500 µl of a LB broth culture (previously grown for 2 days at 28°C) to a TY plate containing the appropriate selective antibiotics and growing at 28°C for 24-48 hours. Once the seedlings had >1 cm of radicle, they were collected and used for the transformation. In a lamina flow hood, seedlings were placed on a microscope slide, in a small amount of sterile diH_2O, and the radicle tip (~3mm) was removed with a scalpel. This cut section was then coated in bacteria by gently scarping it across the *A. rhizogenes* lawn. The seedling was transferred to F media plates containing 25 µg/ml kanamycin, which were sealed with parafilm, slit along the sides to allow transpiration, and returned to the growth room.

6.2.5 – Confirmation of transformation

When transformed plants were 2-3 weeks old, the transformation efficiency was assessed by use of a florescence stereoscope (Zeiss Stemi 508) looking for the emission of the 35S::DsRed visual control used in the constructs (**Figure 6.1**). Around 50% of the roots emerging from the callus showed fluorescence. Non-transformed plants were discarded, and non-fluorescent roots were removed with a scalpel.

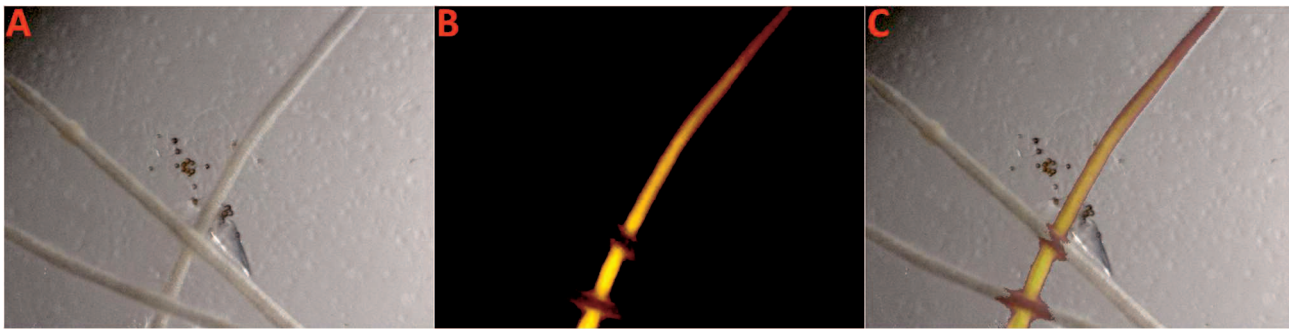


Figure 6.1 – Split plant roots expressing DsRed
Roots of a split plant (NF443 transformed with construct LMP1_65), visualising light field (A), DsRed fluorescence (B) and the merged image (C). Two non-expressing roots are visible, running left to bottom, and a DsRed expressing root runs top to bottom. Some *R. irregularis* spores are visible in the centre of the image. Images taken with an AxioCam MRC5 at 12x magnification.

6.2.6 – Inoculation and colonisation

Transformed composite plants were transferred to F media plates without antibiotics. They were inoculated with *Rhizophagus irregularis*, either as a liquid spore suspension of 100 spores per plate obtained as per **Section 2.3.2**. Colonisation was assessed at 4 and 6 wpi by the ink/acetic acid staining method (see **Section 2.5.1**).

6.2.7 – Quantification of gene expression

qPCR was used to examine expression of RAM1, Mt7g116510 and Mt7g116650 in NF443, NF3438 (*ram1*) and the R108 wild-type control, using the same protocol as **Section 3.2.2**.

For RNA sequencing, plants were grown for 6 weeks in greenhouses conditions in sand:Terragreen with either no inoculum or with a 1/3 volume from a *Rhizophagus irregularis* stock pot with leek nurse plants. RNA was extracted from 200 mg of flash frozen root tissue using RNeasy Plant Kit (Qiagen), per manufacturer's instructions, with RLT buffer and a Retsch MM300 ball mill. DNase 1 treatment and clean-up was performed, and the samples given to York Biology Technology Facility for QC, library prep and Illumina Hi-Seq sequencing. R108, NF443 and NF3438 lines were used, with 4 biological repeats (each from a single plant) per line for the mycorrhizal treatment, and 3 repeats for the non-mycorrhizal plants. This design will allow us to look for genes whose regulation during the symbiosis differs in the mutants relative to the wild type, and see if those genes overlap.

6.3 – Results

6.3.1 – New Noble Foundation lines

The plants obtained from the Noble Foundation were expected to be a segregating population, so all available seeds were genotyped to give the best shot at finding some with the insertion of interest. However, no tnt1 insertions were found in NF15212 or NF19713 in the genes of interest, with the PCR only yielding a single product size, that expected from the no insertion genotype (**Figure 6.2** and **6.3**). As we did not know the orientation of the tnt1 insertion, we could not predict the exact size of the PCR product expected. It was possible that it could have been present at the same size as the 'wild type' product, so a second round of PCR was performed on a randomly chosen subset of lines. Three separate reactions; one with only the primers to the *Medicago* genome to confirm the 'wild type' band, and two with one of the host primers and the tnt1 specific primer. If there was a tnt1 insertion present, one of these pairs should have yielded a PCR product, but this was not seen (**Figure 6.2** and **6.3**).

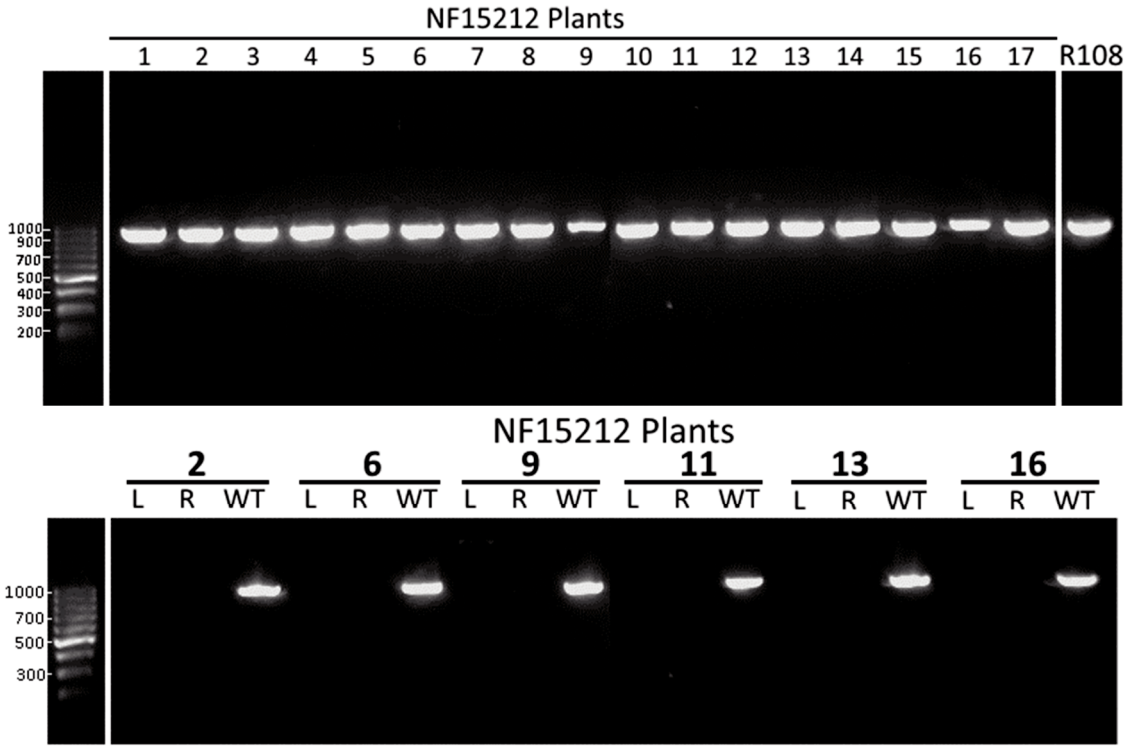


Figure 6.2 – NF15212 plants did not contain a tnt1 insertion near Mt7g116650
Agarose gels for the triplet PCR to genotype NF15212 plants and the WT control (R108). L and R indicate two-primer PCR with one of the genomic primers (left or right) and the tnt1 primer, and WT indicates PCR with just the genomic primers, which was predicted to yield a 1053 bp product.

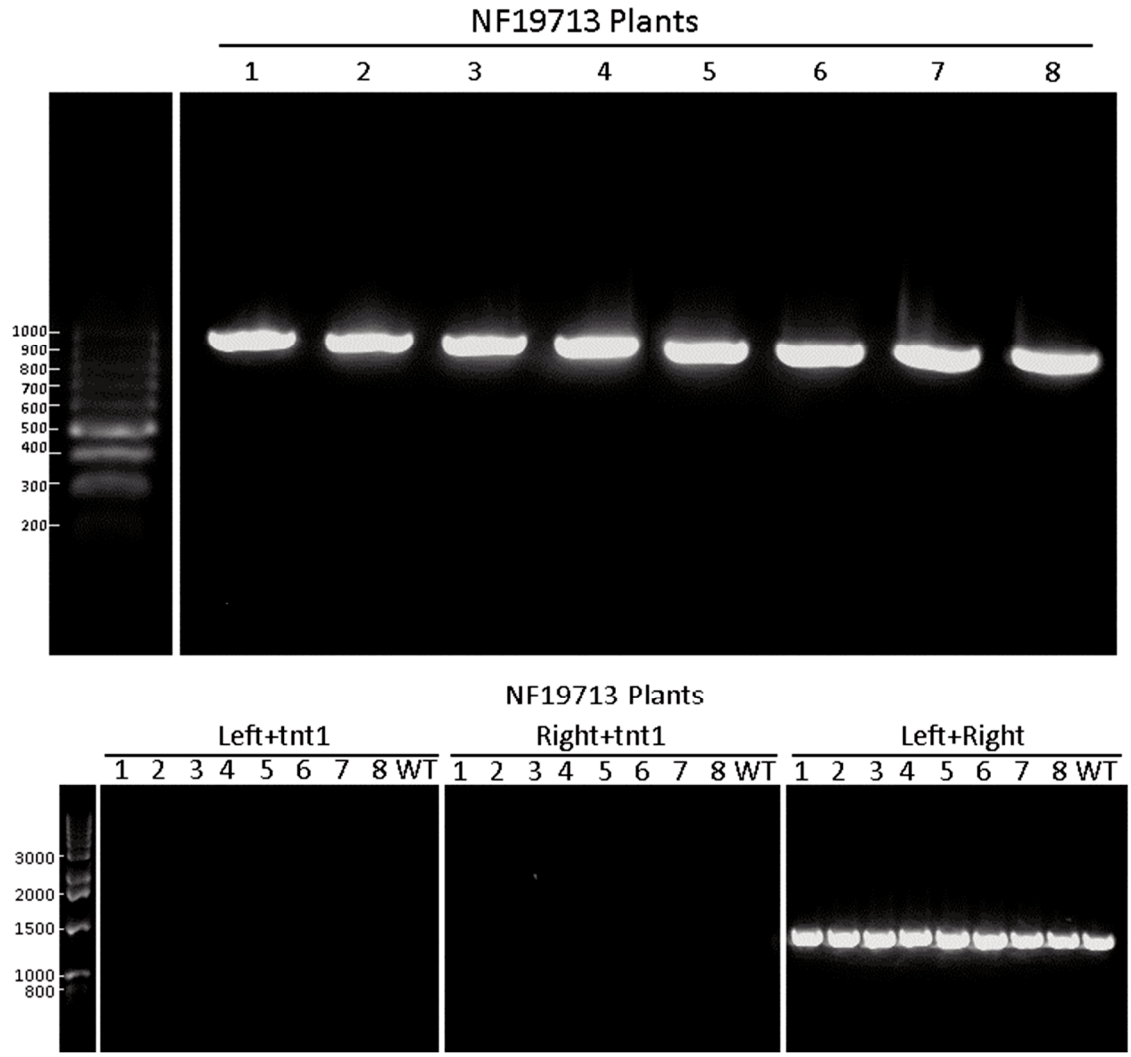


Figure 6.3 – NF19713 plants did not contain a tnt1 insertion near Mt7g116510
Agarose gels for the triplet PCR to genotype NF19713 plants. Left and Right indicate the genomic primers, and the WT plant is a R108 control. The genomic primers were predicted to yield a 1051 bp product.

For NF2901, the triplex PCR amplified bands in 11 of the 16 plants tested, but not in the wild-type control (**Figure 6.4**). These PCR products were also much larger than the predicted WT product, and subsequent paired-primer PCR showed that all the amplification was coming from the genomic right and tnt1 primers (not pictured). It is unknown why this primer pair is generating multiple products in some plants (6, 10, 11). Without a wild type product, we were unable tell if the tnt1 insertion was homozygous or heterozygous in these lines, and as none of the NF2901 plants showed a reduced in mycorrhizal colonisation phenotype, work on this line was shelved.

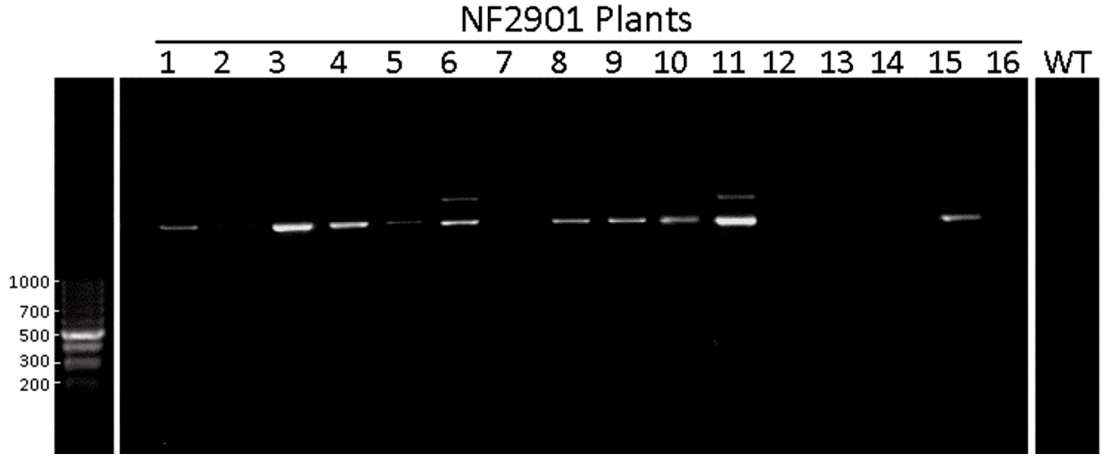


Figure 6.4 – tnt1-related genotype of NF2901 at locus Mt7g446190 is unclear
Agarose gel for the triplet PCR to genotype NF2901 plants and the WT control (R108). Left and Right indicate the genomic primers, and the WT plant is a R108 control. The genomic primers were predicted to yield a 573bp product.

6.3.2 – Generating independent knock-outs

The student produced three CRIPSR/Cas9 T-DNA knock-out constructs, one to each of the genes: Mt7g446190, Mt7g116510 and Mt7g116650. ARqua1 bacteria expressing these constructs were used to produce composite plants. Due to low rates of plant survival and problems with colonisation in composite plant culture (see **Section 6.3.3**), this work was not complete at time of writing.

6.3.3 – Complementing the mutants

Neither NF443 composite plant cultures complemented for Mt7g116510 or Mt7g116650, nor R108 composite plants transformed with the same constructs, were significantly colonised by *Rhizophagus irregularis* at 4 or 6 wpi. Due to this, we could not test the localisation of the localisation of the Mt7g116650 to the periarbuscular membrane. We would suggest that new methods of inoculation be tried for future experiments. This could be done by increasing the spore count per plate, or by adding sections cut from carrot hairy root cultures to increase the potency of the inocula. However, our primary recommendation would be to establish the composite plants into a substrate (sand or sand:Terragreen) before inoculation. This would bypass the two main problems this experiment suffered from, namely contamination of the plates from the spore solution or remaining *Agrobacterium*, and from the rapidity that the composite plants outgrew their plates.

6.3.4 – Changes in gene expression in the mutant lines

Tyrosine kinase (Mt7g116650) has a tnt1 insertion into the sixth of seven exons, and GDSL acetylcholinesterase (Mt7g116510) has a tnt1 insertion into the fourth of five exons. To test if these insertions affected gene expression, we performed qPCR with primers after the insertion point. This confirmed that tyrosine kinase expression is abolished in NF443 (**Figure 6.5**), as the only signal observed was consistent with the no template control (data not shown). However, for the GDSL acetylcholinesterase NF443 shows wild type expression (**Figure 6.6**). This suggests that the primers used were amplifying another sequence. This was not unexpected, as the GDSL family is extremely large (a family with 231 predicted members in A17 *Medicago*[364]). Care was taken to try and avoid sequence similarity with other genes, but the primers remain close enough in sequence that there could still be significant cross-amplification in qPCR conditions.

We know that both NF443 and NF3438 mutants show near abolition of *PT4* and *HA1* expression (**Figure 3.20**). We performed further qPCR to test for any change in *DMI3* and *PHO2* expression (data not presented). As expected, we observed no change in *DMI3* expression in either line (relative to wild type) as DMI3 is constitutively expressed. We also examined the expression of *PHO2* as a proxy of plant phosphate status. Expression of *PHO2*, the *Medicago* ortholog of *AtUBC24*[365], is suppressed under of phosphate starvation[366]. However, we detected no significant change in expression from wild type in either NF443 or NF3438. This may be due to the conditions, with the short (5 week) growth period limiting the time for seed P reserves to dilute out, small pots (allowing the plant to quickly colonise the majority of the potential root volume) and the aqueous P delivery allowing plenty of opportunity for P uptake by the plant intrinsic pathway, preventing a downregulation of *PHO2* even with the absence of P uptake via the mycorrhizal pathway (**Figure 3.20**) in these lines.

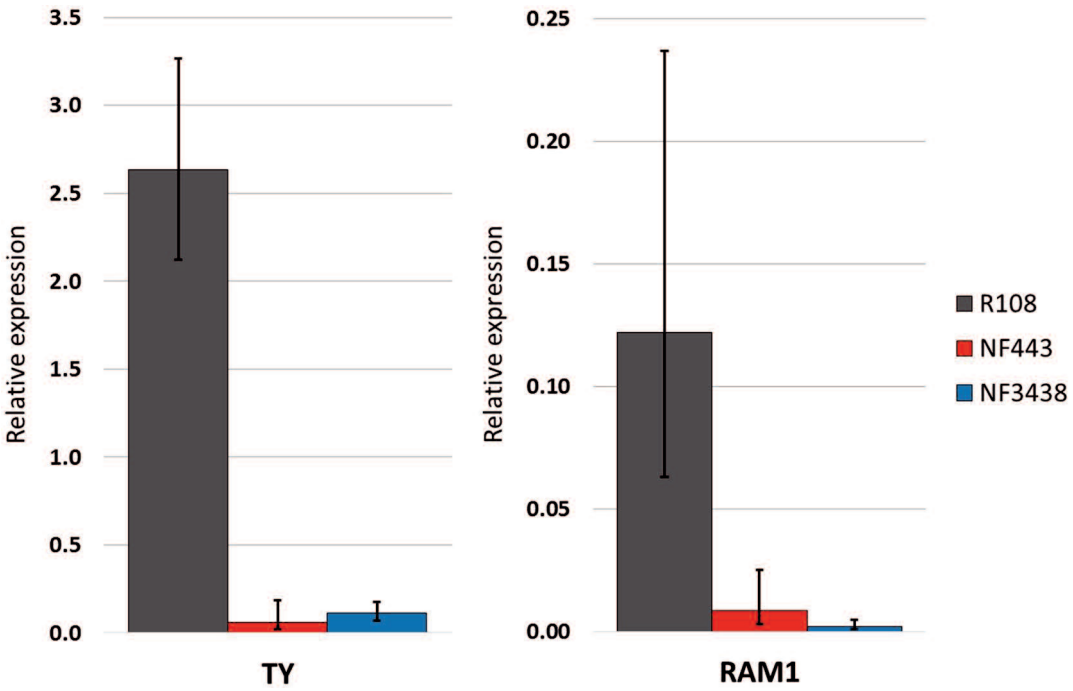


Figure 6.5 – Gene expression in NF443 and NF3438
Expression of tyrosine kinase (Mt7g116650) and *RAM1* (Mt7g027190) in the bulk roots of three *M. truncatula* lines grown with *Rhizophagus irregularis*. Three biological repeats per line, each from a single plant (except NF443, which has 2 biological repeats). The expression was normalised to the expression of housekeeping gene RNA Helicase 1[293]. Error bars show standard error.

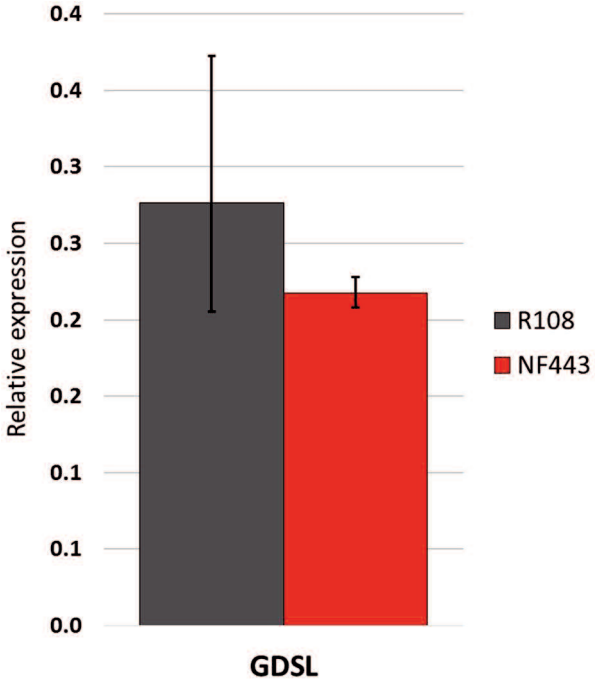


Figure 6.6 – Gene expression in NF443 and NF3438
Expression of GDSL acetylcholinesterase (Mt7g116510) in the bulk roots of three *M. truncatula* lines grown with *Rhizophagus irregularis*. Three biological repeats for R108 and two for NF443, each from a single plant. The expression was normalised to the expression of housekeeping gene RNA Helicase 1[293]. Error bars show standard error.

We then tested both mutant lines for the expression of RAM1. As would be expected, *RAM1* is essentially not expressed in NF3438, where a tnt1 insertion disrupts the first exon (**Figure 6.5**). *RAM1* expression is also strongly reduced in NF443. Interestingly, we also see the reciprocal effect, with NF3438 showing strongly reduced tyrosine kinase expression. A priori, we would have expected the two genes to either be unlinked (showing wild-type expression in the other mutant) or for one to be epistatic to the other (one line showing both genes downregulated, the other showing only one). This is clearly not the case, and we see both genes appear to depend on each other for their expression during the AMS.

6.4 – Discussion

6.4.1 – Gene expression supports the tyrosine kinase, not the GDSL, as the causal insertion

The strong expression of the tyrosine kinase (Mt7g116650) in arbuscule-containing cells made it a promising candidate from the start of the experiment (see **Section 4.5.2**). The GDSL acetylcholinesterase (Mt7g116510) disrupted by i41, on the other hand, is only expressed in the aerial tissues of the plant (**Figure 6.7**). While it is not impossible to conceive a pathway involving long distance signal transmission (e.g. as part of GA signalling, another trans-acting process that influences the AMS), and it is possible that microarray data such as this may miss some changes in expression, we consider this strong evidence that the GDSL is not significantly involved in the AMS. This is of course not proof that disruption of the tyrosine kinase is the causal mutation, but the expression data supports this.

From these data (**Figure 6.7**; further data not shown) we see that the tyrosine kinase is not expressed in the non-mycorrhizal plant. During the AMS, it is only induced in cells containing fungi. It is also switched on during the later stages of the RNS, but not during colonisation by a pathogenic fungus (although this is only one species, so it is too early to say if the tyrosine kinase could be co-opted like certain parts of the AMS machinery[85,93]). It also appears that expression of the tyrosine kinase is not induced following LCO treatment. However, since these data are based on a whole root sample

taken 24 hours after treatment, this may not tell us much, as bulk root expression is low (only twice that of NM root), and similar to the RNS, the tyrosine kinase may not be upregulated until the mature symbiosis has been established. The NF443 physical phenotype (see **Chapter 3**) shows no evidence of derivation from wild type prior to arbuscule formation, which would support this later role.

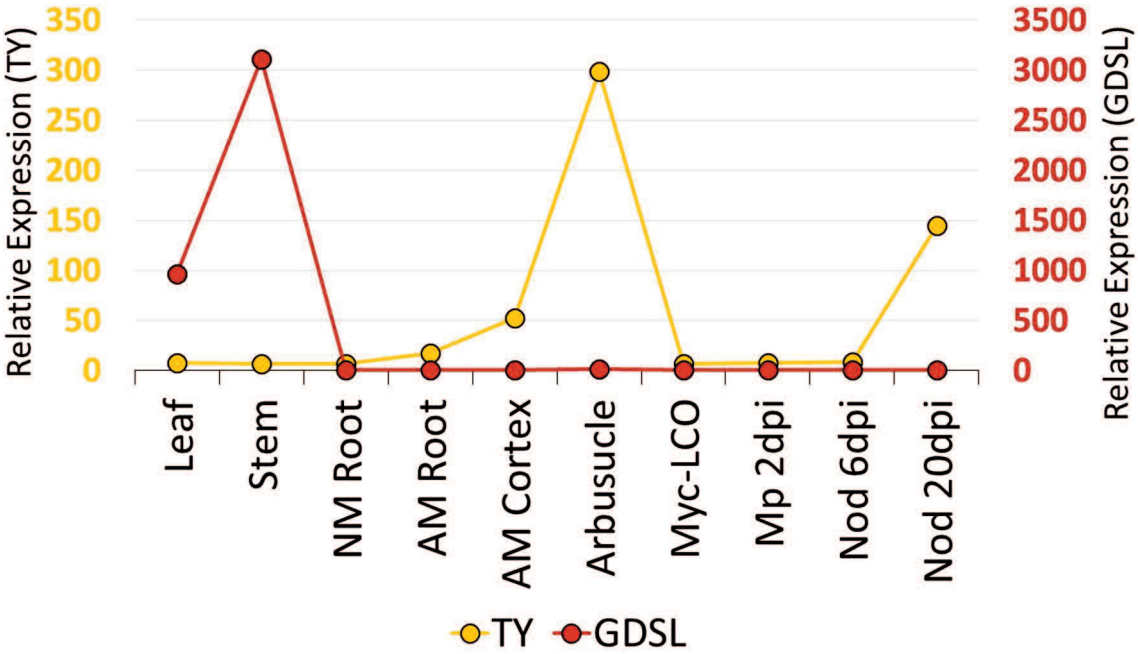


Figure 6.7 – Relative gene expression of NF443 candidates
The expression of the two genes disrupted by insertions that co-segregate with the NF443 phenotype in different conditions. Tyrosine Kinase (TY; Mt7g116650) and GDSL acetylcholinesterase (Mt7g116510). AM AM/NM Root is bulk root tissue 3 wpi, and 'arbuscule' and 'cortex' are samples obtained via laser capture microdissection of cortical cells from a 3 wpi AM root, from cells containing or not containing arbuscules, respectively. 'Myc-LCO' is wild type root 24 hours after exposure to a mix of sulphated and non-sulphated Myc-LCOs. 'Mp 2dpi' is bulk root 2 dpi with fungal pathogen *Macrophomina phaseolina*, and 'Nod' is bulk nodule tissue at 6 and 20 dpi with *Sinorhizobium meliloti*. Graph produced from publically available data collated by the Noble Foundation[313].

6.4.2 – Mt7g116650 has been previously characterised as a mycorrhizal gene

The expression of the tyrosine kinase (**Figure 6.7**) is clearly suggestive of a role in the AMS and RNS. Because of this, it had previously been investigated by colleagues in the Murray lab (John Innes Centre, UK) as part of a reverse genetic screen for kinases involved in the AMS and RNS[367]. This project obtained another Noble Foundation *Medicago* tnt1 insertion line (NF5720) with an insertion into Mt7g116650. However, they found that while this line showed a phenotype of strongly reduced AMF colonisation, and an approximately 50% reduction in nodulation. However, they found that this phenotype did not segregate with Mt7g116650 (which they called *RLCK3*).

During this project, three papers have been published, using a new phylogenetic approach to rapidly search for genes involved in the AMS[368–370]. This approach takes advantage of the recent availability of genomes for many plant species, by looking for genes conserved in plant species that form the AMS, but missing or non-functional in those species which to not interact with AMF. Between the three papers, they have identified a large number of genes that fit this pattern. While these works have a wide variance in the number and identity of genes which they report at mycorrhizal, Mt7g116650 was found in the dataset of candidate genes in Favre et al (2014)[368] and Bravo et al (2016)[370]. The gene also appeared in the initial candidates of Delaux et al (2014)[369], but was excluded from their final list of 174 high-confidence candidates. Favre et al (2014)[368] and Delaux et al (2014)[369] did not attempt to confirm any of their predictions. Bravo et al (2016)[370] did, examining the phenotype of Noble Foundation *Medicago* tnt1 lines for nine of their predicted candidates. This included Mt7g116650 (which they call *KIN3*), for which they report a phenotype (also using line NF5270) of reduced mycorrhizal colonisation and reduced PT4 expression at 5 wpi.

The *Oryza sativa* ortholog of Mt7g116650 (Os11g26140) was described as a candidate mycorrhizal gene by Gutjahr et al (2008)[371], who identified it by a transcriptomics/qPCR based approach, and show it was expressed during the later stages of mycorrhizal colonisation (from 7 wpi onward). This is later than it is first detected in legumes (we detected Mt7g116650 expression at 5 wpi (**Figure 6.5**) and Gaude et al (2012)[100] detected expression at 3 wpi). However, the described morphological phenotype of Os11g26140 (referred to as *AM14* and later *ARK1*) of an initial burst of AM colonisation, followed by a reduction in fungal root occupancy over time[372], is highly similar to that of Bravo et al (2016)[370]. However, the Paszkowski lab report a *ark1* phenotype of normal arbuscule development, and reduced vesicle/spore formation[372], not the reduced arbuscule branching we observed (**Section 3.3.2**). Bravo et al (2016)[370] did not describe arbuscule phenotype, just overall colonisation.

Finally, Aloui et al (2017)[373] did not discover Mt7g116650 among the proteins populating the cell membrane of AMF colonised *Medicago* roots in their recently published proteomic screen.

Where does this leave our work with NF443? Given the low abundance of Mt7g116650 transcript even in its upregulated state during the AMS (**Figure 6.5**), it would not be surprising for it to have escaped detection in the proteomic screen[373]. Bravo et al (2016)[370] report the phenotype of NF5270 at 3 wpi as not significantly different from wild type (with their reduced mycorrhizal phenotype appearing at 5 wpi). This may explain why Jackson (2015)[367] did not observe co-segregation of Mt7g116650 and their phenotype (which they also measured at 3 wpi). However, their initial phenotype was also established at 3 wpi, a difference for which we have no good explanation. We observed a reduced colonisation phenotype at 5 and 7 wpi in NF443 (**Figure 5.3**), similar to the *kin3* phenotype of Bravo et al (2016)[370]. Thus, we see no reason not to conclude that tnt1 insertion 42 into exon 6 of Mt7g116650, is the mutation responsible for the phenotype of NF443. As such, we will refer to Mt7g116650 as *KIN3* for the remainder of this thesis.

6.4.3 – A question of nodulation

From the work of Jackson (2015)[367], and the expression data for *KIN3* (**Figure 6.7**), we would assume a role for *KIN3* in the RNS. But, in our own initial screen (**Section 3.3.1**), lines that showed a distinctive Nod^- phenotype were removed from consideration. However, secondary screening only re-assessed the mycorrhizal phenotype, so it is possible a mild phenotype like a Nod^+ Fix^- mutation could have got through the initial screen. Phylogenetic analysis carried out by the authors (data not presented) suggests that there is no ortholog of *KIN3* in *Lupinus angustifolius* (a nodulating, non-mycorrhizal legume). This conclusion is supported by Bravo et al (2016)[370], who found *KIN3* orthologs in the most basal AM plant in their genome panel (*Selaginella moellendorffii*), but not in *L. angustifolius*. This may mean the apparent activation in late stage nodules is spurious, and may suggest an activation of *KIN3* transcription by factors other than RAM1.

6.4.4 – KIN3 is a receptor kinase

The product of *KIN3* (R108 Mt7g116650) is predicted to be a 540 aa protein comprised of two main domains, either side a single transmembrane helix (**Figure 6.8**). The C terminal region contains an ATP-binding kinase domain, although different protein prediction software suggests that this is either a tyrosine or serine/threonine kinase.

The N terminal region does not match to any known domain[374,375]. Protein prediction also suggests that the C terminus is non-cytoplasmic, and the N terminus cytoplasmic[376]. This would be an unusual configuration, as the standard configuration for membrane kinases is an extracellular domain to detect a ligand, and a cytoplasmic kinase to signal the nucleus. However, extracellular kinases have been found in animals[377] and proposed in plants[378].

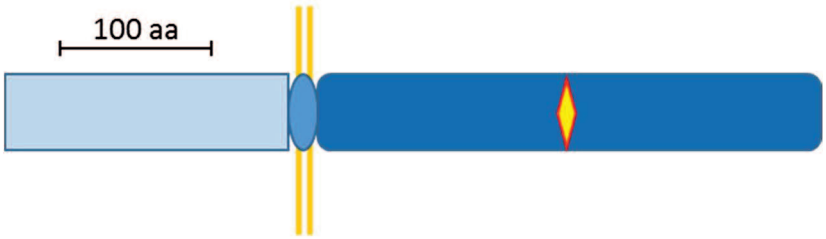


Figure 6.8 – Structure of KIN3
To scale model of KIN3, running from N to C terminus. The double line represents a plant membrane, and the yellow diamond the active site of the kinase (per InterPro[374]).

We observed extensive genetic differences between A17 and R108 (**Chapter 4**), but both these lines are colonised by AMF. Thus, we expect that if *KIN3* is important in the AMS, the mRNA should not show significant deviation between the two ecotypes. The mRNA transcript for *KIN3* is 99% identical between R108 and A17, with 10 SNPs and a 12 bp repeat (coding for Thr/Ser/Thr/Ser) present in two copies in A17 and a single copy in R108, near the C terminus of the protein product. Of the SNPs, 7 are located in the protein coding region, although 5 are synonymous. The remaining 3 SNPs are in the 3' UTR. Thus, relative to the R108 sequence, A17 has two aa substitutions (99Leu→Ser and 180Lys→Asn), and a larger 4 aa insertion occurring in a 16 aa (12 aa in R108) region at the C terminus composed almost entirely of Thr and Ser residues (10 of the 12 aa, with the remaining 2 aa being Pro). Neither substitution affects the predicted function of the TM helix or kinase[374].

Overall, the C terminal domain of KIN3 is highly conserved within the legumes (>90% aa identity), but the N terminal domain shows considerably less conservation (60-75% aa identity). This pattern is repeated in the non-legume dicots, with slightly more sequence variation, as would be expected given the evolutionary divergence between the plant families. One clear difference was observed between legume and non-legume dicot KIN3 orthologs. The aforementioned Thr/Ser rich C terminal region is

conserved in legumes, but not present in non-legume dicots. In these plants, it was replaced with a similar size domain of (10-20 aa) composed of a wider mix of aa, although most retained 2-5 Ser residues in the region. The significance of Thr/Ser rich C terminal region is unclear, but the region bears greatest similarity (by pBLAST) to fungal proteins, either as a *O*-glycosylation motifs in secreted effectors[379], or as a C terminal domain in chitinases-B family proteins[380]. Another potential function would be as a regulatory domain controlled by S/T kinase activity. In the monocot ortholog, the difference is even more striking. The C terminal domain (while lacking the Thr/Ser rich region) is conserved (~65% aa identity), but the N terminal domain is absent from the ortholog, and the sequence is not found in any other gene, indicating that the protein has not been split into multiple genes. This, combined with the lack of conservation or known structure in the legume N terminal domain suggests that it may not be important to protein function.

6.4.5 – Phenotypic differences between NF443 and NF5270

Bravo et al (2016)[370] reported that *kin3* (NF5270) displayed wild-type phenotype at 3 wpi, with fungal colonisation actually decreasing at 5 wpi, where it was significantly lower than wild type (which increased over the same time period). We observed a similar pattern in NF443, with a decrease in NF443 colonisation from 5 wpi to 7 wpi (see **Section 5.6.1**) while the colonisation of the wild type did not change. This trend appeared strongest in terms of arbuscule count, but the variation in colonisation of the 5 wpi NF443 F_0 made this hard to quantify. PT4 expression followed a similar trend, as it was 40% of wild type at 3 wpi, and essentially abolished at 5 wpi (as we seen in our results; **Section 6.3.4**). These phenotypes suggest a failure of maintenance of the symbiosis, which fits with our proposed role of KIN3 (see **Section 6.4.7**), where KIN3 maintains *RAM1* expression and mycorrhizal cell fate.

NF443 appears to have a slightly stronger phenotype than NF5270. This is not what we would expect from the location of the tnt1 insertions. The tnt1 insertion is located in the third of seven exons in NF5270, and in the sixth exon in NF443 (**Figure 6.9**). If the former still produced a protein product, it would contain only the N terminal domain, and thus most likely be non-functional (see **Section 6.4.4**). The potential protein produced in NF443 would be missing only the final ~80 aa. However, without

proteomic evidence it is hard to know if this protein would still be produced, or if the addition of tnt1 sequence would destabilise the resulting mRNA or protein, leading to it being broken down by the plant. Overall, given the clear differences in conditions between Bravo et al (2016)[370] and our own work, the observed difference in phenotype is probably not significant, and we feel it safe to conclude that NF443 does not produce functional KIN3.

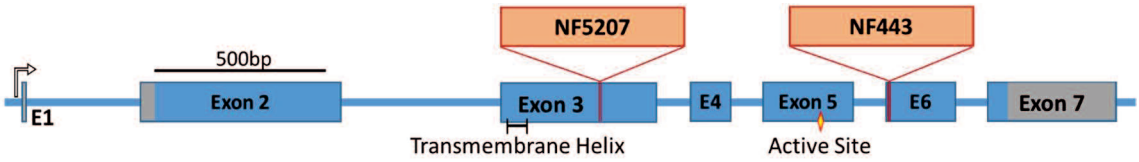


Figure 6.9 – Genomic structure of Mt7g116650 in NF443 and NF5270
To scale diagram showing the exon (boxes) and intron (line) structure of Mt7g116650. The position of the tnt1 insertions (orange boxes) in the two NF lines are marked by red lines (note – tnt1 is shown at 1/10 scale relative to the rest of the diagram). Grey regions of the exons indicate 5' and 3' UTR, and the location of bases coding for the single transmembrane helix and the active site kinase of the C terminal kinase domain are noted with the black bar and yellow diamond, respectively.

6.4.6 – RAM1 expression is significantly reduced in NF443

A key finding of this experiment is that *RAM1* and *KIN3* are both strongly down-regulated when the other is knocked out (**Figure 6.5**). Analysis of the sequence of *RAM1* in NF443 and *KIN3* in NF3438 (data not presented) showed no evidence of tnt1 insertions or SNPs/INDELs within 2 kb of the coding region. It is clear (see **Section 5.3.2**) that the causal mutation in NF443 segregates with the telomeric end of Chr7q, not Chr7p where *RAM1* is located. There are many other mutations (both retrotransposon-related and somaclonal) in these lines, so we cannot conclusively rule out trans-acting effects from these.

Rich et al (2017)[381] supports our finding of downregulation of KIN3 in *ram1*, as they found that expression of the petunia ortholog of KIN3 (Peaxi162Scf00215g00011.1) was abolished in a petunia *ram1* mutant. To the best of our knowledge, we are the first to test *RAM1* expression in *kin3.* Bravo et al (2016)[370] report decreased PT4 and LEC5[382] expression, suggesting non-functional arbuscules and supporting what we see in **Figure 3.20**.

At the time of writing, we are performing an RNA sequencing analysis (see **Section 6.2.8**) to confirm the downregulation of *RAM1* in a *kin3* mutant, and to further study

the role of KIN3 in driving AM related gene expression, and the extent to which the set of genes regulated by KIN3 diverges from that of RAM1.

With our current knowledge, we believe the weight of probability is that *RAM1* and *KIN3* are dependent on each other for their expression, rather than exhibiting an epistatic relationship. The most plausible explanation for this observation is that one of these genes is transiently activated by fungal signalling, and then drives the expression of the other gene. This latter gene then provides a positive feedback loop to maintain expression of the first gene. Downstream functions could be activated by either or both genes in concert. This interaction could be indirect, or involve phosphorylation of RAM1 by KIN3 on a periarbuscular membrane before import to the nucleus.

6.4.7 – A proposed model of KIN3 function

We propose a model of KIN3 function to explain our findings (**Figure 6.10**). Due to time constraints, we have not had the opportunity to test the predictions of the model, but offer below a number of experiments that could falsify the model.

We suggest that RAM1 is activated initially, as binding of IPD3 to the RAM1 promoter has been shown to drive its expression[64]. Thus, we hypothesize that contact with fungal MYC factors would trigger the CSP, leading to Ca^{2+} spiking in the nucleus[75] and assembly of the IPD3 complex[64]. This protein complex assembly is driven by phosphorylation of IPD3 by CaM-bound DMI3[383,384]. We suggest that over time (a few hours) the MYC-factor perception pathway will become acclimatised and the Ca^{2+} spiking response will abate. This in turn would lead to steady de-phosphorylation of IPD3 and disassembly of the protein complex, leading to a transient burst of RAM1 expression. RAM1 assembles with other GRAS transcription factors to regulate the arbuscule formation pathway[70,158]. We propose that RAM1 would induce expression of KIN3. KIN3 would accumulate on the periarbuscular membrane[385], and initiate a positive feedback loop to maintain RAM1 expression (and thus arbuscule formation and function).

How a presumably cell surface kinase would interact with a GRAS TF is not known, but it could be by phosphorylation (assuming the kinase is in the normal cytoplasmic

orientation) of RAM1 or some part of the RAM1-inducing CYCLOPS complex after translation, but before nuclear import. If located on the periarbuscular membrane, then it is possible that KIN3 may be interacting with some ligand, either derived from the fungus, or another plant protein in a membrane raft complex. The former would help to ensure close binding of the fungus, and the latter could imply some role in the regulation of nutrient exchange, by monitoring the activity of other plant proteins. However, given the absence of the N terminal domain in the monocot linage, and the otherwise low conservation and lack of known structures in the legume N terminal domain indicate that it is unlikely to be required for KIN3 function. Thus, our model assumes that KIN3 is either natively active, or is subject to activatory phosphorylation by another protein.

This process could link into the arbuscule degradation pathway described by Floss et al (2017)[73]. The NSP1/MYB1-complex would block expression of KIN3, shutting down the RAM1 complex. DELLA proteins act at all stages of the model, integrating the nutrient status of the plant to regulate arbuscule production and lifespan. This model explains the discrepancy in phenotypes between mutants, with *ram1* showing a stronger phenotype with no arbuscule formation, as RAM1 is the driver of this pathway, and without RAM1 symbiotic gene expression does not occur. The *kin3* mutant shows low levels of colonisation and early stage arbuscule formation. CSP signalling drives transient RAM1 expression, which persists long enough to allow root penetration and the initiation of arbuscule formation. However, without KIN3 to maintain RAM1 expression, the periarbuscular membrane is not populated by the proteins required for nutrient exchange (see the low levels of PT4 or HA1 expression in **Figure 3.20**). AMF fail to sporulate on NF443, suggesting a lack of carbon delivery to the fungus, supporting the observation that arbuscules lack the ability to exchange nutrients in this line.

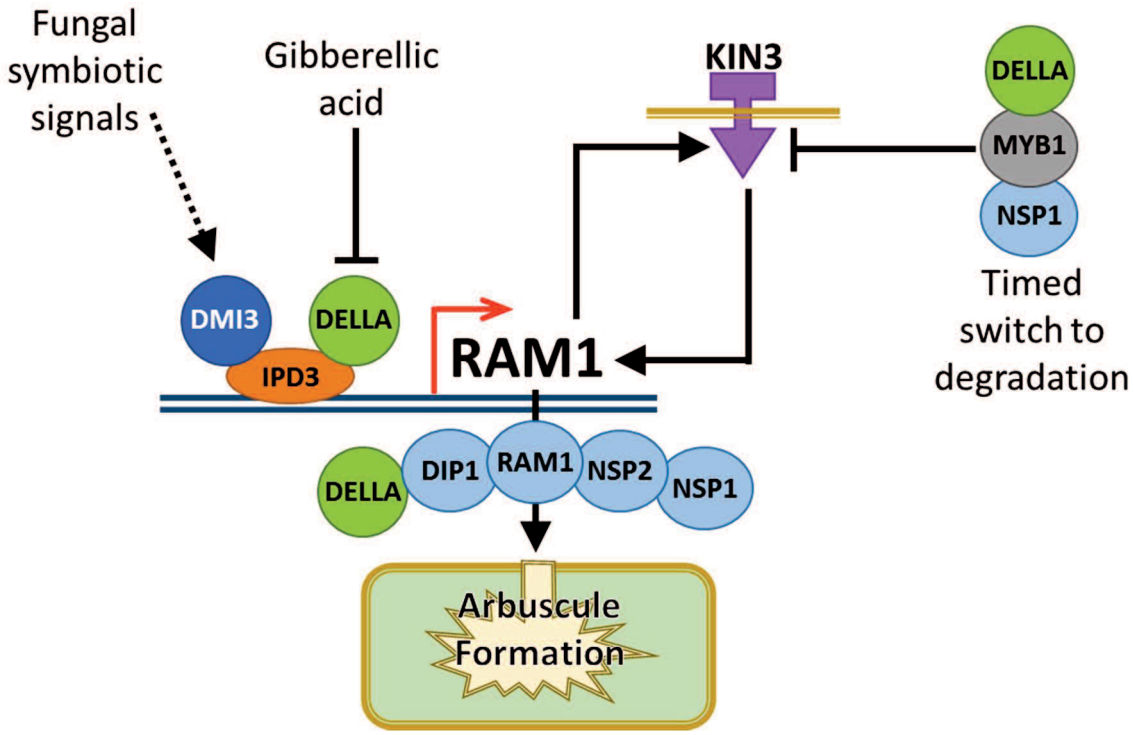


Figure 6.10 – Proposed role of KIN3 in maintain RAM1 expression during the AMS
Diagram illustrating the proposed role of the pro-symbiotic kinase KIN3 in our most parsimonious explanation for the various phenotypes observed. We propose KIN3 is induced by RAM1, then acts in a positive feedback loop to maintain RAM1 expression or function.

6.4.8 – Proposals for validation or rejection of the model

One potential problem with this model is that expression of the *O. sativa* KIN3 ortholog (*Os*AM14) is dependent on DMI1 and DMI3 (as expected from the model), but not IPD3 or the kinase activity of DMI3[371]. In legumes, RAM1 expression is dependent on phosphorylation of IPD3 by DMI3[64]. If both of these dependencies are common across the AM plant linage (as would be our default assumption), then that would argue against KIN3 being activated by RAM1. However, we already see divergence in KIN3 expression between *O. sativa* and *Medicago* (see **Section 6.4.2**), so regulation of *Mt*KIN3 and *Os*AM14 may be different. Assuming the *O. sativa* data to be true in *Medicago*, a possible explanation for the lack of KIN3 expression in *ram1* could be a spurious association (*ram1* lacking functional arbuscules, thus depriving the late expressed KIN3 of some needed pre-requisite). In this alternative model, KIN3 would be initiated by some DMI3-dependent, phosphorylation-independent branch of the CSP, but still act on RAM1 to maintain the symbiosis. The expression of RAM1 is reduced in *kin3* due to a failure to maintain the symbiosis. The alternative model would predict that *RAM1* expression should be like the wild type in *kin3* plants around the time of initial fungal contact, then quickly fall off due to a failure in symbiosis

maintenance. To examine this important possibility, we would propose two experiments. First, to test the expression of *KIN3* in *Medicago ipd3* and *dmi3* mutants, to confirm that legumes follow the same pathway as the monocots. If this confirms the *O. sativa*-like IPD3-independent expression of *KIN3* occurs in the legumes, a secondary experiment should be a time course analysis of *RAM1* expression in *kin3*, to test if there is an initial burst of *RAM1* expression that is not maintained. Another test to compare KIN3 and AM14 function would be to complement a *Medicago* kin3 mutant with a proKIN3::AM14 construct. If the phenotype was rescued, it would indicate that KIN3 and AM14 have the same role in the AMS, and that the KIN3 N terminal domain and the C terminal Thr/Ser rich region (see **Section 6.4.4**) were not necessary for KIN3 function.

A number of other experiments could be undertaken to test other predictions of the model. Since we hypothesise that KIN3 acts to maintain RAM1 expression, we would predict that constitutively expressed RAM1 would remove the requirement for KIN3. Thus, a line overexpressing RAM1 in a *kin3* background should show the same phenotype as a RAM1 overexpressor, not the expected *kin3* phenotype or some intermediate level of colonisation. This assumes that the sole function of KIN3 is maintaining RAM1 expression, as it may also activate other effects downstream of RAM1 in our model.

If KIN3 maintains RAM1 expression, then we would expect *kin3* to have the same directional effect on gene expression as *ram1* (i.e. the same set of genes should be downregulated or upregulated relative to the wild-type plant, although the fold expression changes may be smaller in *kin3* due to the initial burst of RAM1 expression predicted in the model). To test this, we are at time of writing performing a RNA sequencing experiment to test the expression of *ram1* (NF3438) and *kin3* (NF443) mutants, relative to a R108 wild-type control, in both mycorrhizal and non-mycorrhizal conditions. An alternative expression pattern we may see would be that *kin3* reduces expression of a subset of RAM1-linked genes, specifically those related to arbuscule fine branching and maintenance of nutrient exchange, but does not affect genes that are purely involved in the initial root penetration event. This could occur under the model as during these initial contact events RAM1 expression (and thus the expression

of its downstream genes) is controlled by signalling from the CSP. KIN3 only becomes important in a cell hours after initial fungal contact, so genes active before this would be relatively unaffected in *kin3*. The differences in gene expression between *kin3* and *ram1* may also help us understand any downstream effects of KIN3 separate from maintaining RAM1 expression.

The model would also suggest that ectopic expression or overexpression of KIN3 should lead to an increase in RAM1 expression. We suggest this would manifest as a priming effect, rather than cause ectopic RAM1 expression, as the function of KIN3 is most likely to be via post-translational modification of another protein, which would also need to be present to see a phenotype. The potential link into the MYB1 arbuscule-degeneration complex would be harder to examine, as we are currently not aware of the post-translational modifications that control MYB1 function[73]. From this model, we would expect KIN3 expression to decline along with RAM1 (assuming this regulation is transcriptional) during arbuscule degradation. Both of these theories would be best examined using qPCR for various marker genes from a time course along arbuscule development.

The cellular localisation of KIN3 could be examined using the YFP fusion lines we have established. While light microscopy would lack the resolution to distinguish between the sides of the protein, anti-YFP antibodies should allow immune-gold labelling under electron microscopy to detect which side of the cell membrane the YFP (and thus the C terminal kinase domain to which it is attached to) is located. Testing the N terminal domain for ligand binding function would be more challenging without any leads to the identity of a potential ligand. Thus, a pull down assay using plant root exudates and AMF GSE followed by mass spectrographic identification would be the obvious starting point (assuming the N terminus is apoplastic, as would be norm for such a protein).

Another experiment that would be difficult, but highly enlightening, both for our model and for the understanding of control of arbuscule lifespan in general, would be to examine the role of GA in arbuscule degradation. Specifically, do DELLA proteins have an inhibitory or synergistic function in the MYB1-NSP1 complex[73]. Thus, do changes in GA levels induce early arbuscule degradation, or is the lifetime of the

arbuscule set at its creation (for example, with a molecular oscillator), or controlled by local nutrient exchange? Our assumption would be that GA induces degradation of the arbuscule (i.e. inhibitory DELLA function), or that arbuscule lifespan would instead be tied to a local measure of nutrient exchange, but the opposite (GA suppresses arbuscule degradation) occurred, that would support a key role for arbuscule degradation in nutrient uptake. This would likely require live imaging of arbuscules to prove conclusively, a technical challenge to distinguish arbuscule-linked fluorescence from the root background without being able to clear the root. If this could be achieved, single growing arbuscules could be monitored to give a true estimate of lifespan (rather than the estimates from ratios of mature to degrading arbuscules). Then, the root could be flushed with Pi or GA, to test if this affects the arbuscule lifespan.

Chapter 7 – Testing the universality of the CSP

7.1 – Introduction

7.1.1 – Implementing improvements in agriculture

Mycorrhizal benefits to plants are extremely variable, depending on plant and fungal genotype, as well as soil microbiota and inorganic conditions[386]. While AMF supplementation is starting to show some success[26,201], there is unlikely to be any AMF genotype that is universally highly beneficial for all plant species. Integration of AMF into farming practice will require consistent and reasonably profitable results, something that has not currently been achieved[26]. One step towards solving this problem is the increasing ease of genomics, allowing for the profiling of fungal symbionts in field root samples, and the profiling of soil microbiota to add to abiotic data. However, this will be limited by our ability to understand these interactions. Additionally, even with falling costs, this data gathering is well outside the reach of subsistence farmers, those most vulnerable to increases in fertiliser prices and to climate change, and thus most likely to benefit from improvements in sustainable agricultural practice. If altering fungal populations is unfeasible over large temporal-spatial scales, then the plant is the obvious target for breeding or genetic modification efforts to ensure a productive symbiosis, as it would be relatively easy to spread the product of such an effort through existing plant seed distribution systems.

One problem that will need to be addressed for such a project is how best to quantify plant benefit. Mycorrhizal growth response (MGR) is a measure of the change in plant growth between mycorrhizal and non-mycorrhizal plants. While MGR is one of the most common metrics in the literature, it is clearly substandard. MGR is not directly related to yield, and is often measured over short periods. This may fail to explain the lifetime response to the AMS, as symbiont preference has been observed to change through the lifespan of the plant host[104]. This further distances experiments with a single AM genotype inoculation from real world findings. In a crop system, yield changes in response to mycorrhizal colonisation could be measured (mycorrhizal yield response (MYR)), albeit with increased monetary and time costs. MYR should also be

seen as only part of the picture, with the nutritional value of the crops (e.g. delivery of sufficient mineral nutrients or vitamins to the human or animal consumer) and the ability of AMF to mitigate pest or abiotic stress induced depression of yield, are both measures that need to be accounted for in a holistic measurement of farming benefit. For a wild ecosystem, lifetime fitness (the ability of an organism to survive to produce viable offspring)[387] should be the ideal measure, but this is equally hard to measure.

An alternative approach to application of mycorrhiza inocula to fields that may already contain substantial AMF populations, is to engineer plants to improve their symbiotic capacity with as broad a range of symbionts as possible. To do so, we need to understand the methods of communication (either by direct signalling or via effector mediated changes to the other partner) between the plant and fungus.

7.1.2 – How common is the common symbiosis pathway?

We would hypothesise that the Glomeromycotina, descended from a symbiotic common ancestor, should all utilise the same pathway to interact with the plant host. However, there is a scattering of evidence that opposes this hypothesis, enough to warrant a formal investigation.

First, Gao et al (2001)[388] found that *Diversispora versiformis* colonised and formed arbuscules on the tomato *rmc* mutant (a large deletion including *Sl*CYCLOPS, the tomato homolog of IPD3). While the colonisation was reduced to 60% of that seen on the wild type tomato, the commonly used lab strains of AMF (*Rhizophagus irregularis*, *Funneliformis mosseae* and *Gigaspora margarita*) all showed a reduction in arbuscule number on the mutant to around 5% of that seen on the wild type. IPD3 is a key part of the CSP in both the AM and nodule symbiosis in legumes, initiating RAM1 expression in the AMS[64]. Yet in tomato its function is apparently unnecessary for some fungal isolates. Since *Rhizophagus* requires *Sl*CYCLOPS to colonise the host, it is unlikely that *Sl*CYCLOPS has simply been duplicated and rendered redundant in tomato. We can hypothesise that *Diversispora* has an effector that allows it to bypass this part of the CSP, or may form the symbiosis entirely via a hypothetical CSP-adjacent pathway (see **Section 1.3**), but this is speculation, and the authors are not aware of any work exploring a genetic basis for this difference between *Rhizophagus* and *Diversispora*.

Secondly, while some mutants like *ram1* or *ram2 s*trongly reduce colonisation, others have more minor phenotypes, suggesting a level of redundancy within the symbiosis. There are multiple pathways for fungal perception (CO4/8-based, LCO-based and the unknown signal detected by D14L)[66,103]. These different pathways do not all depend on DMI1-3 for function, and drive mycorrhizal associated phenotypes (Ca^{2+} spiking, lateral root branching) to different extents[40]. Preference for decoration of fungal signalling molecules also appears to differ between plant species[65]. Yet, the same fungal species are detected and 'invited in', with little specialism, by all these plant species, even when using signals that function as PAMPs in other contexts[9]. In the RNS, LCO decoration and LysM RK co-evolution drives specificity, yet fungal exudates contain broad spectra of highly decorated LCO, CO of different chain lengths, and other yet unknown signals[66]. Therefore, we suggest that generalism in the AMS is retained by overlapping, redundant production of both signals and receptors by both partners.

There are massive changes in gene set between different isolates of *Rhizophagus*[389], so we would expect to see an even greater level of divergence at the genus level. That could provide space for the differences in effector or signalling suite between fungal genotypes (**Figure 7.1**). While this might seem to make breeding a plant for improved AM nutrition hopeless, the existence of non-AM plants tells us that some parts of the AM genetic pathway are essential, and that these are likely the most productive targets for future work.

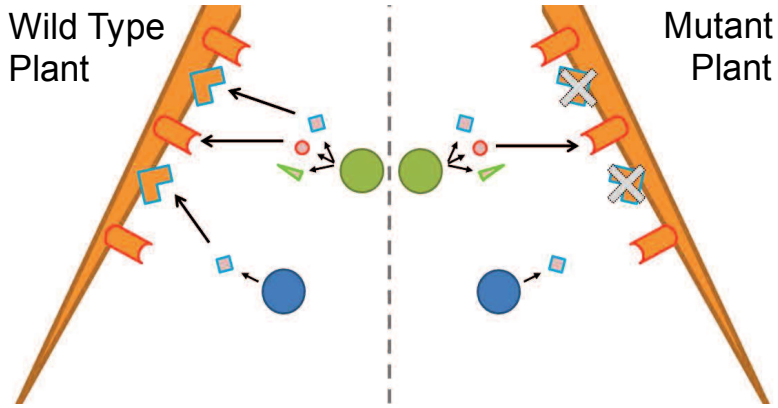


Figure 7.1 – Diverse arrays of signalling molecules and/or effectors maintain generalism
The chance loss of a plant receptor or effector target (right; either a mutant or another species) prevents the specialist fungus (blue) from forming the symbiosis, but does not affect the generalist fungi (green).

7.1.3 – A broad-scale experimental exploration of plant mutants on fungal genotypes

Lab AMF lines have been highly selected for disruption tolerance and abundant sporulation, and it can be assumed that this will have led to considerable co-selection of other traits. It is likely that lab strains of AMF, like in most other groups of organisms, only comprise a small percentage of the total genetic diversity in the whole group. We would expect that a wild population of AMF should have at least as wide a range of phenotypes as the lab lines seen in Gao et al (2001)[388].

We designed an experiment to test the findings of Gao et al (2001)[388] in our model *Medicago*, and on a much larger basis. We created a panel of *M. truncatula* mutants impaired in various stages of the symbiosis, and grew these with either *Rhizophagus irregularis* or a wild population of AMF obtained from a local loam grassland. Total colonisation keep visually assessed on stained roots, and total root DNA extracted, amplified with AM specific primers and sequenced. This would show us if certain AMF operational taxonomic unit (OTU) became more abundant on certain mutants (indicating an ability to bypass that part of the AMS signalling pathway).

Using a non-cultured AMF sample allowed us to maximise the potential diversity in the sample. We used a variety of techniques to try and capture the mycorrhizal diversity of the soil, using homogenised bulk soil in different concentrations in the first experiment, and a spore extract in the second, supplemented both times by fragments of roots. These grassland roots, showed heavy AM colonisation, and were assumed to be a good source of infectious potential for the AM symbiosis, especially for AMF species with low sporulation rates that would otherwise be filtered out of a non-soil inoculum.

We acknowledged the problem of measuring a successful symbiosis above (**Section 7.1.1**), but like all experimentation, time and money limit the scope of an initial exploratory investigation such as we undertake, and as such it is necessary to use simple growth measurements as a proxy for success in this work.

7.2 – Methods

7.2.1 – Summer 2015 Experiment

An initial test experiment was undertaken to test the colonisation of the R108 wild type *Medicago* with our wild inocula. 8 plants were grown for 6 weeks in 80 ml pots (P24 disposable trays, H. Smith Plastics Ltd, Essex, UK) in glasshouse conditions (see **Appendix 4**). The substrate was 1 part sand, 4 parts washed Terragreen and 5 parts sieved soil samples from Walmgate Stray (**Figure 7.2**; location 2), a fertile loam grassland common in York (UK). The roots from this soil sample were chopped into short sections and mixed evenly through the substrate.

For the main experiment in 2015, soil and turf samples 20x20 cm across and 25 cm deep were taken from four locations around Walmgate Stray (**Figure 7.2**). From these sample sections, green plant material was removed, and the soil sieved with a 2 cm sieve and pooled. Roots from this soil were cut into 5-20 mm sections and added back into the soil. The soil and roots were mixed either 1:1 or 1:4 with the sand:Terragreen mix. Two sizes of pot were used, 80 ml pots containing a single *Medicago* plant or 560 ml pots (cut down 2 litre/6½" pots, Richard Sankey & Son Ltd, Nottingham, UK) with seven individual plants (one of each genotype as described in **Table 7.1**) planted in a ring 2 cm from the edge of the pots, the order of planting randomised to reduce neighbour effects. Six repeats were performed per soil treatment. All pots were watered daily with diH_2O, and fertilised with 10 ml of R no P solution per plant weekly. They were grown for in the standard glasshouse conditions for 6 weeks, with the tray positions randomised weekly to mitigate block effects.

Figure 7.2 – Location of wild inoculum sampling sites (2015) on Walmgate Stray, York

Table 7.1 – Lines used in wild inoculum experiment (Summer 2015)

Line	Mutation
NF443-3	*kin3* (see **Chapter 6**)
NF807	*ram1*^{72}
NF1436-7	Unknown (*deformed arbuscules phenotype*)
NF3209-4	Unknown (*delayed colonisation phenotype*)
NF3438-1	*ram1* (see **Chapter 5**)
NF905-3-5	Wild-type control
NF905-5-5	*ha1*^{138}

M. truncatula mutants derived from the Noble Foundation's tnt1 mutant library used in the wild inoculum experiment (Summer 2015). The wild-type control and HA1 mutant line were derived by backcrossing and segregation of the NF905 parent line, which was MYC^{-}. Thus, the wild-type control will have several tnt1 insertions, although none cause a mycorrhizal phenotype.

Plants were harvested by washing roots in water to remove the soil, then dried with paper towel. Shoot and root fresh weights were measured (with plants cut at the point of where the stem turned white). The roots were sectioned into three for ink staining to measure overall colonisation, WGA-AF488 staining to observe arbuscule structure, and for measuring the dry weight. DNA was extracted from 200 µg of root tissue from three plants per treatment per genotype, using a Qiagen DNeasy kit, as per manufacturer's instructions. The shoot tissue and root section were dried at 80°C for one week to measure the dry weight, from which total root dry weight was estimated.

7.2.2 – Summer 2016 Experiment

Soil samples were collected as in **Section 7.2.1** from five locations on Walmgate Stray (see **Figure 7.3**). Spores were isolated from 1 kg of mixed soil, washed through a sieve stack (with 600, 500, 300, 150, 100 and 50 µm pore sizes). The smallest four sieves were then backwashed with autoclaved diH_2O to collect the fungal spores. A bacterial filtrate was prepared by passing autoclaved diH_2O through 100 g of mixed soil and filtering (#1 Filter Paper, Whatman, Maidstone, UK). Roots from the soil samples were cut into 5-20 mm sections and pooled. To provide the *Rhizophagus irregularis* control, a defined spore solution was produced (see **Section 2.3.2**) from *Daucus carota* hairy root culture, and sections of *Daucus carota* root were chopped into 5-20 mm sections.

Figure 7.3 – Location of wild inoculum sampling sites (2016) on Walmgate Stray, York

Pots (Synprodo 7x7x8cm pots; Synprodo BV, Wijchen, Netherlands) were filled with 350 ml of sand:Terragreen mixed with either chopped pooled roots from Walmgate Stray, or chopped *D. carota* hairy roots. Once *M. truncatula* seedlings were planted, these pots were watered with 50 ml of bacterial filtrate and 50 ml of either Walmgate Stray AM spore wash (for the pots containing Walmgate Stray roots; forming the 'wild inoculum' treatment) or the purified *R. irregularis* spores (for the pots containing *D. carota* hairy roots; forming the 'control' treatment). 9 *M. truncatula* lines (**Table 7.2**) were grown in these treatments; 8 individuals per line for the wild inoculum treatment and 6 for the control treatment. Excess R108 wild type plants in wild inoculum

treatment were also grown, and harvested weekly from 5 weeks to assess total colonisation. Plants were watered with diH_2O and fertilised weekly with 40 ml of fertiliser solution (**Table 7.3**). Plants were grown for a total of 12 weeks in a growth cabinet (Snijders Labs, Tilburg, The Netherlands; see **Appendix 4; Table 9.19** for conditions), with location in the growth cabinet randomised weekly to reduce the effect of the drop-off in insolation from the centre of the cabinet (from roughly 320 to 250 µmol photons/m^2/second). Plants were harvested after 12 weeks, with DNA extracted from root samples and the same measurements as the previous year's experiment (root/shoot fresh and dry weight, and ink and WGA-AF staining to assess colonisation).

Table 7.2 – Lines used in wild inoculum experiment (Summer 2016)

Line	Ecotype	Mutation
EP10007	A17	*dmi1*[55]
EP10013	A17	*dmi2*[95]
EP10017	A17	*dmi3*[316]
EP10058	A17	*ram2*[93]
NF443-3	R108	*kin3* (see **Chapter 6**)
NF3209-4	R108	Unknown (*delayed colonisation phenotype*)
NF3438-1	R108	*ram1* (see **Chapter 5**)
NF18637-10	R108	*bfp1*[390]
R108	R108	Wild type control

Plants used in the wild inoculum experiment (Summer 2016). Due to constraints on mutant availability, some lines used were T-DNA insertion mutants in the A17 genetic background, and some tnt1 insertion mutants in the R108 background.

Table 7.3 – Fertilisation regime for the wild inoculum experiment (Summer 2016)

Week	1	2	3	4	5	6	7	8	9	10	11	12
Fertiliser	R^{noP}	R^{noP}	R	R^{noP}	R^{noP}	R	R^{noP}	R^{noP}	R	R^{noP}	R^{noP}	R

40 ml of R solution (green) or R [no Phosphate] solution (orange) was added weekly to 350 ml of growth media

7.2.3 – PCR to identify AM fungal DNA

PCR was used to assess the total root DNA for the presence of Glomeromycotina rDNA10. 10 ng of root DNA from was amplified using 35 cycles of PCR with the AM1 & NS31 primers[22] or the AML1 & AML2 primers[391]. This used the standard Phire Hot Start polymerase mix (**Section 2.4**), with a protocol adapted for this enzyme from that of Lee et al (2008)[391], with cycles of 1 minute at each of 98°C, 58°C (for AM1/NS31) or 50°C (for AML1/AML2) and 72°C, preceded with a 5 minute initial melting step of 98°C

and followed by a 72°C 10 minute final extension step. After these methods did not give amplification sufficient for gel visualisation, a nested PCR approach was used, with 30 PCR cycles using AML1 & AML2, a 20x product dilution and then 30 PCR cycles with AM1 & NS31.

7.3 – Results

7.3.1 – Pilot wild inoculum colonisation experiment

Walmgate Stray is a fertile loam grassland long used for seasonal cattle grazing, formally since 1759, and likely since at least the 16th century[392]. This long term management has been without tilling, and with limited fertilisation by the grazing cattle, and occasional spraying with liquid manure. This lack of disruption should ensure a strong mycorrhizal population, supported by the strong colonisation seen in the pilot experiment (**Figure 7.4**). Sampling sites were chosen as areas that differed in insolation, drainage and vegetation characteristics (such as ratios of grasses to Forbes and other dicots, as well as distance to large trees). These differences mean the sites will likely have genetic differences in the mycorrhizal populations, allowing us to capture a greater AMF diversity.

In the pilot experiment, six of the eight R108 individuals survived the growth period, and had consistently high levels (55-75%) of total root colonisation, with around two thirds of this colonisation housing mature arbuscules (**Figure 7.4**). This is in line with a strong inoculum as seen with the R108 plants grown with the *Rhizophagus irregularis*/leek stock pot inoculum (see **Chapter 3**). Problematic, however, were the large differences in overall growth, and inconsistent root:shoot ratios (**Figure 7.5**). This suggested the plants were struggling to grow well in the 1:1 soil and sand:Terragreen mix, likely due to high water retention of the loamy soil, inconsistent with the free-draining conditions preferred by *Medicago truncatula*. Because of this, a 1:4 ratio mix of soil and sand:Terragreen was used for half of the plants in the summer 2015 experiment. Since the initial test plants were well colonised, it was predicted that this dilution of the inoculum would not reduce the number of infectious particles per pot enough to start causing significant differences in species abundance across the pots.

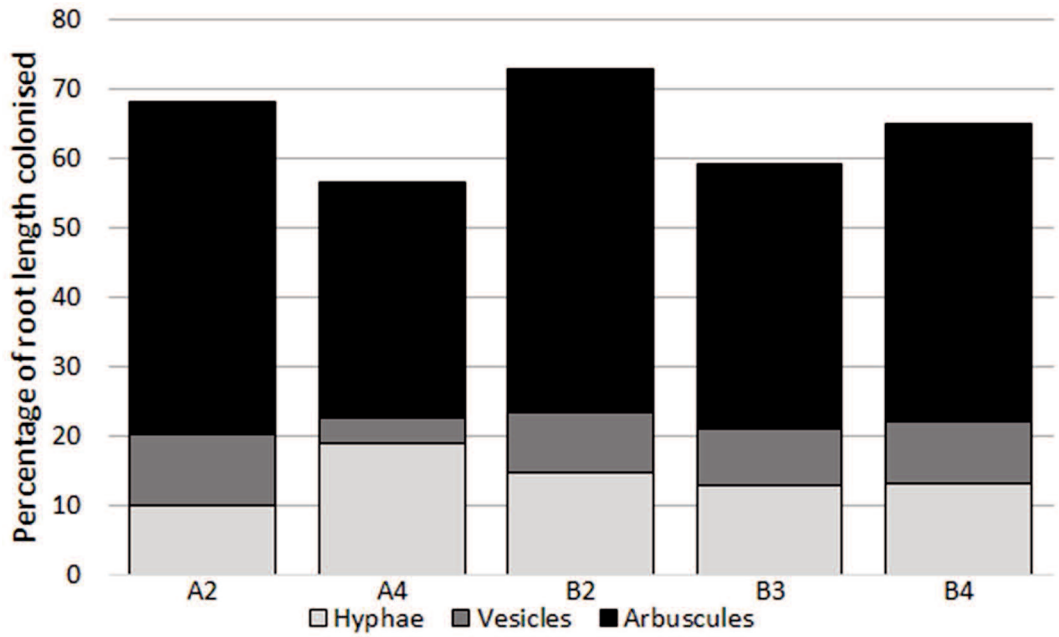


Figure 7.4 – Colonisation of WT test plants 6 wpi with wild inoculum
Levels of AM colonisation of roots of five R108 plants after 6 weeks of growth in initial pilot experiment for the Walmgate Stray soil inoculum. The letter-number codes relate to the position of the growing plants for block effecting, which had no significant effect on the results.

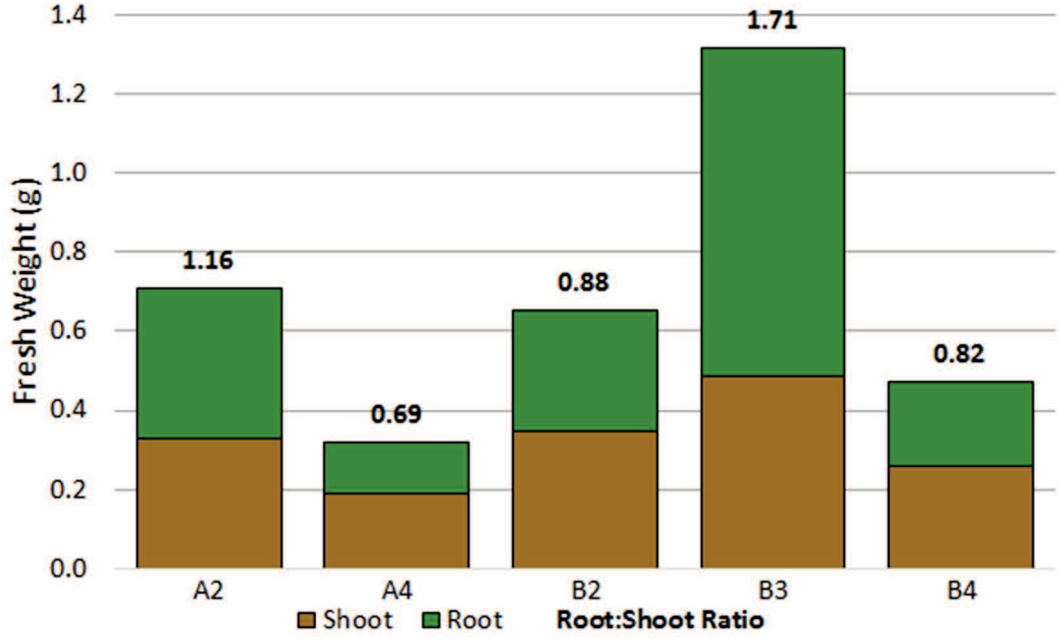


Figure 7.5 – Fresh weight of WT test plants 6 wpi with wild inoculum
Fresh weight of the R108 plants after 6 weeks of growth in initial pilot experiment for the Walmgate Stray soil inoculum, and the derived root:shoot ratio. The letter-number codes relate to the position of the growing plants for block effecting, which had no significant effect on the results.

7.3.2 – Summer 2015 Experiment

Plants were grown in the glasshouse for six weeks, during which time a major whitefly infestation occurred, which was resistant to the current control measures. There was extensive insect reproduction and sundew excretion on the plants, and heavy damage

to the leaves, with most exhibiting brown patching. This effect was not quantified on a genotype or treatment basis, but there was no clearly visible effect of either on the success of the whitefly infestation. While we lack controls to quantify the actual effect of the infestation of overall plant fitness, it is likely that plants were significantly smaller than they would have been without the infestation, as very few healthy leaves were formed during the final 2 weeks of the growth period. The effects on mycorrhizal colonisation were not as clear, but it seems likely that the infestation would have reduced the availability of sugars to feed the AMF, thus reducing their contributions to the plant. Additionally, examination of the roots showed highly variable colonisation and a large amount of non-AM fungal colonisation, likely pathogenic. We measured the plant masses during this experiment to use as a secondary measure of mycorrhizal success, as colonised *M. truncatula* should have a greater availability of P, leading to an increase in growth relative to an uncolonised plant when both are P limited[393]. Given the growing conditions in this study, we would expect that P should be the most limited of the macronutrients. Given the heavy pathogen and herbivore load observed on of the *Medicago* plants, we decided not to invest the time in counting the colonisation.

The most obvious difference between the genotype and treatment effects was the size difference between those plants in the 1:1 soil treatment and the 1:4 soil treatment (**Figure 7.6**). Plants grown in a greater proportion of Walmgate Stray soil (the 1:1 soil treatment) were significantly smaller (in terms of both fresh and dry weighty). The reduction of root mass was stronger than the effect on above ground biomass. This difference was consistent across both single and mixed pot treatment ($p=9.42x10^{-8}$ and $p=1.13x10^{-4}$ respectively, with a student's t test). Plants grown in the 1:1 soil treatment are expected to have a greater availability of mineral nutrients from the rich loam soil of Walmgate Stray, compared to the nutrient poor sand:Terragreen mixture (theoretically lacking P or N content). The decline in root:shoot ratio with higher nutrient content fits with previous descriptions of *Medicago truncatula* growth habit[394]. However, while root:shoot ratio declines with soil content, so did the overall mass and shoot mass, in contradiction to Sulieman et al (2013)[394], suggesting that changes in nutrient availability are not the cause of this observation.

We hypothesise that the cause of the decrease in mass is a necessary increase in investment of resources in roots, due to an increased rate of photosynthate loss from roots. Visual examination of stained roots showed a large variety of non-Glomeromycotina fungi on the roots of plants grown in Walmgate Stray soil, and evidence of root damage from presumably pathogenic activity. The higher soil content of the 1:1 soil treatment would have increased pathogen load on the plants, reducing the resources available for plant growth. Additionally, the 1:1 soil treatment showed a propensity towards waterlogging, which would have further stressed the plants (see **Section 7.3.1**), and lead to increased turnover of roots. The considerably lower rate of plant survival in 1:1 soil treatment (80% compared to 94% in the 1:4 soil treatment) supports the idea that increased soil content stresses the plant.

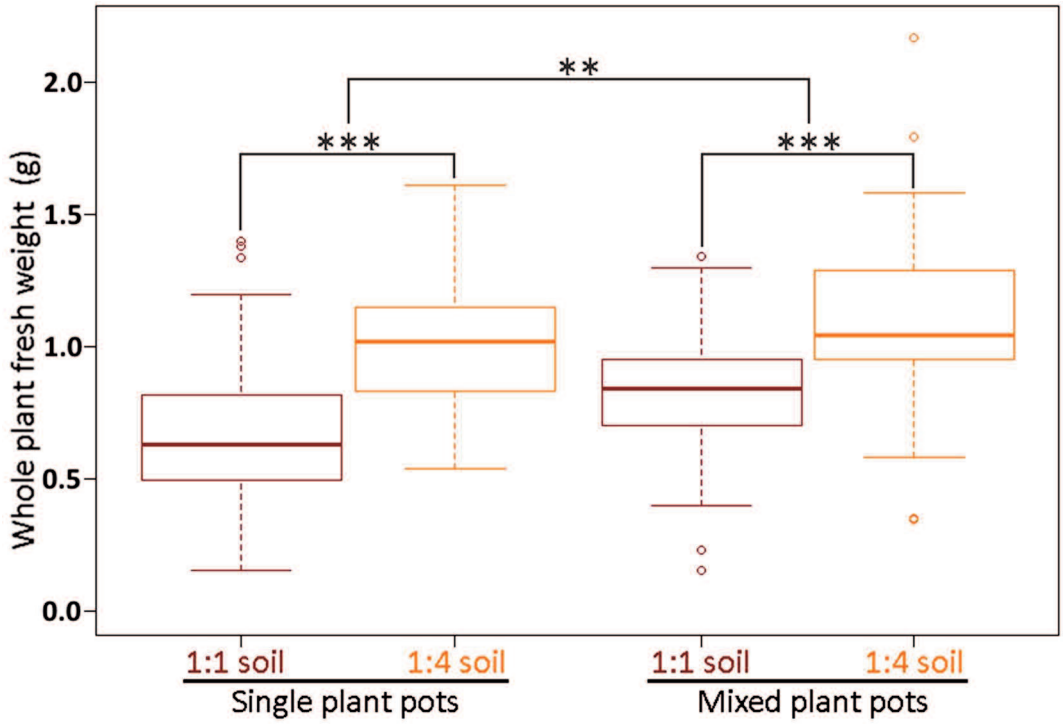


Figure 7.6 – Whole plant fresh weight by treatment; wild inoculum 2015
Whole plant fresh weight for the different soil treatments for the 70ml Single Pots and the 560ml Mixed Pots from the summer 2015 experiment. This collates 48 individual plants of all genotypes for the 1:1 soil treatment and 53 for 1:4 soil in the single pots, and 32 plants across 5 pots for the 1:1 soil treatment and 40 plants across 6 pots for the 1:4 soil in the mixed pot experiment.

Figure 7.6 also shows that plants in the 560ml mixed pots grew larger ($p=0.0088$; student's T test) than those in 80ml single plant pots. All genotypes in the mixed pots had smaller root:shoot ratios (less than 0.85). The smaller root:shoot ratio may be due

to non-self root inhibition, where roots avoid growing to near to roots of other plants, even of the same species, which can decrease overall growth, even though each plant had the same 80ml theoretical volume to exploit in both pot treatments[395]. This remains hypothetical, however, as this response varies widely between plant species (e.g. maize, soybean and onion show this response, but wild strawberry roots actively grow towards others). The considerable increase in shoot biomass in the mixed pots is harder to explain. It seems unlikely that this is a nurse plant effect, as the wild type plants increase in whole plant fresh weight in 1:4 soil treatment pots to 132% of the 1:1 soil treatment pots, which was not significantly different to the mutant lines (147±24%), with similar values for the increase in shoot dry weight (131% in the wild type and 140±21% in the mutant lines). The nurse plant effect would be expected to increase the mass of the mutant lines relative to the wild type, as it was contributing most of the carbon to the shared mycelium. The pot size may be a confounding factor, as the larger soil volume produced a more consistent environment, and may have led to smaller changes in soil moisture between watering, which would have aided shoot growth.

NF905-3-5 (wild-type control) produced the largest plants in the 1:4 soil treatment in single plant pots, both in terms of fresh (**Figure 7.7**) and dry (**Figure 7.8**) weight, although this was not seen in the 1:1 soil treatment. When the difference between soil treatments as controlled for with a mixed linear model, the wild type growth increase was not significant. Percentage dry mass was consistent between genotypes and treatments, as would be expected for plants of the same growth stage and not water deprived.

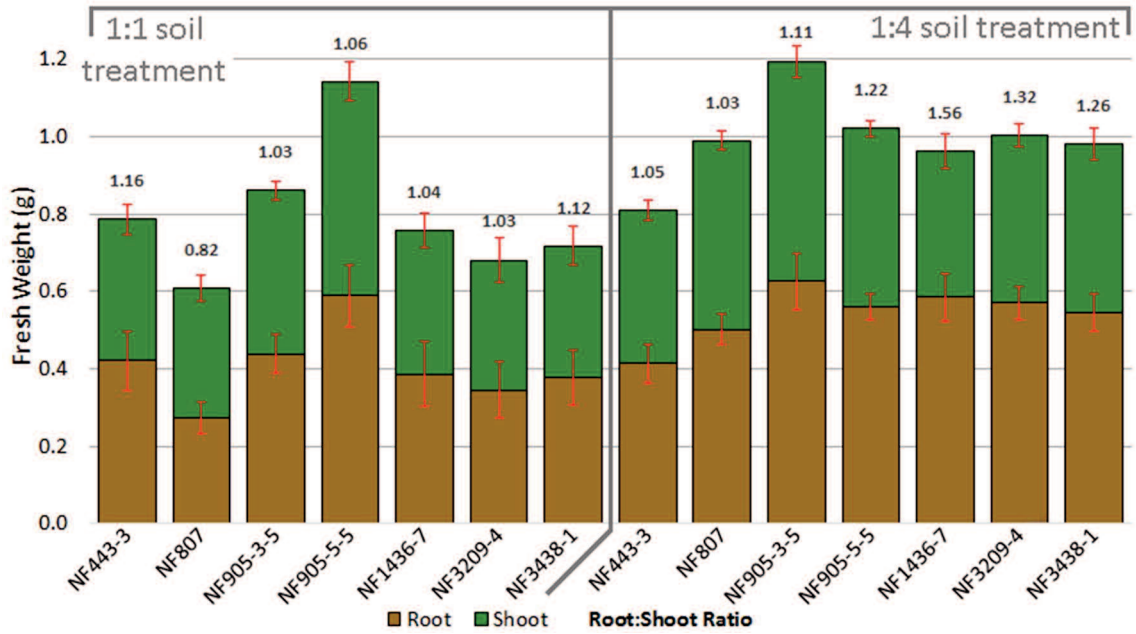


Figure 7.7 – Plant fresh weight by genotype; single pots; wild inoculum 2015
Average fresh weight of each plant genotype from the summer 2015 single pot experiment, split by soil treatment. Displaying standard error and root:shoot ratio of each genotype.

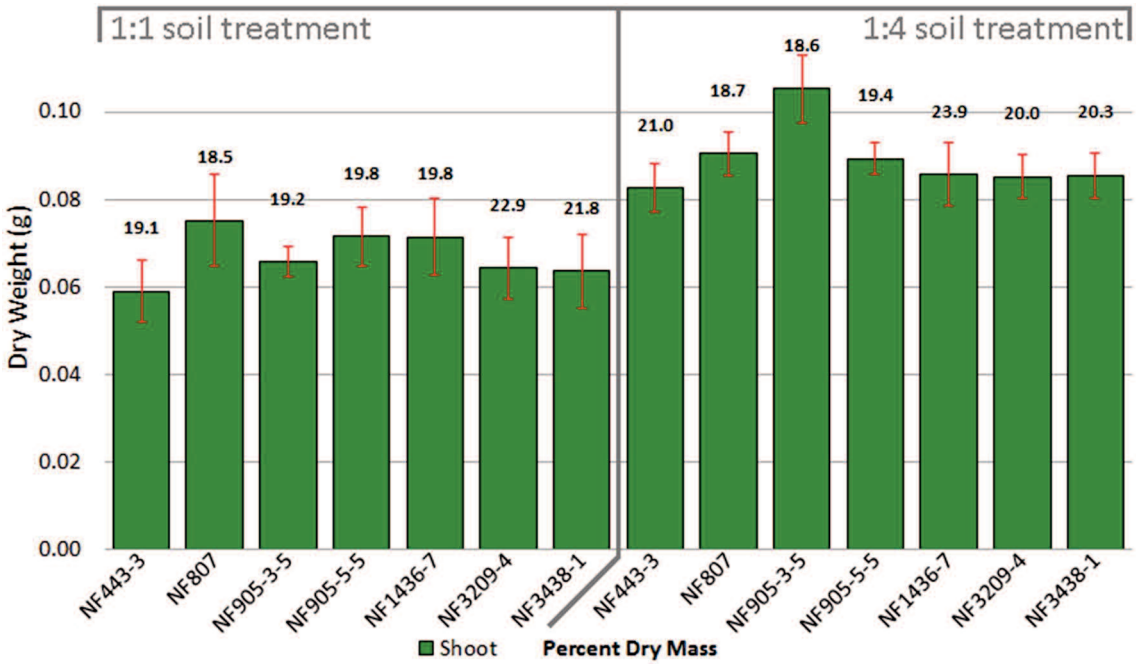


Figure 7.8 – Shoot dry weight by genotype; single pots; wild inoculum 2015
Average shoot dry weight of each plant genotype from the summer 2015 single pot experiment, split by soil treatment. Displaying standard error and the percentage of dry shoot mass remaining for each genotype.

More differences between the genotypes are apparent in the plants grown in the multi-plant pots. Within the 1:1 soil treatment, the NF1436 plants had a significantly smaller fresh weight than the wild type NF905-3-5 (mixed linear model, p=0.025; **Figure 7.9**). However this decrease was not significant in the 1:4 treatment, where NF1436 was indistinguishable from NF3209 and NF3438. Additionally, in the 1:1 soil

treatment there was a significant block effect from the pots, with plants from pot 3 significantly smaller and those of pot 5 larger than the average. This was not seen in the 1:4 soil treatment pots. All these results were consistent with fresh shoot or root weight alone.

With the different soil treatments added to the same mixed linear model, several pots were significantly smaller (2 & 3) and several larger (9 & 11) than the average (**Figure 7.10**), fitting our expectations from the treatment effect alone (**Figure 7.6**). Overall, the mutants NF1436 and NF807 were significantly smaller (p=0.013) than the wild type (NF905-3-5) and NF905-5-5, and the decrease in mass in NF3438 was nearly significant (p=0.051). NF905-5-5 was indistinguishable from NF905-3-5, and NF443 and NF3209 were intermediary between the high and low fresh weight clusters. Again, this was largely consistent with just the shoot and root fresh weight, although the NF1436 fresh root mass is not significantly reduced. The lack of difference between NF905-3-5 and NF9055-5 may suggest that much of the observed differences in growth are due to the effects of other tnt1 insertions in these lines, rather than a mycorrhizal affect.

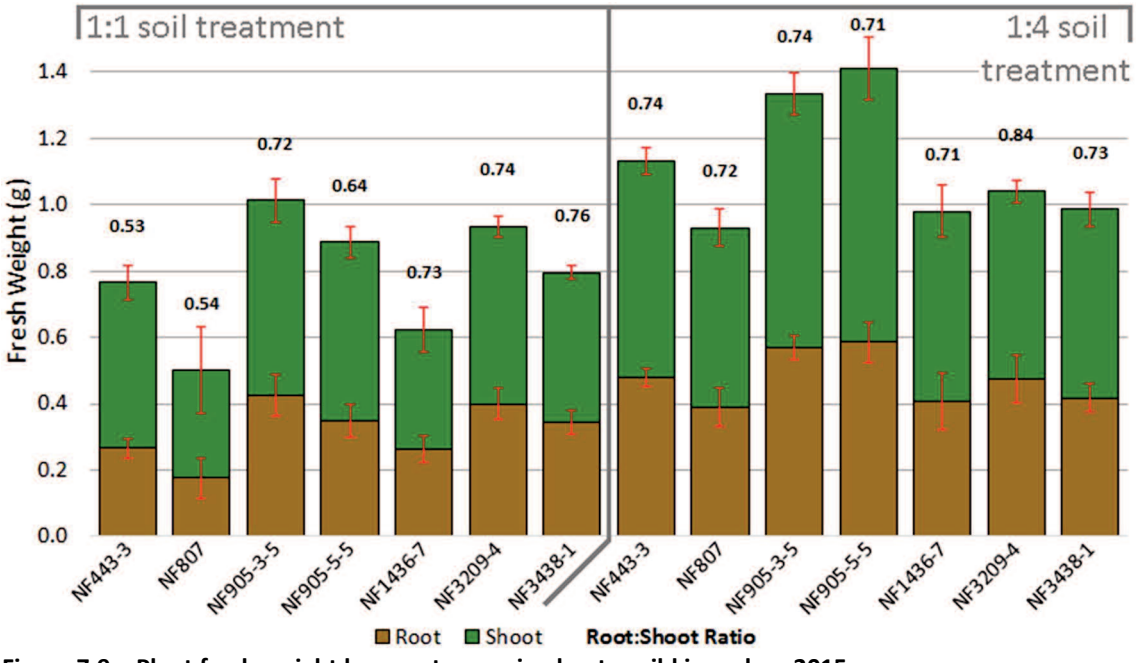


Figure 7.9 – Plant fresh weight by genotype; mixed pots; wild inoculum 2015
Average fresh weight of each plant genotype from the summer 2015 multi-plant pot experiment, split by soil treatment. Displaying standard error and root:shoot ratio of each genotype. The results for NF807 in the 1:1 soil treatment have not been analysed as many of these plants died (resulting in a surviving n of 2).

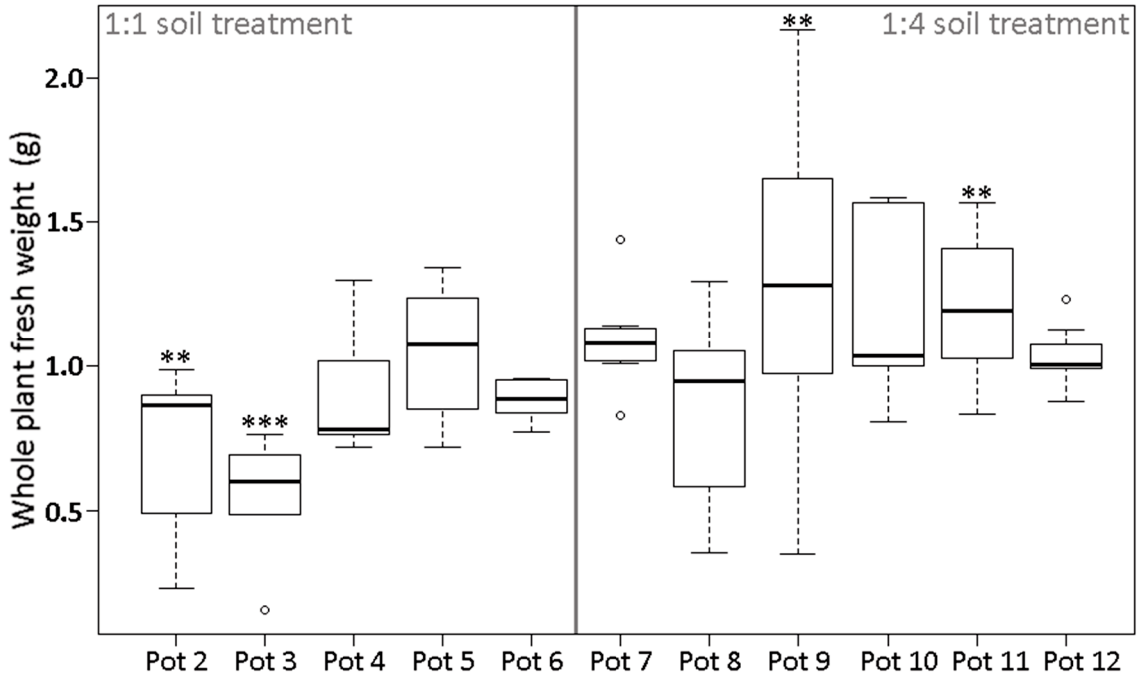


Figure 7.10 – Pot effect; wild inoculum 2015
Whole plant fresh weight for the different individual pots for the 560ml mixed pots from the summer 2015 experiment.

Overall, the plants showed the same pattern in the shoot dry mass (**Figure 7.11**). There were tendencies for certain mutant lines (notably NF1436) to be smaller than the wild type, but this was not consistent between the different soil and pot treatments. This does not seem to be an effect consistent with reduced benefit from mycorrhizal colonisation, as NF1436 has one of the weaker MYC phenotypes of the lines studied (no reduction on hyphal colonisation and deformed arbuscules), so would be expected to perform better or the same as, not worse, than mutants with more extreme mycorrhizal phenotypes like NF3438 or NF905-5-5. NF807 and NF3438, which were both mutants in the same gene (RAM1) showed similar weights (excluding the issues with the 1:1 soil treatment mixed pots), but NF3438 showed a consistently higher root:shoot ratio. In the mixed pots, NF905-3-5 and NF905-5-5 were closely grouped, although less so in the single pot treatment.

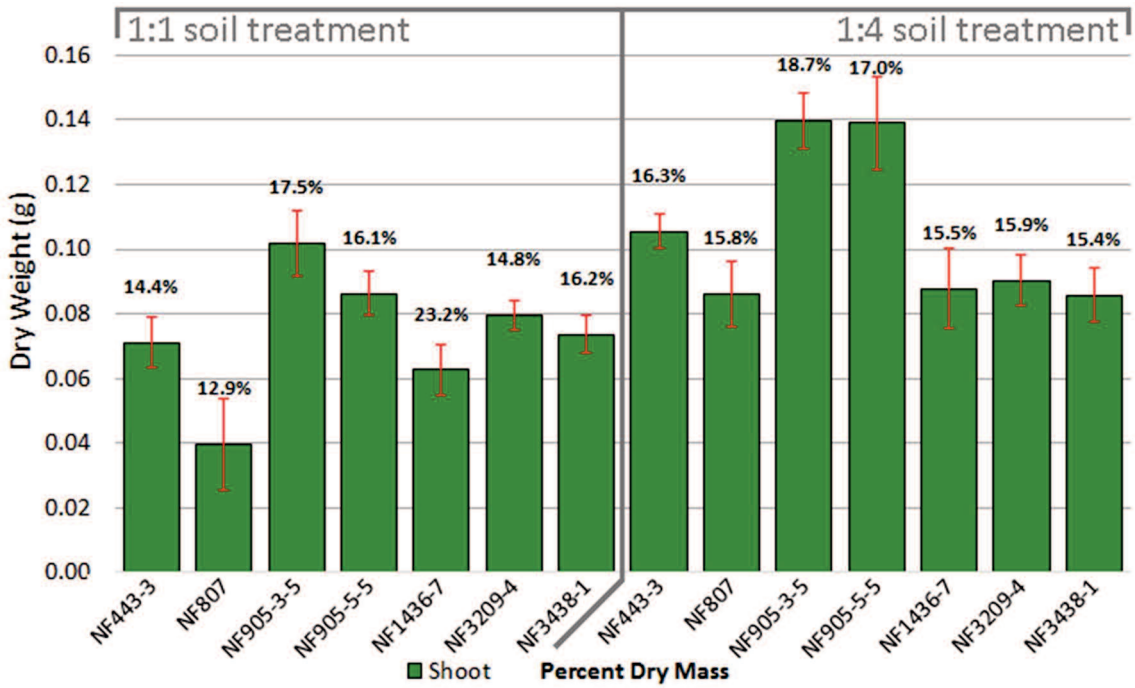


Figure 7.11 – Shoot dry weight by genotype; mixed pots; wild inoculum 2015
Average shoot dry weight of each plant genotype from the summer 2015 multi-plant pot experiment, split by soil treatment. Displaying standard error and the percentage of dry shoot mass remaining for each genotype. The results for NF807 in the 1:1 soil treatment should be treated with caution, as many of these plants died (surviving n=2).

In summary, there was no clear and consistent difference in mass between the wild type and mutants that would indicate a significant growth benefit from mycorrhization, although we did not have non-mycorrhizal controls to see an overall difference. The serious whitefly infestation limited the use of data that could be obtained, adding an uncontrollable variable to the experiment. The biggest differences seen were due to the soil treatment, and thus we decided to try to limit this effect by sieving out spores in the repeat experiment, and to add in a *Rhizophagus irregularis* colonised control, to ensure that we empirically assess any changes in colonisation phenotype with the mixed inoculum.

7.3.3 – Summer 2016 Experiment

We decided to repeat the experiment, taking further controls to reduce the chance of predator infestation. The plants were grown in a growth cabinet to limit the opportunity for contamination with insect predators. The homogenized soil was replaced as the inoculum source with sieved spores and chopped roots, which may reduce the presence of soil pathogens and prevent the problems with water-logging seen in the 1:1 soil to sand:Terragreen treatment. A control of *Rhizophagus irregularis* spores and infected carrot roots was included so we would establish the 'expected'

level of colonisation for each mutant line, and see if the mixed fungal colonisation differed from this before deciding to sequence the fungal DNA in the roots.

By this point, we knew the causal mutation in NF3438 was in RAM1, so we dropped NF807 from the repeat experiment to avoid duplication. NF1436 and NF905 were also dropped, the former as we thought it unlikely we would learn the causal mutation by the end of the study, and the latter because mutations in plant nutrient uptake during the symbiosis were considered the least likely to show differences in colonisation with different AMF species, as the benefit we were selecting for with the nutrient conditions used (P uptake) would be removed in this HA1 mutant. We obtained a number of other mycorrhizal mutants (T-DNA knock-outs for DMI1-3 and RAM2) from a collaborator[390], as these, like CYCLOPS in Gao et al (2001)[388]. We hypothesised that these signalling genes would be more likely to show fungal genotype specific redundancy. We also included NF18637, another reduced AM colonisation mutant, at the request of the collaborator, who was looking to gather phenotype data on that line (See **Table 7.2** for the full list of lines used). With NF905-5-5 dropped from the experiment, NF905-3-5 no longer seemed an appropriate control, so R108 (the parent ecotype used to establish the NF lines) was used instead. One complication with this was that the T-DNA lines available for DMI1-3 and RAM2 were in the A17 background, not the R108 background. Rather than grow two sets of controls, it was hoped that while ecotypes would have an effect on the plant growth character[396], the *R. irregularis* control pots would account for any difference in AM colonisation between the R108 wild type and the A17 lines.

Control plants (R108 with Walmgate Stray inoculum) were harvested weekly from 6 weeks of growth, but showed consistently low levels of colonisation (<10%). After 12 weeks of growth, control plants continued to show little colonisation, and no detectable increase over this time period. Additionally, the plants showed signs of serious stress, especially those in the R108 background, which were growing with 3-4 long shoots with little foliage except around the top, suggesting they were nutrient limited. Therefore, the plants were harvested at this stage, and DNA samples taken. There were large differences in the mass of plants of the different genotypes used in the study, and between the root:shoot ratios of each mutant genotype.

In this analysis, there are several issues to take into account. First, *Medicago* mutants from two different ecotypes were used in this study, and this may have influenced the results. A student's t test suggested that genetic background was a significant factor for fresh weights ($p=2.9x10^{-4}$), but this effect was not significant in the dry weight ($p=0.061$). The cause of this, and the overall inflated root:shoot ratios observed when calculated with fresh weight may have been caused by incomplete drying of the roots. The larger root mass than in the previous experiments made drying with paper towel difficult, as the roots matted and may have retained surface water on the interior of root tangles (**Figure 7.12**). Another possible cause is the longer growth period leading to nutrient depletion, and thus increased invested in root growth as part of the starvation response.

While the students t test is inconclusive at confirming a mass difference between the genetic backgrounds, looking at the root:shoot ratios does suggest a difference there. Plants in the A17 background had a higher root:shoot ratio (2.9-3.7 in fresh weight; 1.2-1.5 in dry weight), so had relatively larger root systems than R108 individuals (with root:shoot ratios of 1.5-2.0 in fresh weight, and 0.6-0.8 in dry weight). A student's t test confirmed this difference ($p=2.2x10^{-16}$). Taking this apparent link between growth character and genetic background, while the plants showed many significant differences in root and shoot weight (see **Figure 7.12** & **Figure 7.13** for significance categorisation), most of the differences were non-significant when ecotype was controlled in a mixed linear model corrected for multiple testing. EP10007 (*dmi1*) had a significantly smaller fresh root mass than EP10013 (*dmi2*) or EP10017 (*dmi3*), and NF443 (*kin3*) had a smaller fresh root mass than NF3438 (*ram1*). For the potentially more accurate dry mass data, the remaining significant difference was that EP10007 was smaller ($p=0.045$) than EP10013. However, while fresh leaves showed no difference once ecotype was controlled for, in the dry masses NF443 was significantly smaller ($p=0.005$) than R108 and EP10013 was smaller ($p=0.018$) than EP10058 (*ram2*).

Overall, as in the summer 2015 experiment, no clear and consistent change in mass we observed between the MYC^+ wild type and the reduced AM colonisation mutants. However, given the low colonisation of all individuals it is unclear if AMF would have had significant effect on growth character in this experiment, regardless of how mycorrhization would change this growth.

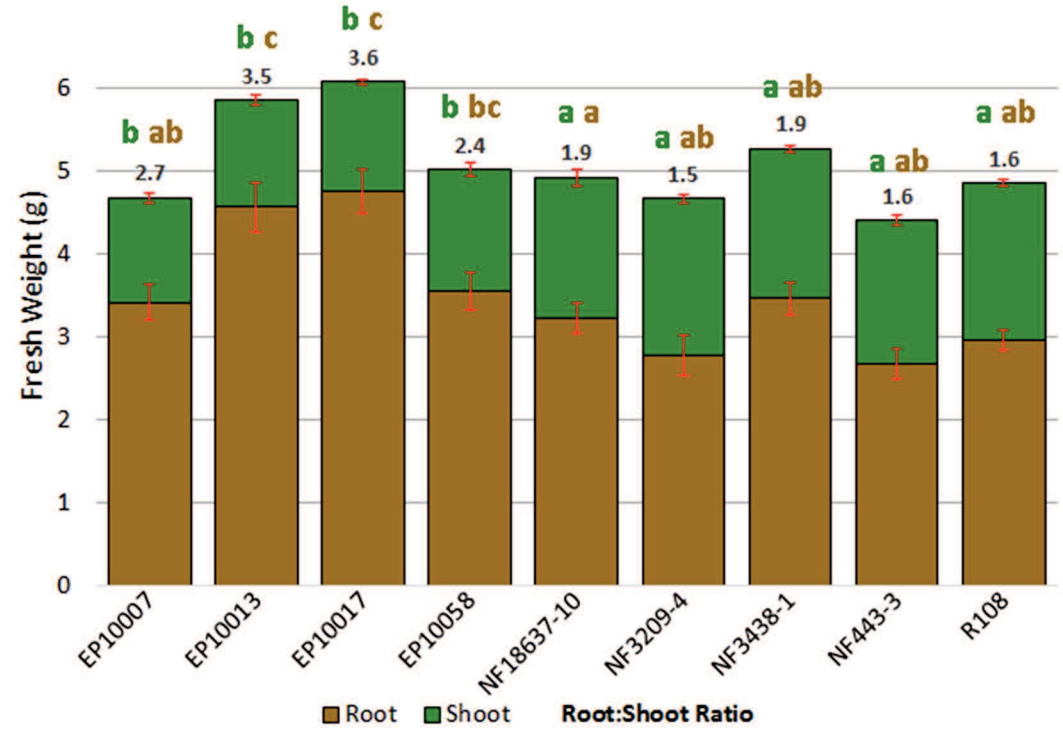


Figure 7.12 – Plant fresh weight by genotype; wild inoculum 2016
Average fresh weight of plants and resulting root:shoot ratios of each genotype from the summer 2016 wild inoculum experiment. Significant differences categories are to $p<0.05$ with a multiway ANOVA and Tukey HSD post-hoc test. The EP# lines were in the A17 background, and the other genotypes in the R108 background. The R108 genotype is the MYC^+ control.

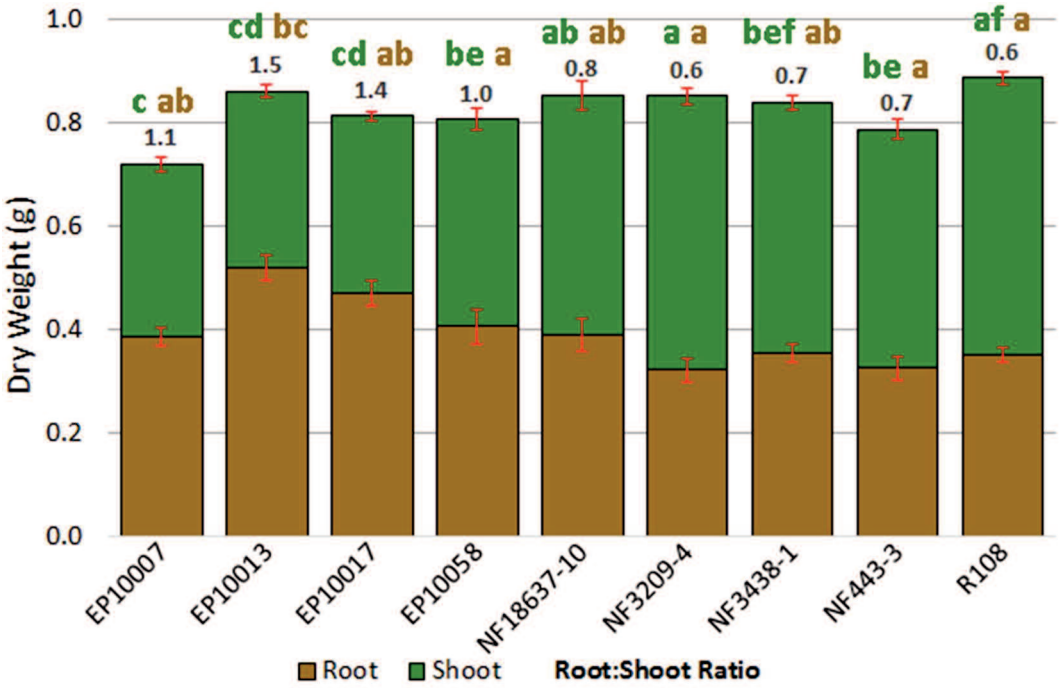


Figure 7.13 – Plant dry weight by genotype; wild inoculum 2016
Average dry weight of plants and resulting root:shoot ratios of each genotype from the summer 2016 wild inoculum experiment. Root dry weight is estimated from the dry weight of a known fraction of each plants total root. Significant differences categories are to $p<0.05$ with a multiway ANOVA and Tukey HSD post-hoc test. The EP# lines were in the A17 background, and the other genotypes in the R108 background. The R108 genotype is the MYC^+ control.

An initial PCR to find Glomeromycotina rDNA10 in wild-type control plant roots from 2015 and 2016 showed little colonisation in the roots (**Figure 7.14**). However, the presence of the strong band pattern in the expected region (550 bp) in a negative control of leaf DNA suggested that the AM1 & NS31 primers[22] were producing off-target amplification from general fungal rDNA10, with leaf endophytic fungi the likely culprit. The experiment was performed again with a different set of primers (AML1 & AML2[391]) considered to be more specific to the Glomeromycotina. These primers alone did not visibly amplify any fungal DNA after 30 PCR cycles (data not presented), so nested PCR with was performed, and yielded visible amplification from most tested samples, although still from leaf DNA control (**Figure 7.15**).

Assuming that the amplification is from Glomeromycotina, then the colonisation of the individual can be inferred from the amplification. Colonisation in the summer 2016 experiment is highly variable, even within the wild type control, suggesting that the main determinant of the colonisation was the starting population of AMF in the sample. Twelve weeks of growth should be ample for spores to colonise roots and begin to replicate in many AMF species, this means that the overall infective potential (loose spores and fungus in plant root fragments) is low. Thus, that there is unlikely to be an even spread of AMF species diversity across the pots, and the colonising species will likely be more correlated with those that started in the pot. In these circumstances, we would not expect to see a consistent genotype effect across the repeats. Because of this, we decided that sequencing the root DNA to establish an OTU population for each sample was unlikely to provide useful data, and thus not worth the cost of proceeding.

Specific amplification of only AMF DNA from plant roots was not successful, with visually colonised plants not producing much fungal sequence (**Figure 7.15**; NF905-3-5 control plants) while amplification from leaf DNA samples was consistent. This made it impossible to judge the origin of the amplicons without sequencing, which was not deemed worthwhile.

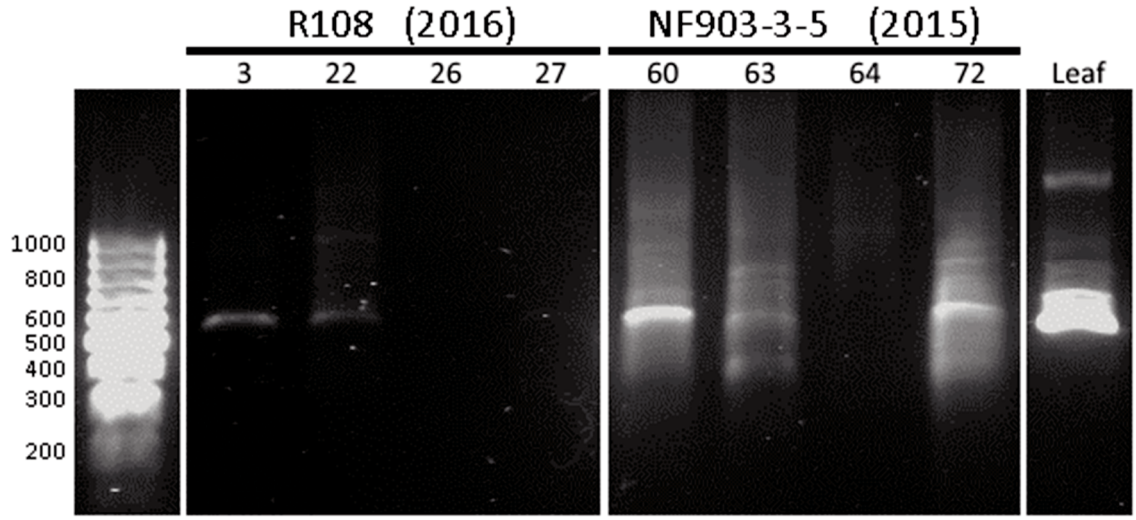


Figure 7.14 – Detection of AMF rDNA10 in wild type plants
AM1+NS31 PCR to amplify a ~550 bp band from the small subunit of Glomeromycotina rDNA10 from *Medicago* roots from the wild inoculum experiments. R108 and NF905-3-5 both have a wild type AM colonisation phenotype with *Rhizophagus irregularis*. DNA from a *Medicago* leaf is included as a negative control.

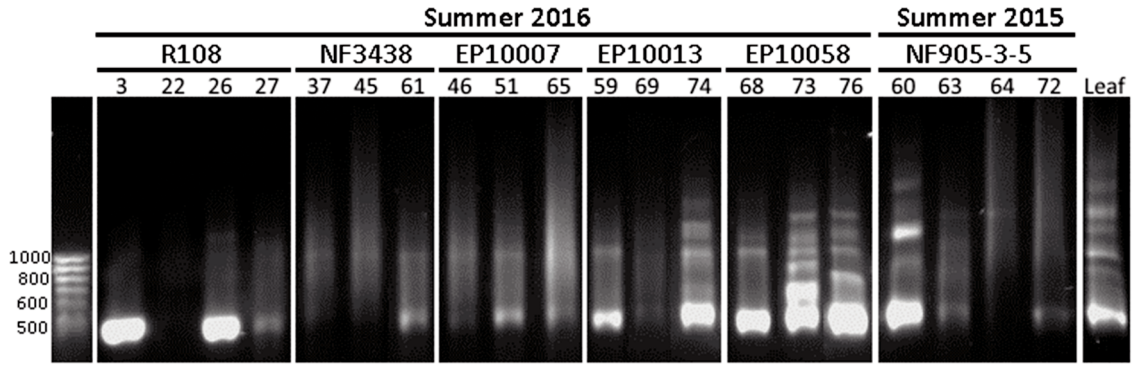


Figure 7.15 – Detection of AMF rDNA10 in mutant plants
PCR to amplify a ~550 bp band from the small subunit of Glomeromycotina rDNA10 from *Medicago* roots from the wild inoculum experiments. R108 and NF905-3-5 both have a wild type AM colonisation phenotype with *Rhizophagus irregularis*, and NF3438 (*ram1*), EP10007 (*dmi1*), EP10013 (*dmi2*) and EP10058 (*ram2*) are all impaired in colonisation with *R. irregularis*. DNA from a *Medicago* leaf is included as a negative control.

7.4 – Discussion

7.4.1 – Can changes in plant mass tell us anything about AM colonisation

Overall, any differences that could be found in the plant masses were inconsistent, and seemed to depend more on soil or pot treatment (2015 experiment) or ecotype background (2016 experiment) than any correlation between mycorrhizal phenotypes. We conclude that whatever differences between the lines in terms of overall growth, they are more likely to be due to natural variation. While it is possible there may be some effect from other tnt1 insertions, as all NF lines carry 20-60 tnt1 insertions.

Although no other disrupted gene appears to have caused an immediately visible phenotype in these lines, it is possible that the disruption to other plant processes could cause the alterations in growth habit seen in the 2015 experiment. But since the R108 control (which lacks any tnt1 insertions) in the 2016 experiment fails to grow significantly better than any other R108 derived line, this effect is likely small.

7.4.2 – Outcomes of the experiment

Both our experiments failed to produce data capable of addressing our hypothesis, the first due to whitefly infestation and the second to the low infection potential of the AMF. The latter may be due to inexperience with the preparation of the soil-free spore mixture, although the root sections would have been expected to provide some inoculation pressure given their visible colonisation. A possible explanation is that *M. truncatula* is a poor host for many of the AMF found in Walmgate Stray, given the differences in the ecosystems they are adapted to. This may have also made *M. truncatula* more susceptible to root pathogens and waterlogging in the 2015 experiment. Alternately, the sand:Terragreen substrate of the 2016 experiment may not have suited to loam adapted AMF. However, the wild type was well colonised in the initial experiment (**Figure 7.4**), so the former explanation is probably incorrect.

While the experiments failed to produce any concrete data, we believe they support the need for further work in this area. **Figure 7.15** suggests that several mutants with reduced *R. irregularis* colonisation phenotypes show equal or greater colonisation than the wild type when infected with the Walmgate Stray inoculum. This hints at support for our hypothesis, that at least some AMF genotypes in the mixture are less sensitive to loss of core MYC genes, including EP10058 (*ram2*). This is most striking in EP10013 (*dmi2*), which fails to show any root penetration with lab strains *R. irregularis* or *Diversispora versiformi*[96,397]. Others mutants like NF3438 (*ram1*), and to a lesser extent EP10007 (*dmi1*), have lower levels colonisation than the wild type on the Walmgate Stray inoculum, suggesting that they may be part of a more essential 'core' gene set. If an experiment as we attempted here could be properly controlled, and produced reliable data that recapitulated **Figure 7.15**, then we submit that it would lend credence to our hypothesis and to that suggested by Gao et al (2001)[388].

7.4.3 – Future directions

A range of possible improvements to this method could be made. First, soil samples could either be taken from multiple ecosystems to provide an even greater range of AMF species, or from an ecosystem that the host plant is native to. British grassland native *Plantago lanceolata* has been used for mycorrhizal studies in the past, but there are few genetic resources available for this system. A model crop system like potato or *Brachypodium distachyon* may yield better results, although again with comparatively less genetic tools available. The use of AMF from soils in the native range of *M. truncatula* or *Lotus japonicus* is preferable if they could be obtained. The use of large soil cores, or planting of knock-out mutants directly into the native environment, present promising solutions to the risk of damage to the native AMF populations by soil homogenisation, but may lead to nurse plant effects obscuring the weaker mutant phenotypes, as well as preventing the use of a *R. irregularis* control. Additionally, only non-GM mutants (i.e. those produced by CRISPR/Cas9 or an EMS-TILLING screen) such as the A17 lines used in this chapter could be grown in this fashion.

We consider the next appropriate step to be a continuation of this methodology with bulk soil (partially diluted by another unreactive substrate) and root fragments, with the origin of the soil matched to the native range of the plant species being tested. This would provide a way to quickly screen for specific AMF genotypes (looking for those that colonise many mutants, and those that seem most strongly affected by plant mutants) that could be studied further to look for a genetic basis (e.g. differing effector complements) for these phenotypes.

7.4.4 – Effectors and the evidence for mutualism?

Overall, we return to the question of the extent of mutualism in the symbiosis. If a fungus can bypass certain steps in the plant's recognition cascade by an expanded effector suite, does that make the interaction more 'hostile'? The fungus is dependent on the plant for carbon, so should be under strong selection pressure to colonise a plant host. Two fungal life history strategies may act against this selection pressure. First, many Glomeromycotina species have demonstrated the ability to return to spore dormancy in the absence of a compatible host. Second, the benefit of not colonising a plant that will not deliver carbon should select against a drive to forcibly colonise the

first root the fungus detects. EP10058 (a *ram2* mutant that shows little arbuscule development and a large reduction in colonisation with *Rhizophagus irregularis*[85,155]) shows more apparent colonisation (**Figure 7.15**) than the other mutants, and at least as much as the wild type. This could indicate differences in carbon metabolism among wild AMF, as well as the expected differences in effector and signal suites.

The selective pressure on the plant to form the symbiosis is weaker, but the abundance of the interaction across evolutionary space shows that it provides a strong benefit to plant fitness (see **Section 1.7**). Over evolutionary time, cheaters on both sides of the symbiosis may arise, but these would be weeded out by selection for increasing stringency of selectivity in the partners. Plants clearly have a large incentive to maintain a complex signalling system and controlled reciprocity of nutrient exchange with their symbionts to ensure that a pathogen does not exploit the limits the symbiosis imposes on plant immunity. We can see that both symbionts would benefit from pre-contact 'handshaking' to ensure they only invest in a highly mutualistic partner. Does the presence of fungal effectors in the interactions imply they are forcing entry on the plant, or should they be viewed as simply a more direct signal to induce plant gene expression changes? The former view could be proved by finding either evidence of co-evolutionary selection of plant R genes to an AMF effector, or a core fungal effector, that if knocked out would lead to AMF colonisation inducing a hypersensitive response in the plant root. One possible candidate for this is *R. irregularis* R5, which shows sequence similarity to the hypersensitive response blocking Six6 class of *Fusarium* effectors[398], although the specific function of R5 in the AMS is currently unknown. The latter view is probably impossible to provide positive evidence for, but would be supported by the existence of a core set of plant genes that remain universally required for the AM symbiosis across all AMF genome-space. This would support the plant playing the active role in downregulating its immune system to allow for fungal colonisation. The CSP has been co-opted into many other symbioses, with recent evidence supporting its involvement in the EcMS and actinorhizal symbioses as well as the rhizobial symbiosis. If this held true for ErMS, OMS or other fungal endophytes, this would support a universal mutualistic pathway to enable symbiosis re-used across the plant kingdom.

Chapter 8 – General Discussion

8.1 – Goals and achievements of this work

8.1.1 – A general summary

The aim of this project had been to improve our understanding of the genetic networks that function in the arbuscular mycorrhizal symbiosis, specifically in the parts not duplicated in the legume common symbiosis pathway. We have had the mixed blessing of working in an era (2013-2017) that has seen an explosion in knowledge of the symbiosis, from the discovery of new signals and receptors[9,40,45,65,66], to the revelation that fatty acids are the primary carbon source of the fungal symbiont[85,86], to new techniques like phylogenetic analysis that have used the wealth of new information from the genomics revolution to discover mycorrhizal genes at an incredibly rapid rate[368–370]. Our approach to gene discovery was to use a forward genetic screen using the Noble Foundation collection of *Medicago truncatula* lines mutagenised by the retrotransposon tnt1[272]. This broad and unbiased approach identified many lines with strong mycorrhizal phenotypes, and has previously led to the discovery of *RAM1*[72] and *HA1*[138].

For this project, we studied five of the lines initially identified as showing mycorrhizal phenotypes and retaining wild-type nodulation. We showed that one of these lines (NF472) had been misidentified, and described the mycorrhizal phenotype of the remaining four (**Chapter 3**). Initial attempts at insertion discovery by invPCR proved overly time consuming, so we devised a screening pipeline by use of Illumina WGS to locate the insertions in our lines (**Chapter 4**). We established co-segregating F_2 populations for the four lines, and attempted to identify the causal insertions (**Chapter 5** and **6**). Unfortunately, the two lines where we achieved this proved to be new alleles of already discovered mycorrhizal genes. NF3438 is an allele of *RAM1*, first discovered by Gobbato et al (2012)[72] (in NF807 and a A17 fast-neutron deletion line)[72]. NF443 proved to be an allele of *KIN3*, described in Bravo et al (2016)[370] (in NF5270). While the former has been extensively studied, we are not aware of work on *KIN3* beyond a basic description of the phenotype. Thus, we hope we have provided greater context on the *kin3* phenotype and the protein's role in the AMS. In regard to the

other two lines, we have shown that none of the tnt1 insertions into NF3209 are causing the phenotype, and provided guidance on the likely location of a hypothetical somaclonal mutation. For NF1436, screening for the relatively mild phenotype was considered time consuming enough to prioritise the other three lines. We have shown that qPCR for phosphate exchange related genes (**Section 3.3.5**) is not a good alternative to visual screening for NF1436–derived populations, as the mutation does not significantly affect this aspect of the symbiosis. Importantly, however, none of the insertions found in NF1436 appear in the list compiled by Bravo et al (2016)[370] or Delaux et al (2014)[369]. This suggests that the causal mutation of NF1436 may be conserved in non-mycorrhizal plants, having some additional function outside the AMS, as such genes cannot be found by phylogenetic approaches. The tnt1 insertions we in located NF3209 also are not in any genes known from Bravo or Delaux, but as that line likely has a casual somaclonal mutation, we cannot comment on its novelty until SNP calling has been performed.

The period of the study has also seen ever more evidence of the major disconnect between our genetic knowledge of the symbiosis, and the inability to predict the outcomes of the symbiosis. The effects of the symbiosis on plant growth appear defined by complex genotype x genotype x environment interactions (see **Chapter 7**). For example, multiple fungal isolates of the same morphospecies will have different effects on plant growth[399]. Changing plant ecotype or soil environment has similarly different effects on the host with the same fungal isolate. Without being able to have at least some predictability in interactions between plant, fungal or environmental variables, efforts to promote the symbiosis as a biofertiliser will be limited. We attempted to screen the effects of plant/fungal genotype interactions, using plant KO mutants and fungal OTU abundance as a proxy for colonisation success. The aim was to establish if certain plant mycorrhizal genes or pathways were bi-passable by certain fungal species, as current evidence hints at[388]. This would inform future plant genomic engineering efforts to improve the symbiosis, by establishing which genes were universally important for the symbiosis. However, due to technical difficulties, we were unable to finish this work within the time period of the study.

8.1.2 – Further understanding the role of KIN3 in the AMS

We hope to better answer this question in the future with a number of experiments. AM colonised *Medicago* composite plants expressing YFP tagged KIN3 will allow us to confirm the predicted periarbusculuar localisation of KIN3 (and potentially the orientation in relation to the membrane, see **Section 6.4.4**). The RNA sequencing of *ram1* and *kin3* lines should allow us to examine the respective transcriptional effects of each protein, and see if they overlap (as predicted by the model; **Section 6.4.7**). Alternatively, it may show us that KIN3 only influences a subset of RAM1-related functions, or has additional roles outside of RAM1 function.

8.2 – Many gaps remain in our current knowledge of the AMS

8.2.1 – Locking the parts of the jigsaw together

One aim of this project was to try and link up the otherwise disparate parts of the AMS signalling network (**Figure 8.1**). While we have not added any novel genes to our knowledge, we believe that KIN3 represents a potential link-point, maintaining RAM1 expression and mycorrhizal cell fate past the cusp of CSP signalling and into the control of nutrient exchange.

We maintain that current knowledge of the AMS remains, despite the recent explosion in knowledge, bound to discrete sections. We know, for instance, that intercellular accommodation of the fungus (arbuscule, PPA etc.) requires the CSP, but that the promotion of lateral root branching only requires a subset of CSP genes. Clusters of genes are known to be involved in facilitating hyphal spread through the root, and in suppressing host immunity, but what the immediate regulator of these processes (and how it links back to fungal signalling) is unknown. These connections remain an important question that needs to be answered.

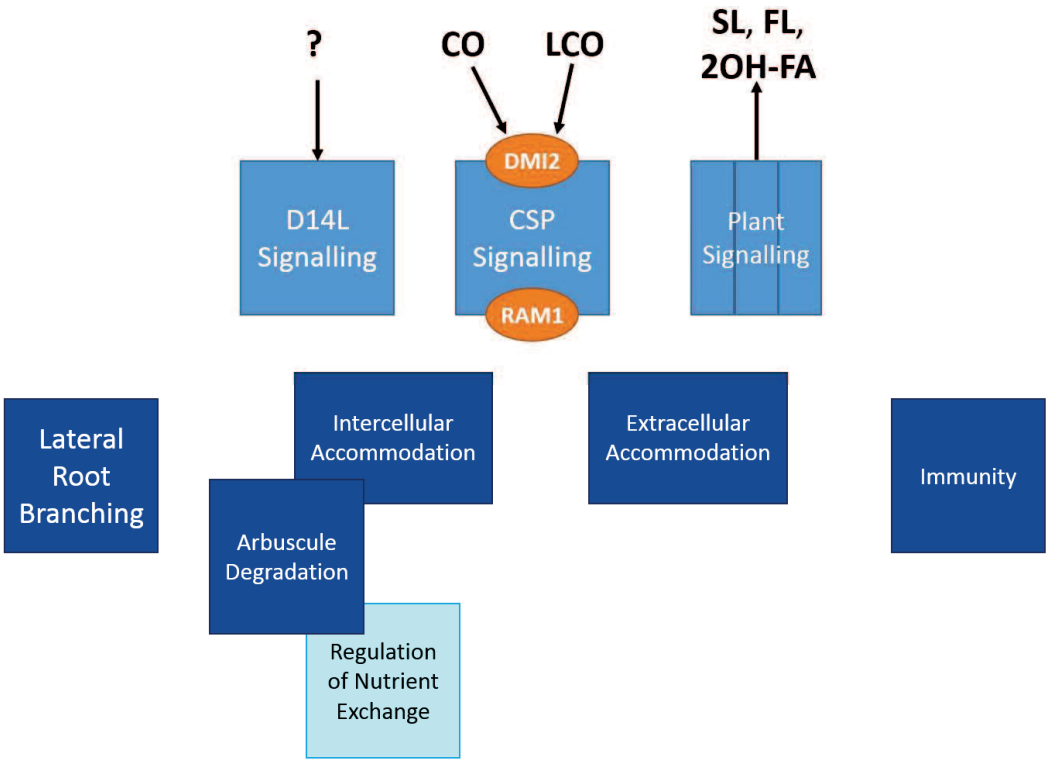


Figure 8.1 – Knowledge of the AMS tends to collect into discrete modules
Connecting the modules of AMS function will be needed to deliberately engineer a more efficient symbiosis into crop plants.

8.2.2 – Invitation or invasion: how mutualistic is the AMS?

Surveying the mycorrhizal genes known today, one thing stands out – the paucity of strong symbiotic phenotypes. MacLean et al (2017)[400] report only 7 of the 67 genes with a negative colonisation phenotype result in complete abolition of fungal colonisation. While the fungus cannot complete its life cycle without a number of other genes (*RAM1, FatM, KIN3* etc.), most mutants show only a reduction in colonisation or increased arbuscule turnover, often having less than a halving of wild type colonisation.

The seven genes are involved in the CSP (*DMI1* (*CASTOR*/*POLLUX* outside *Medicago*), *DMI2*, *DMI3*, *NENA*), a receptor for an unknown fungal signal (D14L)[66] and a cell surface transporter (*NOPE1*)[67] (see **Section 1.3.2** and **1.3.3**). The proposed LCO/CO receptors of the CSP (which include genes like *LYK3* and *LYR3*) do not produce as strong a phenotype as their co-binding factor *DMI2*, supporting reports of semi-redundancy along this pathway. The other two 'choke points' of the symbiosis, prone to prevent AMF sporulation, are arbuscule formation (e.g. *RAM1, VAPYRIN*) and arbuscule maintenance/degeneration (produced by genes involved in reciprocal

nutrient exchange like *PT4, HA1, FatM, DIS, RAM2* etc.). Other processes (see **Section 8.2.1**) seem less critical to the symbiosis, including extracellular accommodation of the fungus and downregulation of host immunity. While this could be an artefact of redundancy, it could also suggest that these processes are more likely to be under the control of the fungus (via effectors). As the strongest phenotypes are mutations in genes involved in the detection of fungal pre-contact signals, it seems clear that the plant must detect a fungal 'handshake', without which the plant can block fungal root entry. This does leave open the ability of the AMF to 'cheat' the plant host if they can pass this first hurdle, as the plant appears unable to remove a fungus entirely once it has entered (e.g. a *pt4* mutant).

The exception to this categorisation of genes involved in detecting pre-contact fungal signals is NOPE1. Its function is unclear, but it is likely exporting some factor required by the fungus. Contrary to the description in MacLean et al (2017)[400], Paszkowski et al (2006)[401] observed that while most AMF hyphae that encountered the root surface did not penetrate, occasional patches of wild type colonisation did occur. Paszkowski et al (2006)[401] ascribe this to epigenetic instability in the mutation, but this phenotype could also be caused by a fungal requirement for a NOPE1-transmitted signal to initiate the root penetration process. In this hypothesis, the occasional lapses in the non-colonised phenotype would be caused by chance leaks in surface cells allowing efflux of the hypothesised signal, or by chance fluctuations in a stochastic process controlling the switch to initiate the penetration event by the fungal appressoria. Whatever the identity of the plant signal, it likely has other functions, as *Arabidopsis* root exudates rescue the *nope1* phenotype, indicating conservation of the signal in non-AM plants. If this hypothesis is correct, it would further emphasise the mutualism of the symbiosis, as the obligate biotrophic AMF must both ask and receive permission to enter the root.

These phenotypes may be specific to certain fungal genotypes, a behaviour restricted to the lab strains of the old *Glomus* clade. With our wild inoculum (see **Section 7.4.2**) we observed potential AMF DNA signatures inside the roots of *dmi2*. It is also unclear how these seven mutants would react when in nurse culture. AMF with abundant carbon have been shown to force entry to *Arabidopsis*[193], an interaction that is wholly detrimental to the plant, but still involves some AMS-like genes expression. This latter

fact, along with the abundance of potential effectors being discovered[402], still supports a strong role for the AMF in controlling plant reactions despite their apparent requirement to pass the initial 'handshake'.

8.2.3 – Conservation across Embryophyta: to what extent is the AMS universal?

Phylogenic approaches to finding mycorrhizal genes rely on conservation of the symbiosis across plant species. But to what extent is the genetic machinery of such an ancient symbiosis actually conserved in plants with functional conservation? In *Medicago*, we see a single gene having the role of several in other plants (i.e. DMI1 or *Lj*CASTOR and *Lj*POLLUX; PT4 or *Os*PT11 and *Os*PT13), as well as the inverse (i.e. three DELLA genes in *Medicago* and one in *Pisum sativum*). Does this represent a functional change, evolution to fill some niche, or simply genetic flux around a stable point, with genes gained and lost with random events (e.g. whole genome duplications), while function remains stable over time? While, for example, phosphate transport employs PT family proteins in all known species, other processes like sugar transporters use different gene families (SUT family proteins in most studied plants, but SWEET family proteins in *Solanum tuberosum*)[135]. This is more indicative of co-evolution of later traits (e.g. ancestrally, AMF took up apoplastic sugars that naturally leaked from the root cells[8], and later the plants evolved deliberate export to the periarbuscular matrix to supplement this supply to reward for cooperatively) than duplication or copy number reduction of PT transporters.

Did the AMS have a central, ancestral function (e.g. exchange of carbon for phosphate), with other functions (e.g. deliberate plant sugar export, fungal water transport, mycorrhizal induced resistance) acquired later? Similarly, different AMF species show differing effects on plants[237,403,404], and we expect to see this variability in genotype x genotype interactions to apply to plant mutants as well (see **Chapter 7**, Gao et al (2001)[388]). Phosphate nutrition would be default assumption as the 'base' pathway from which the symbiosis was elaborated. In extant AMF species, phosphate delivery to the plant can be quite variable[405], but this could be explained as a secondary loss in species that offer other benefits. The extent to which AMF use the same general pathway of interactions with their host is a topic of much importance for future approaches to use the AMS in crop plants, which will have to interact with many different fungal genotypes in the field.

8.3 – Looking towards the future

8.3.1 – How useful are forward genetics to our understanding of the AMS?

This project is a good showcase for the current problems with forward genetics as a technique for gene discovery, namely that it is often slower than reverse genetic approaches such as gene expression analysis or phylogenetics. We thus must answer the question – are forward genetic approaches a useful tool for gene discovery going forward?

Forward genetics works best on a poorly understood system. As with invPCR (see **Section 4.1**), we are essentially engaged in a coupon collector's problem. Every gene discovered increases the effort needed to find the next, as we rediscover new alleles of known mycorrhizal genes. How many mycorrhizal genes are there? MacLean et al (2017)[400] collates 73 genes with proven phenotypes, and we are aware of several more. Phylogenetic approaches will not find genes conserved for other functions (e.g. strigolactone synthesis or karrikin signalling related genes), and potentially could be confused by convergent evolution, where different genes perform the same function in the symbiosis in different plant species (e.g. PT4 transports Pi in *Medicago*, and controls arbuscule degeneration, whereas in rice this function is divided between PT11 and PT13[406]). However, phylogenetic estimates are also highly variable (from a high of 1632 in Farvo et al (2014) to a low of 138 in Bravo et al (2016)). Estimates vary depending on algorithmic stringency and the plant genomes tested. Additionally, there has been little attempt to quantify the success rate of these predictions. Bravo et al (2016)[370], which had the most stringent methodology, examined mutants of 7 previously undiscovered genes, and found mycorrhizal phenotypes in 6 of them. On top of this variation, there is little agreement between the three studies[368–370]. Bravo et al (2016: Supplementary Figure 2)[370] nicely illustrates the extent of this problem, which they partially blame using different non-AM controls, and are partially unable to explain.

Gene expression based estimates are even more varied. Gaude et al (2012)[100] and Handa et al (2015)[311] suggest that there are somewhere in the region of 900 to 1900 genes[x] differentially regulated during the AMS in *Medicago* (see **Section 4.4.2**), but Bravo et al (2016)[370] puts the number at over 5100[xi]. These studies will also miss genes with post-transcriptional regulation, but this should be accounted for in proteomic screens such as reported by Aloui et al (2017)[373]. Cell specific expression (e.g. expression of KIN3 in arbuscule containing cells but not in bulk root tissue; see **Section 6.4.1**) is also a confounder of such studies, but again current technological advances such as laser capture microdissection and the decreasing amount of sample necessary for mass spectrometric identification will help solve this problem, and allow cell and cell compartment specific 'omics to be carried out. Still, genes like DMI3 (whose root expression is not regulated by the AMS[313]) would not be found by this method.

Forward genetics is potentially less biased than reverse genetic approaches, but have unfindable targets as well, as there is little likelihood that a pair of genes with redundant function would both be affected in the same line, even for high mutation number approaches like EMS. Genes that are downregulated or have proteins targeted for degradation as a part of the AMS are also important, and likely to be hard to find with forward genetics, as are genes with only mild mycorrhizal phenotypes, as the variability of colonisation makes increases or mild decreases in colonisation difficult and time consuming to screen for.

There is also a third method of gene discovery, namely looking for the associations that known mycorrhizal genes possess. For example, a transcription factor like RAM1 must bind to a specific DNA sequence, so any gene with that sequence in the promotor is a good candidate for playing a role in the AMS. Similarity, proteins that are bound to or post translationally alter others are also good candidates (e.g. assays for kinase targets as described by Jayaraman et al (2017)[407], or similar assays looking for modifications like ubiquitination by LIN[123]). These techniques cannot find novel pathways by themselves, but have the advantage of being work that we would like to do anyway. In

[x] Estimates of 3% and 6% of the genome respectively, based on the 31,500 high confidence genes reported in the A17 4.0 genome assembly[295], rounded to nearest 100.

[xi] 3,597 upregulated and 1,543 downregulated genes, of a total of 23,496 genes examined.

establishing the function of important genes (after they have been identified as being involved in the symbiosis), we will naturally come on to test transcriptional and post-translational control (e.g. we are aware of both promotor binding screens being performed on RAM1 and CBX1).

In the end, the future usefulness of different techniques for assigning gene function will come down to a cost-benefit analysis. Bravo et al (2016)[370] and Delaux et al (2014)[369] leave the community with over a hundred potential candidate genes that have never been examined, and many more (e.g. those phenotyped in Bravo) for which no clear idea of gene function exists. We suggest that expending large amounts of effort on further forward genetic screening is wasteful, when most candidates discovered will likely be recapitulations of these phylogenetic candidates or previously described genes. Study of the function of these phylogenetic candidate genes would seem the more efficient use of resources, especially when focused on likely signalling or regulatory genes like kinases or transcription factors. Finding the targets of these genes will allow us to build up a map of the mycorrhizal growth program, and potentially fill in gaps missed in the phylogenetic studies.

In short, phylogenetic approaches are likely missing many mycorrhizal genes, but will only improve as more plants are sequenced and algorithms are refined. Given the huge pool of predicted candidate genes now known, and the high success rate at confirming these predictions[370], we believe efforts should be concentrated away from gene discovery and onto understanding the role of the candidates in the AMS.

8.2.2 – AMF into agriculture: what is the best approach?

One of the main reasons for studying the AMS is for its potential utility in agricultural practice, both to improve phosphate use efficiency in the face of dwindling reserves of high quality rock phosphates, and to improve crop resiliency to pathogens and climatic stress. While the community aims to eventually breed crop lines to more efficiently utilise the symbiotic uptake pathway, there are two ways to approach the fungal side of the equation. Agricultural soils are often heavily fertilised, fallowed and tilled, all practices that reduce the population size and diversity of AMF communities. Despite this, most agricultural fields that have been tested retain significant pools of AMF[22,408].

Thus, we can either supplement or replace these existing field stocks with cultured AMF, or try to conserve and promote the resurgence of existing stocks.

While there are no known AMF species that are always detrimental to the host, there is a wide range in the degree to which the host benefits. Ideally, we would want to maintain soil stocks of AMF species which give the greatest benefit to the host (be that in terms of mycorrhizal growth response or stress resilience). This is complicated by the fact that fungal morphotype and OTU-type are not predictive of host plant effects, with isolates of the same species displaying different growth effects[399,409]. While some field trials of AMF inoculation have produced positive results[26], there has been significant difficulty translating the results we see in the lab into positive gains in the field, either in conservation biology or in agriculture.

Mass cultivation of AMF stocks is difficult and expensive, making it hard to profitably supplement fields. This cost will likely be reduced in time as methods improve. However, while the supplementation effect reported by Hijri (2016)[26] was positive on average, some 20% of the trials showed non-profitable or negative effect on yield. All farmers will have different soil and climatic conditions, even between fields, and there will be multiple genotypes of each crop in production across a country, and different weather effects year on year. Thus, mass producing a single fungal isolate optimised on a certain crop may struggle to perform in the field, regardless of how much technological improvement can bring down costs. An alternative to broad application of AMF might be the addition of spores to seed coats. AMF spores are hardy, so provided they could be transported with or grow down the growing seedling root from a seed coat, this approach could deliver plant and fungus combinations optimised for genotype interactions, if not necessarily matched exactly to the environment. Hormones are already included in some of these seed coats, so pro-symbiotic signals could also be applied in this manner.

The other route, that of conservation and management of existing field stocks, would rely on the production of seedstock with a broad range of host AMF, and modification of farming practice to try and promote AMF survival without also allowing pathogen proliferation (one of the main reasons for the use of practices detrimental to AMF like

tilling, crop rotation and fallowing). The main logistical effort for this approach would be to help farmers to understand the conditions of their land, and reliably predict which plant lines would be best for those conditions. This approach has the potential to be cheaper for the farmer, assuming governmental or NGO aid in regard to education and soil testing. This will be important for subsistence farmers in the global south, who may otherwise be shut out of some current approaches. Cheaply delivering fertiliser and pesticide to these regions is already difficult, and AMF spore transport will require more temperature control and have a shorter shelf life than these.

Again then, the problem comes back to our current inability to predict genotype x genotype x environment interactions. Without at least some predictability, we cannot make useful recommendations for crops to plant or fungal supplementation. Without predictable outcomes there will be little incentive for farmers to change their practice. Comparisons to the three body problem aside, this is a busy area of current research, but is unlikely to really be 'solvable' in the foreseeable future given the inherent complexity of the system. A more 'blind' approach to plant genomic engineering here may also pay off. Crossing landraces of crops adapted to a specific climate/soil combination, and thus co-evolved with the fungal community (AMF communities vary more on local condition and soil type than by continent[410]), into high yielding varieties is a possibility. However, as the AM interaction involves many genes across the genome, this effect is unlikely to be concentrated into a single locus trait that would allow easy breeding. A possible alternative may be guided evolution experiments, using plants made more dependent on the symbiosis (i.e. by knocking down direct P uptake transporters like PT1[411]) to allow easier population screening, and with an increased mutation rate to speed up the work (e.g. by altering DNA synthase proofreading or chemical mutagenesis of seeds). These lines could be grown through a number of generations in soil/climate pairings of interest, and the changes seen in the most successful lines integrated into elite varieties by genetic engineering.

Appendices

Appendix 1 – Plasmids

A1.1 – List of all plasmids used

List of all Addgene plasmids (**Table 9.1**), PCR amplicons (**Table 9.2**) and new plasmids (**Table 9.3**) generated for producing complementation mutants (see **Chapter 6**).

Table 9.1 – Addgene plasmids for complementation

Addgene Number	Role	Level	Description
pICH49477	CDS	0	DsRed intron-less coding sequence
pICSL50017	C_t Tag	0	YFP fusion
pICSL11024	Gene	1	*E. coli* nptII gene with *A. tumefaciens* nopaline synthase promoter & 5' UTR (270 bp) and octopine synthase 3' UTR & terminator (732 bp)
pICH51266	Promoter	0	CMV 35S promoter (1338 bp) & TMV Ω 5' UTR (62 bp)
pICH85281	Promoter	0	*A. tumefaciens* mannopine synthase promoter & 5' UTR (383 bp)
pAGM8031	Structural	M	Level M Position 1 Acceptor
pICH47742	Structural	1	Level 1 Position 2 Acceptor
pICH47751	Structural	1	Level 1 Position 3 Acceptor
pICH50892	Structural	1	Level M Endlinker 3
pUAP1	Structural	0	Universal Level 0 Acceptor
pICH41414	Terminator	0	CMV 35S 3' UTR & terminator (204 bp)
pICH77901	Terminator	0	*A. tumefaciens* mannopine synthase 3' UTR & terminator (252 bp)

Table 9.2 – PCR amplicons for complementation

Name	Role	Description
L0_51_CDS_NS	CDS	Mt7g116510 coding sequence (inc. introns) with STOP codon removed
L0_51_CDS_S	CDS	Mt7g116510 coding sequence (inc. introns)
L0_65_CDS_NS	CDS	Mt7g116650 coding sequence (inc. introns) with STOP codon removed
L0_65_CDS_S	CDS	Mt7g116650 coding sequence (inc. introns)
L0_51	Gene	Mt7g116510 in native context (+2000 bp at start, +500 bp at end of predicted gene)
L0_65	Gene	Mt7g116650 in native context (+2000 bp at start, +500 bp at end of predicted gene)
L0_51_Pro	Promoter	2 kb upstream of Mt7g116510 START
L0_65_Pro	Promoter	2 kb upstream of Mt7g116650 START
L0_51_Ter	Terminator	500 bp downstream of Mt7g116510 STOP
L0_65_Ter	Terminator	500 bp downstream of Mt7g116650 STOP

Table 9.3 – Plasmids produced for complementation

Name	Assembly	Enzyme	Description
L1P2_51	pICH47742 + L0_51	Bsa1	Intermediate assemblies for the different genes of interest (Mt7g116510 & Mt7g116650) in the three combinations (native context '51', native + YFP C_t fusion '51Y' and overexpression + YFP C_t fusion '51YM')
L1P2_51Y	pICH47742 + L0_51_Pro + L0_51_CDS_NS + pICSL50017 + L0_51_Ter	Bsa1	
L1P2_51YM	pICH47742 + pICH85281 + L0_51_CDS_NS + pICSL50017 + pICH77901	Bsa1	
L1P2_65	pICH47742 + L0_65	Bsa1	
L1P2_65Y	pICH47742 + L0_65_Pro + L0_65_CDS_NS + pICSL50017 + L0_65_Ter	Bsa1	
L1P2_65YM	pICH47742 + pICH85281 + L0_65_CDS_NS + pICSL50017 + pICH77901	Bsa1	
L1P3_DsRed	pICH47751 + pICH51266 + pICH49477 + pICH41414	Bsa1	Visible/positive selection marker
LMP1_51	pAGM8031 + L1P2_51 + pICH50892	Bpi1	Finished T-DNA constructs for the different genes of interest (Mt7g116510 & Mt7g116650) in the three combinations (native context '51', native + YFP C_t fusion '51Y' and overexpression + YFP C_t fusion '51YM')
LMP1_51Y	pAGM8031 + L1P2_51Y + pICH50892	Bpi1	
LMP1_51YM	pAGM8031 + L1P2_51M + pICH50892	Bpi1	
LMP1_65	pAGM8031 + L1P2_65 + pICH50892	Bpi1	
LMP1_65Y	pAGM8031 + L1P2_65Y + pICH50892	Bpi1	
LMP1_65YM	pAGM8031 + L1P2_65M + pICH50892	Bpi1	

A1.2 – Mt7g116510 complementation plasmids

The R108 genomic sequence of the putative GDSL acetylcholinesterase Mt7g116510 (1864 bp), and sufficient native sequence to encompasses the promoter (2000 bp) and terminator (500 bp), was obtained by aligning the A17 sequence[303] to R108 sequence data obtained by WGS of NF1436 line (as this does not contain any tnt1 transposon insertions at the region of interest). The sequence contained two restriction enzyme recognition sites in the promoter (a Bsa1 site at -1850 bp and a Bpi1 site at -1400 bp) and another in the terminator (a Bpi1 site 230 bp after the end of the coding sequence). It was pre-amplified as a single 4870 bp fragment from R108 leaf gDNA, and adaptor ligation was carried out in three sections:

- **51-1** (a 414 bp stretch of the promoter from the Bsa1 site to the Bpi1 site)
- **51-2** (a 3487 bp stretch between the Bpi1 sites, comprising the majority of the promoter and the whole CDS)
- **51-3** (a 358 bp stretch from the terminal Bpi1 site to the assumed end of the native terminator sequence, 500 bp after the STOP codon)

These sections were joined into a Level 0 module (**L0_51**) by digestion ligation, and fragments corresponding to the promoter, terminator, CDS and CDS lacking a STOP codon were produced as their own L0 modules. The Level 0 components could then be assembled into Level 1 Position 2 parts, to fit into the standard transformation construct. This consists of three Level 1 parts inserted into a Level M Position 1, with a standardised antibiotic marker in Position 1 (pICSL11024) and a visual marker flanking the complementing gene (L1P3_DsRed). This combination of chemical and visual markers would allow selection of hairy roots that had been transformed with the complete complementation construct. Alternative LM plasmids were produced with for fluorescently tagged (GDSL::GDSL-YFP) and overexpressor (*At*Mas::GDSL-YFP) constructs.

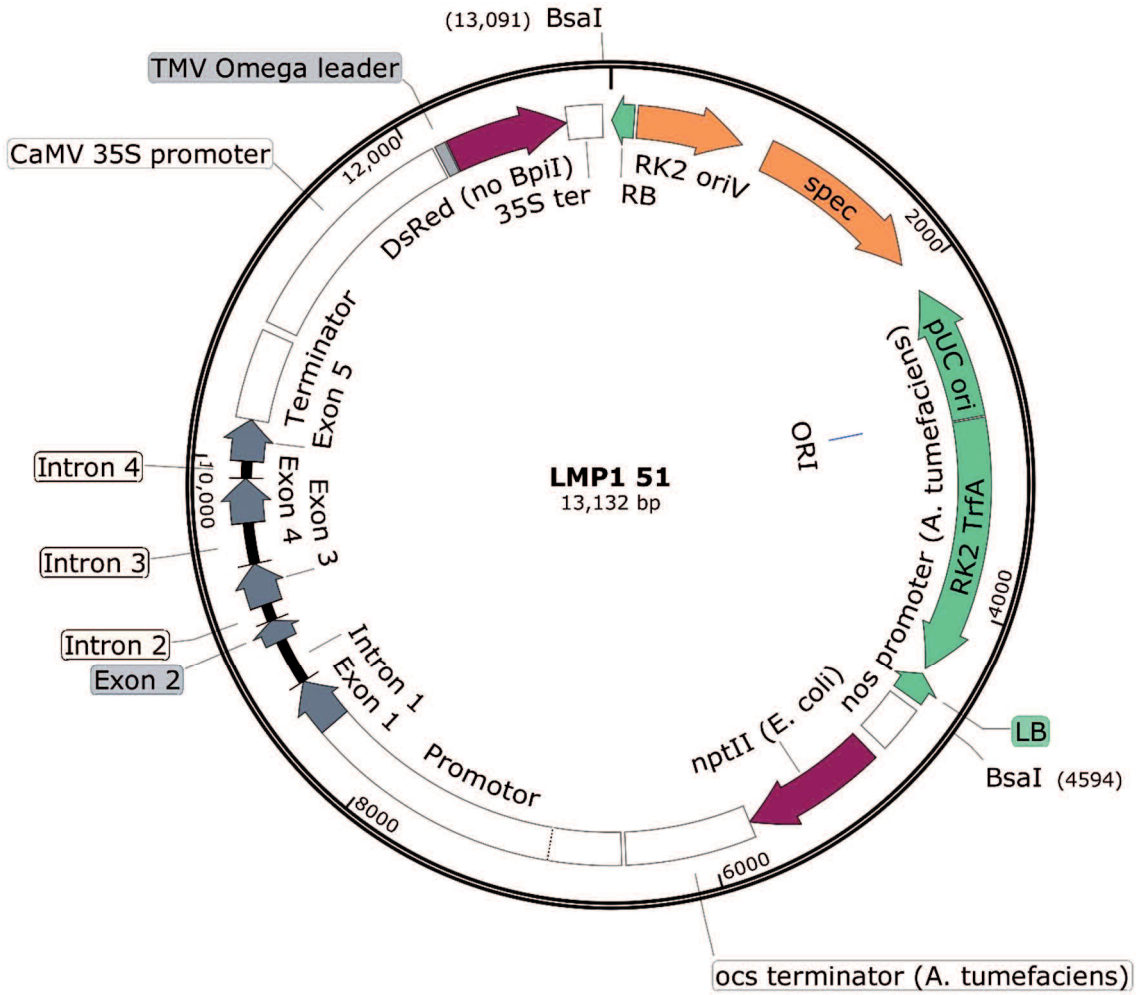


Figure 9.0.1 – LMP1_51
Level M Position 1 finished construct made from the negative antibiotic selectable marker pICSL11024, the genomic sequence of 7g116510 in its native context, and the positive visual selectable marker L1P3_DsRed.

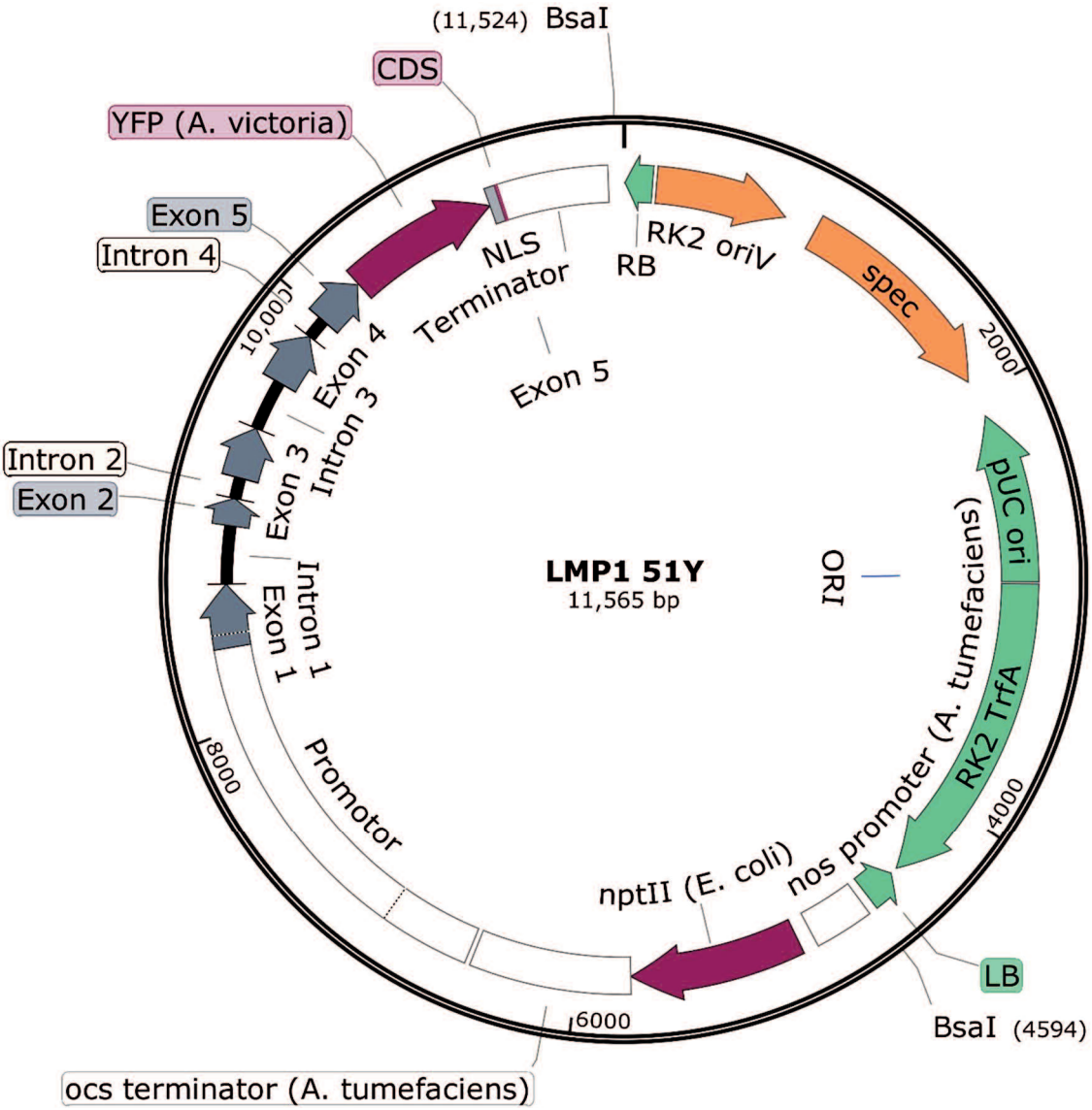


Figure 9.0.2 – LMP1_51Y
Level M Position 1 finished construct; as LMP1_51 with a YFP C terminal fusion added to 7g116510

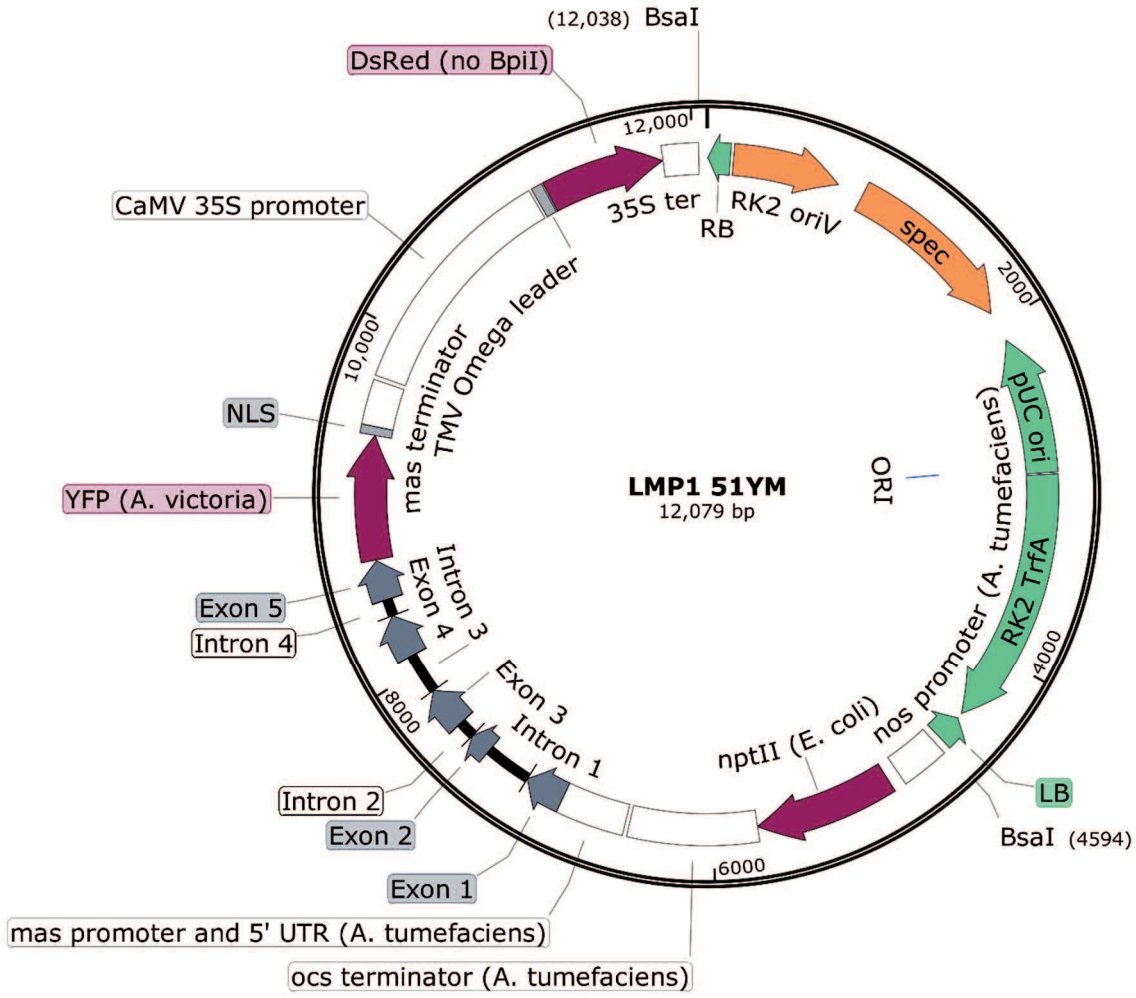


Figure 9.0.3 – LMP1_51YM
Level M Position 1 finished construct; as LMP1_51 with a YFP C terminal fusion added to 7g116510 and the native promoter, terminator and UTR replaced with those derived from the *Agrobacterium tumefaciens* mannopine synthase gene

A1.3 – Mt7g116650 complementation plasmids

The sequence of the putative cell wall associated tyrosine kinase Mt7g116650 (3174 bp) and its native promotor and terminator was obtained as in **A1.2**. The sequence contained a Bpi1 recognition site 600 bp into the coding sequence (in Exon 2). Given the length, it was pre-amplified as two sections (3036 and 3278 bp respectively). The whole gene, especially the promoter, is extremely GC poor (25% compared to 40% in the exons), so some difficulty was had siting primers, which reduced the promoter used to 1853 bp. Additionally, a 150 bp region of the promoter beginning at -260 bp is conserved across several different genomic locations, although these do not generally appear near genes. The two sections produced from adaptor ligation:

- **65-1** (a 2458 bp stretch including the promoter up to the Bpi1 site in middle of the 2nd Exon)
- **65-2** (a 3074 bp stretch from the Exon 2 Bpi1 site and the end of the hypothetical terminator)

As in **A1.2**, these were assembled into a L0 constructs, then L1 and LM.

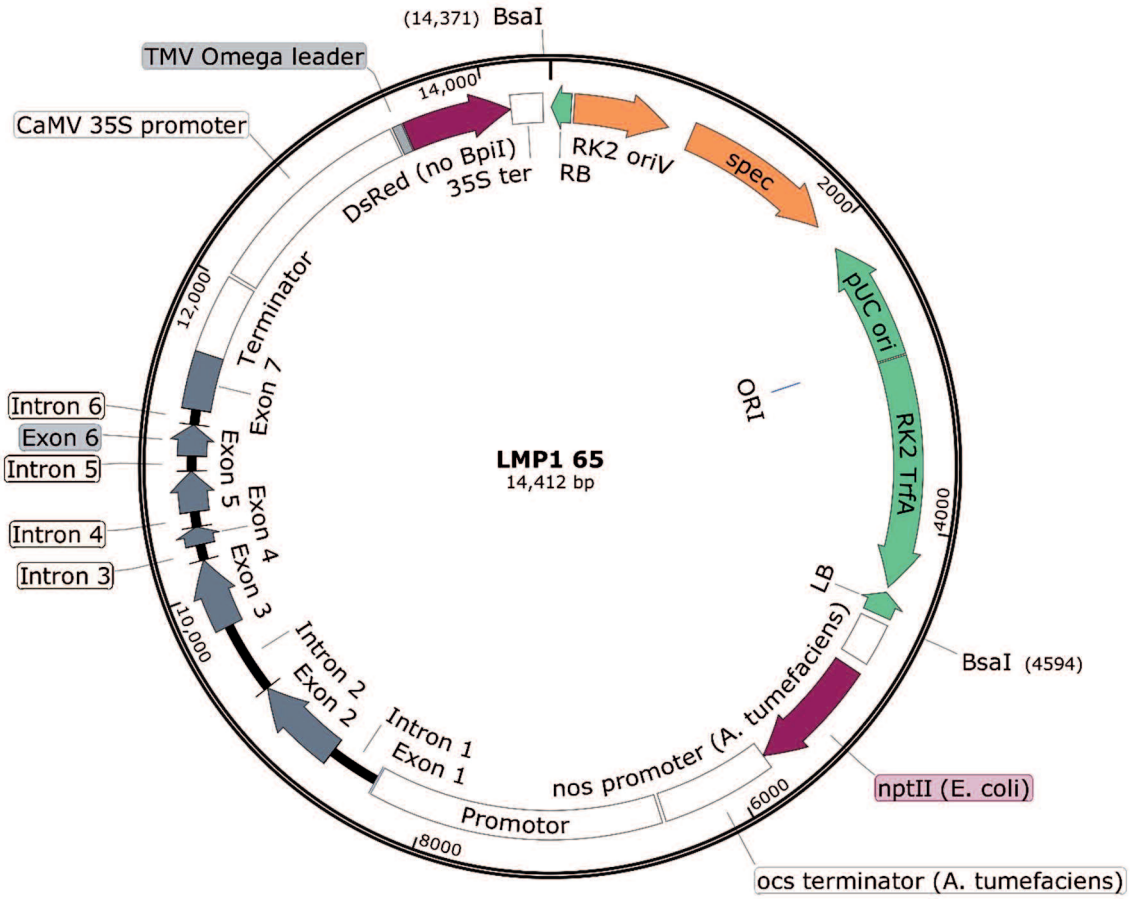


Figure 9.0.4 – LMP1_65
Level M Position 1 finished construct made from the negative antibiotic selectable marker pICSL11024, the genomic sequence of 7g116650 in its native context, and the positive visual selectable marker L1P3_DsRed.

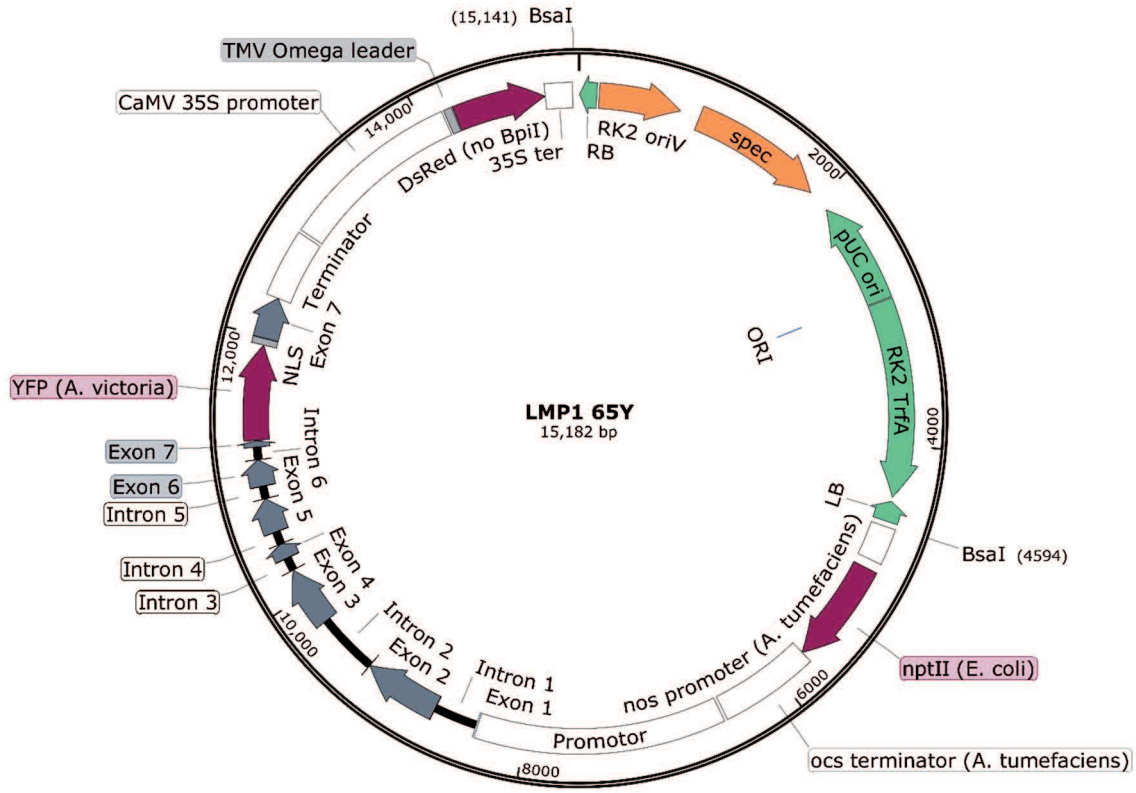


Figure 9.0.5 – LMP1_65Y
Level M Position 1 finished construct; as LMP1_65 with a YFP C terminal fusion added to 7g116650

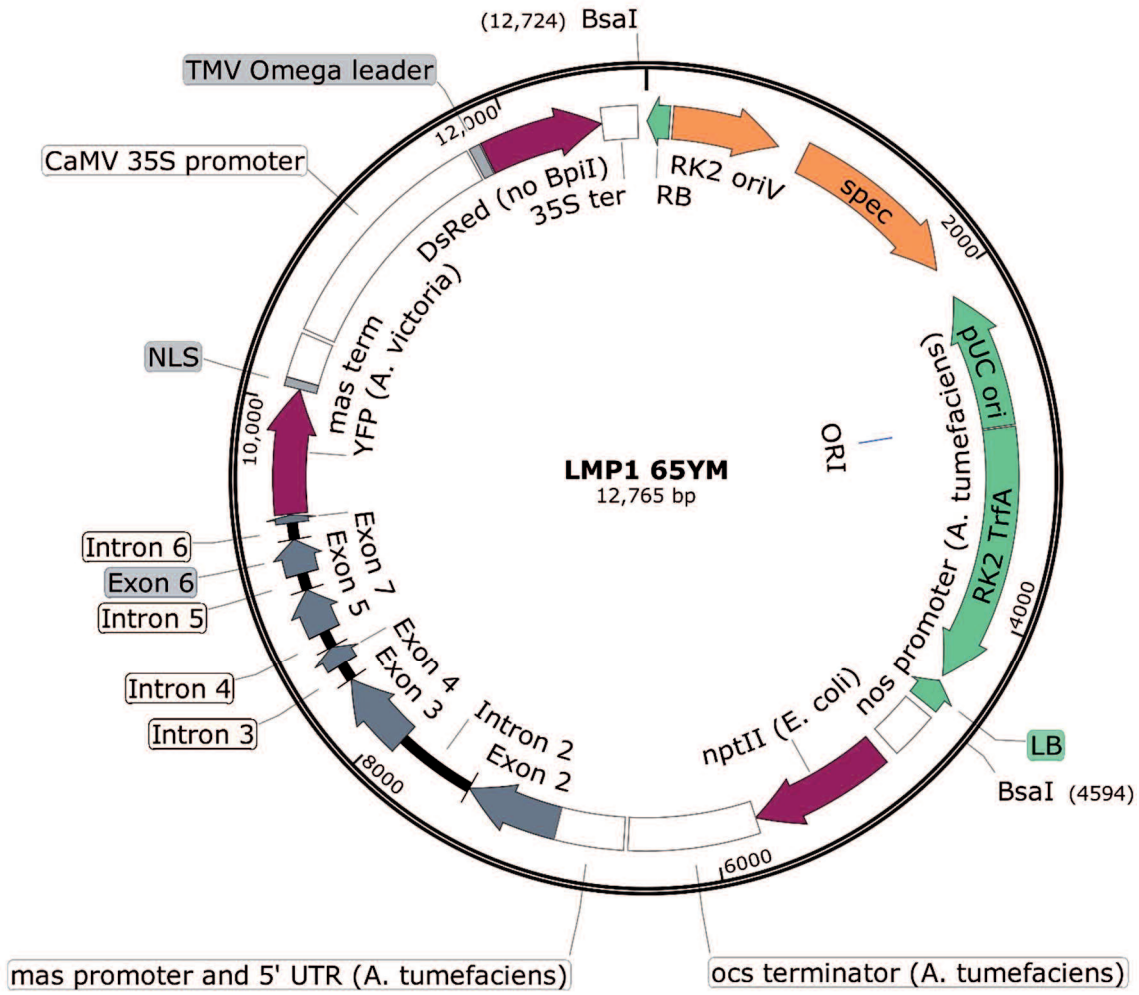


Figure 9.0.6 – LMP1_65YM

Level M Position 1 finished construct; as LMP1_65 with a YFP C terminal fusion added to 7g116650 and the native promoter, terminator and UTR replaced with those derived from the *Agrobacterium tumefaciens* mannopine synthase gene

Appendix 2 – Primers

A2.1 – Genotyping Primers

All genotyping was performed as a triplex PCR (see **Section 2.4.3**) with a pair of gene specific primers, and the tnt1 Universal Left promoter (below). A chromosome lacking the tnt1 insertion (hereafter 'WT') would give a band of one size, and a chromosome with the tnt1 insertion (hereafter 'Mut') would give a different sized band. A heterozygous plant would present a pair of bands on the agarose gel, and a homozygote a single band whose size would identify the genotype.

tnt1 Universal Left primer
GTATTTACCTCCGACCTACAAAGTG

Primers for genotyping NF443 (**Table 9.4**), NF3209 (**Table 9.5**), NF3438 (**Table 9.6**) and the Noble Foundation lines obtained for secondary screening of genes disrupted in NF443 (**Table 9.7**)

Table 9.4 – NF443 genotyping primers

inv-PCR derived primers				
Insertion	**Left Primer**	**Right Primer**	**WT size**	**Mut size**
1 (WGS i44)	GCCATACACAGAAAACTAAGCCATT	CCGTACCCATACCTATGCAACA	261	455
WGS derived primers				
Insertion	**Left Primer**	**Right Primer**	**WT size**	**Mut size**
i2	ACATTCAACCTTGACCTAATTGTG	ACCTATTCTTGCACCTCCCG	758	477
i4	AGCAGCTTCTTCTGGTGTTG	AATTCCCAGCTGCCACTAGC	935	673
i6	TCAACATCATCGTCCTCGGC	TGCTGCCACTATGCTCACG	627	294
i7	CGCGCGTTTGTTACCGATTG	AGCCCAGCACCAGAATTGTC	670	257
i10	TTGCATTTGTTCCTGTCGCC	GACAATCCACGGCACTCAAG	227	427
i13	AGCTCTAGTCGTGGAATTGCG	AAGGTGATGCAGAGTGGACG	600	263
i16	GCAGTGTATCCCACGAGTGC	CTGGCAAATGGAAGGCTTGTG	700	251
i19	CATACAGCAAGGCCAACAGC	GTGAGGGCGTGACTCATCTG	774	536
i20	CAACACTCACTTCTAGTTTCTTCTC	CGCGGTTATGTGGCTGTTG	450	631
i21	ATCCAGTGCGACGTCTGTAAC	GGAGAGTGCAGCTTGATCCC	650	512

i22	CGAGTACGCATGGCATTGAC	CTGGTGGCCCGACGATATG	536	351
i24	GATACCGGCATGTAGGGCTG	AGGGCTGGAAATGAGTCGAG	762	934
i25	GGTGAATGCATGAGCGGTTG	AACTGCAGCACCAACAAAGG	847	598
i26	GATACCGGCATGTAGGGCTG	AGGGCTGGAAATGAGTCGAG	762	934
i27	TCAAAGGAAAGGCACATTCGC	ATGCTTGCTCGGTAGGTGTC	934	326
i28	CACGTGACACCATCCCACTC	ACTGCAAGTTATTCGGGACC	900	795
i29	TTGAAGGTGCTGGACTGCTG	ATGACAGGTTGGGTGGGATG	632	346
i30	ACTCAACGACCAGCAAAGGG	GGACTCCACCTCTTGCAGTG	718	485
i33	AATGGCAGCTCACACCCTTC	AGATGTCGCAGAACCACGTC	960	310
i36	AGCTGCATGAATTATTGCCACTC	AGAGAAGGAGGGCGAGATCC	636	754
i36 v2	TTAGTTGTGTCGGCGTGTCC	ACGTGATTCCAAACCAGTTGC	743	353
i37	ACTAGACGCTCCAACTCCTTTAC	GGAAGAGACAACAATGCCGAG	333	466
i38	GCCTAGGGTTTCTCTGCCAC	GCAACTGGGTTATGTGCGTG	461	562
i39	AATATTTGCGATACGTCAAGTGG	TCAACAGAGGATGGAGAAGGAAC	752	642
i40	TTGTTAAATTGCAGAAAGAAATGGC	TCTGTTGTTTCCGCGACCAC	739	939
i41	TGGGCACAAGTTTGAGGCTG	GGGAGCTCGTAGGGTTCTTG	426	272
i42	CCCGTGCTTGATTTGGCAAC	GGGCATGGAAGAGAGCAAGG	421	522
i43	AGGATGGTCAGGACATACACAC	TGGTTCTCATACCTTGCCTGC	545	288
i43 v2	TGCAAATTTCCTGATAATACAAGGC	ATTGGTAGGCTGCAAGGTGG	296	496
i45	GGCGGAGAGTGTGGTGAAG	ACACAAGAACAAACCACAGCC	646	567

Table 9.5 – NF3209 genotyping primers

Inv-PCR derived primers				
Insertion	**Left Primer**	**Right Primer**	**WT size**	**Mut size**
1 (WSG i17)	ACCTGTAGCGACGGTCATTG	GCCACAACGATGATGATGCC	207	312
3 (WGS i10)	GCGTCGGACACGGGAATAAT	GCAAGTGATTAAGAGAGAGCGT	241	337
3 (WGS i10) v2	GGGAGCAGGGAATGGAGAAC	GCCGTTGTTGTTGATGAGCC	708	527
5 (WGS i26)	CAACTCCGTTAGTGCATGACC	CTTGTATGAGACAGTAGAAATGAGC	504	448

12 (WGS i8)	AGCATCATTATCAAACTCACCAGC	ACGCCACCCTTTGCAACTC	212	342
14	AGAAGCTGGTATGCGAGAGAG	TAAGACAAGGCGTCCCATGC	534	693
14 v2	TGGTCACAGTCCACATTGAAAC	ACGTTTACCTGATTGTCATACCC	572	652
15	AGTGGAGGAGAGTGTAATTCAAGC	GAGGCGGCTTAGCAGAAGAG	593	354
16 (WGS i3)	TTGTGGCCAAGATTATGCAAAC	ATTCCCTCCAACATCAAGCC	379	581
18	ACCTCACGATCAAACTCCCG	CCCTTGCACATGGTTCTTTCTC	503	695
23 (WGS i2)	CCCACCAAATACAAAGGTAGCC	AGCAATGGGTCCAAGGTGAC	434	533
24 (WGS i7)	TGCAGAGAGCATTGTTGAATGG	TTTGGAGTGACAAAGACGCC	463	562
27	GAAAGAAGTGTTAGAGAGGAGAAACTC	CCAAATCGTTGGGACAAGTG	490	336
28	CGTCAATGCATACGAAGCCG	TCACTGCCAATACCACCACG	314	426

WGS derived primers

Insertion	Left Primer	Right Primer	WT size	Mut size
i1	CTCCAAAGCTAGGTTGAAGCG	TGCTTGTCATGGAAGGGTGG	669	577
i2	GGATTGGAGGAGATGCTGGG	TTGGTGGTGCTGATCCAAGG	891	727
i3	CTTGGAACACCATGGAACCAC	ATTCCCTCCAACATCAAGCC	918	581
i6	ACTACTTTATGTGGTGCATACTTGC	GGTTAATGCAGATGGTGACCC	762	899
i7	GCCCAATTCACCCTCTTCAAC	TGGAGTGACAAAGACGCCAC	385	563
i8	ACGCCACCCTTTGCAACTC	TCTGCTACACTGCACCCAAC	329	442
i12	TCCCGCGGAATTAGGAAATC	TTGGTACCTCGCCTTCCAAC	656	336
i13	CTAACCCCTTCCGCTTTACC	ATGGATCCAACACAAGCACA	782	874
i13 v2	AGCATTTTGGTCCTCCATTG	GCAAATTAACATTTCATTTCCATC	528	264
i15	GTGGATCCGCAATGCATACTC	TCAACAGAGGATGGAGAAGGAAC	740	497
i15 v2	AGACTCTTTGCATCCACCACC	ACACCTTGAGGCATTCACCG	673	324
i17	CGGGTCTGAGGCAGTTGTAG	GGGAGATGGTTCTTCGACGG	571	412
i18	TCCACAGCTGCTCAAGTTAGTC	GCCGGGTTAAAGGGAAGGTC	441	543
i19	ACCAACACAGTTACACGAAGG	CTGCTACTAACCTGGTGCGG	401	587
i20	CCCACAATAAGCAACTACCTGTC	GAATGGTTTGGTGAAGGCAGC	204	353

i24	TCACCAAAGACGTAGCCTGC	GTTCCGGGTCGAATCCTGTG	734	444
i25	GGAGAAATGGAGTGTCCGGG	TGGTACCCAATCAATTGTAGTCTCC	657	260
i25 v2	TTTGTCGCTACCAACTGAGC	TGGTACCCAATCAATTGTAGTCTCC	606	260
i26	GTGGTAGGGTCATCGTCGTG	TGCTCTCTCAGAAATCTCCTTCC	952	502
i27	TTCTGTTGAAGGTGGAGCGG	AGGTGTGTGCAGTGTCCAAC	429	569
i27 v2	TTCTGTTGAAGGTGGAGCGG	GGTGTCGGCGTGTCAGTATC	269	409
i30	TGGTATCGCTTTGGTCCCTTAAC	CTGCTTGTTCCTCCGTCTCC	984	614
i31	ATGTCGGTGTCGTGTGGAAC	GGCACCTCTGTCGTCATCAG	729	530
i33	TGGAGGGTTGTGGAGCAAAC	CGTCCTACCATCATGCCACC	249	360
i35	CATGTGCAGATAAGGGCTTT	GCCCATAAATTGGAAGGATG	166	294
i38	GCTTCGAGGCTTCCCAGTTC	ACGCATCTCTGGTTCTGCTG	752	534
i39	AGTCGAGTCCAGACGAATCC	GGGCCTTGGTGTTAGAATGG	377	533
i40	CTTTGTCAGCTGGAATGTCCC	CAGGGTGAGTGTGAAGTGGG	225	262
i41	TCCATTCAGTGCAGGGTTGC	ATTCCTAATCGGCTGGCGTG	634	255
i42	TCGTTGCTGGAGTGCTTCTC	CCAAACCTCAATGTGCACCC	599	341
i43	TGTTCAACGTATTGAGCATGGC	TGCGCCATAGATACATAGATACGC	490	687
i44	TAAGCAGCACGTGAGGTAGG	TAAGCAGCACGTGAGGTAGG	408	257

Table 9.6 – NF3438 genotyping primers

inv-PCR derived primers				
Insertion	**Left Primer**	**Right Primer**	**WT size**	**Mut size**
9	GAAGCCTCCTAGGCCATTCAG	GGATCTCCTTGCCTATTGCCC	519	721
9 v2	AGCACAAATGGGTTGTTGTTCC	TGGAGATGAAGAATGCCAAGAGG	475	607
10 (WGS i4)	TGTGTCAAACTGTCAACTTCATTTC	ACAGCAAGGATTGGGTGTCAG	420	332
10 (WGS i4) v2	CTTCTTATCTTACCCTTCACTTCCTG	AGGTAACGTGCCCTTGTGAG	579	410
11 (WGS i6)	GGCTCCGTTTATGGTGGGAG	AGTATTTGAGCAAGTTCTAACACCC	518	399
12	AGGCATTCCTAAACCTGTCCAC	CCGTAGGACAAATCGTGCAG	402	303
12 v2	GGCCCTAGTAATCCAGAGTTCG	AGATACTGGCAAAGCGTGGG	555	756

WGS derived primers				
Insertion	**Left Primer**	**Right Primer**	**WT size**	**Mut size**
i2	AGTCAGCAGAGTGTCCACAAC	CTCTCCACCACCGTTTCTCC	718	907
i3	TCCCTCCTCTACCACCTTCC	CGTACAAAGTGTCTCATTGATGCC	942	768
i5	AATGAGGGTCTGCAACACGG	GTAGGAGTGGCAGTTGGGTG	412	319
i8	TCCTGGTGGTGGCATTGAAG	AACTGGTAATGCCCTCAGCG	307	497
i9	CAACCCAACATTGGTGAGAGAG	CGGGTTGCATGCCATTCTTG	565	310
i13	TAAGAAGCACTCTCGCACCC	TGTCTGTGTGCTTACCTCGC	573	753
i15	TATGGCCCGTCACCAAGAC	TACATGGGTCGTCCTGCAAC	400	501
i16	ACACTGAACTGAATCTTCTCTGC	GCGAAACTTCCCGATGAAACG	274	363
i21	TTTAACCTGCAGCGACTCTC	ATTTCTGTGCCGTCTCTGCG	516	721
i23	ACTGTTGAGTGTGAATTTGAGAGAG	GCTGACATCACCTAGCTCCC	412	602
i25	ATGTTCCGAATCTCCGGTGC	TTGAGGCTGAAGAGCGTGTC	541	463
i27	CCACTTGGGCATGCTTGATG	TGCCTCAAGGGAACTTTCGG	415	332
i28	GGATCCAAGTCATACTCCTGC	TGCGGTGAGTGAAGTGATGC	401	600
i34	TTCCCACTTCCATGTGGCGG	AGCCAATGTGGTCCGAAAGG	561	737

Table 9.7 – New NF line genotyping primers

Insertion	Left Primer	Right Primer	WT size
NF15212	ATAAACCTTGGACGCGGGAG	CCCGTGCTTGATTTGGCAAC	1053
NF2901	GCTCGCATGTTTGGAAGTGG	GTGCCGTAACATGACACACC	573
NF19713	ATAAACCTTGGACGCGGGAG	TCCGACAATGGACGCTTTCC	1051

A2.2 – GoldenGate Primers

Primers for producing PCR amplicons for Goldengate cloning (**Table 9.9**) for gene complementation (see **Chapter 6**) and primers for LightRun Sanger sequencing (GATC Biotech AG, Kolm, Germany, see **Table 9.8**) to confirm the GoldenGate plasmid sequence.

Table 9.8 – Primers for LightRun sequencing

Primer Name	Primer Sequence	Primer Name	Primer Sequence
51-LR-L1	GATTTTGAGAGACGACCCT	65-LR-L3	GCGACCAAAAGTAGTGATT
51-LR-L2	CCACCTATTCAAACACAACT	65-LR-L4	CTCTCTGCTGCAGAAATG
51-LR-L3	GTGTAATGATAAGCTTT	65-LR-L5	ATTAGTCTGCCATAGGCTC
51-LR-L4	CCTAGTCAACAGAGCACT	65-LR-L6	GCGCTTAAGTATCTCCATC
51-LR-L5	GGGCTTTGTACACCAGC	65-LR-L7	CGGTTTATGCAGAAGGC
51-LR-R1	TGTGTTCACGGGGACAG	65-LR-L8	GCAGAAGAGTCCAGATGG
65-LR-L1	CTCAGGAGGTGGCAGCA	65-LR-R1	GGAACAAAGCCTACTCCC
65-LR-L2	CAAGGTCAGCTCAATTTCC	65-LR-R2	AAAATCCGACAGTTTGGC

Terminology note: 51 and 65 refer to the genes Mt7g116510 and Mt7g116650 respectively, which are disrupted by insertions i41 and i42 in NF443.

Table 9.9 – Primers for producing the GoldenGate cassettes

Mt7g116510		
Primer Name	**Left Primer**	**Right Primer**
AdLig-1	GTGAAGACTACTCAGGAGCCTGTCTATACAAG AGTGACC	GTGAAGACGTCTCTCAAAATCAAATGACTC
AdLig-2	GTGAAGACTGAGAGACGACCCTCTTAAATGT	GTGAAGACGAAGGCATTTTATTACTGACAATTT G
AdLig-3	GTGAAGACGTGCCTTCCTTTGATTCAATCCTA	GTGAAGACTACTCGAGCGAGATCAGCTACAGA TTAG
AdLig-CDS_NS-R	GTGAAGACTACTCAAATGGCTACTTCTTTGGTG	GTGAAGACTACTCGCGAACCAGTCCTGGAATC TAAGGC
AdLig-CDS_Stop-R	GTGAAGACTACTCAAATGGCTACTTCTTTGGTG	GTGAAGACTACTCGAAGCTCAAGTCCTGGAAT CTAAGG
AdLig-Pro	GTGAAGACTACTCAGGAGCCTGTCTATACAAG AGTGACC	GTGAAGACTACTCGCATTGCAATGCAAAATAAT TGAG
AdLig-Ter	GTGAAGACTACTCAGCTTCAATCCATCCAAGAC ATTTG	GTGAAGACTACTCGAGCGAGATCAGCTACAGA TTAG
PreAmp	ACACTGGAGGAGGGTAGCTG	AGACAGCCTTCCTTGTCGTG
Mt7g116650		
Primer Name	**Left Primer**	**Right Primer**
AdLig-1	GTGAAGACTACTCAGGAGGTGGCAGCAACCTC TG	GTGAAGACGACGACAGTGACATCCCTGG
AdLig-2	GTGAAGACATGTCGTCATGTGGCTTTGATAAC	GTGAAGACTACTCGAGCGTTCACTTCTCTAACT TGCTG
AdLig-CDS_NS	GTGAAGACTACTCAAATGCCTTTTGGTGAATGT G	GTGAAGACTACTCGCGAACCGTTAGAGCTAGA TGTTGATG
AdLig-CDS_Stop	GTGAAGACTACTCAAATGCCTTTTGGTGAATGT G	GTGAAGACTACTCGAAGCTCAGTTAGAGCTAG ATGTTG
AdLig-Pro	GTGAAGACTACTCAGGAGGTGGCAGCAACCTC TG	GTGAAGACTACTCGCATTATATTCATTTTGGTA TGCATTG
AdLig-Ter	GTGAAGACTACTCAGCTTGTCCTCGACGAAGT GGT	GTGAAGACTACTCGAGCGTTCACTTCTCTAACT TGCTG
PreAmp-1	GTCAAGATGTGGATTGGTGCG	CGCAGGAGGTGGCAAGCAG
PreAmp-2	AGTGGCAGTGATGGTTTCCC	GGCTAATGCAAGATCATTGGCTC

Terminology notes: AdLig – Adapter Ligation; primers to add a Bpi1 recognition site and 4 bp overhang to insert the fragment into the universal level 0 acceptor plasmid pUAP.

- Numbered fragments – assemble into the gene in native context (2000 bp promoter and 500 bp terminator) with Bpi1/Bsa1 sites removed with the L0 pAUP.
- Pro/Ter/CDS – amplified from L0 plasmid produced by the numbered fragments to produce the promoter+5' UTR, 3' UTR+terminator and the coding sequence (with other native STOP codon or truncated to remove this, and replace it with a two amino acid (Gly-Ser) bridge.
- PreAmp – Pre-amplification; primer set to amplify the region from gDNA prior to adaptor ligation.

A2.3 – qPCR Primers

qPCR using primers in **Table 9.10** was used to assess the expression of certain genes (see **Chapter 3** & **6**). UBC9 and Helicase were chosen as housekeeping controls based off the recommendations by Kakar et al[293].

Table 9.10 – Primers for *Medicago* MYC gene qPCR

Gene ID	Gene Name	Left Primer	Right Primer	Product Size
Mt1g028600	PT4	TGCTTGTGCCATTTGCTCTG	CCACCAATACCAAATCCAAGCC	102
Mt4g088835	PHO2	TGCTTTCCTTGTCACAGCCA	CGAAGTGCTCTTCCACCAGT	86
Mt4g134790	Helicase	AGCAGAGACCAGCATAACAAT	AGGACGCATGAGCTTTTCAA	220
Mt7g027190	RAM1	TGCAAGCTCAACAACAGAGTC	TCACCTTTTGCCACTGCTTC	121
Mt7g116510	GDSL Lipase	ATCTTCGTTGCTGCTGATGC	TTGAGGCTGGTGTACAAAGC	139
Mt7g116650	Tyrosine Kinase	TGCAAGAGATCAGCTCACCAG	TCGCGAGTCTACAAAATCCG	79
Mt7g116940	UBC9	AGCCCCGCTTTGACAATATC	AGCAGTGGTCTCATACTTGGAC	135
Mt8g006790	HA1	GCGTGGAATCTGGTGGTAAAC	TTGTGCTTCTCTGCAAACCC	149

A2.4 – OTU-typing primers

Glomeromycotina OTU-typing (**Chapter 7**) was carried out based on rDNA10 sequence (**Table 9.11**).

Table 9.11 – Primers for OTU typing primers Glomeromycotina

Primer Name	Primer Sequence	Source
AM1	GTTTCCCGTAAGGCGCCGAA	Helgason et al, 1998[22]
AML1	ATCAACTTTCGATGGTAGGATAGA	Lee at al, 2008[391]
AML2	GAACCCAAACACTTTGGTTTCC	Lee at al, 2008
NS31	TTGGAGGGCAAGTCTGGTGCC	Helgason et al, 1998

Appendix 3 – Co-segregation graphs

No insertion discovered by invPCR (**Figure 9.7**) or WGS (**Figure 9.8**) of NF3209 segregated in the homozygous or heterozygous state with the mycorrhizal phenotype.

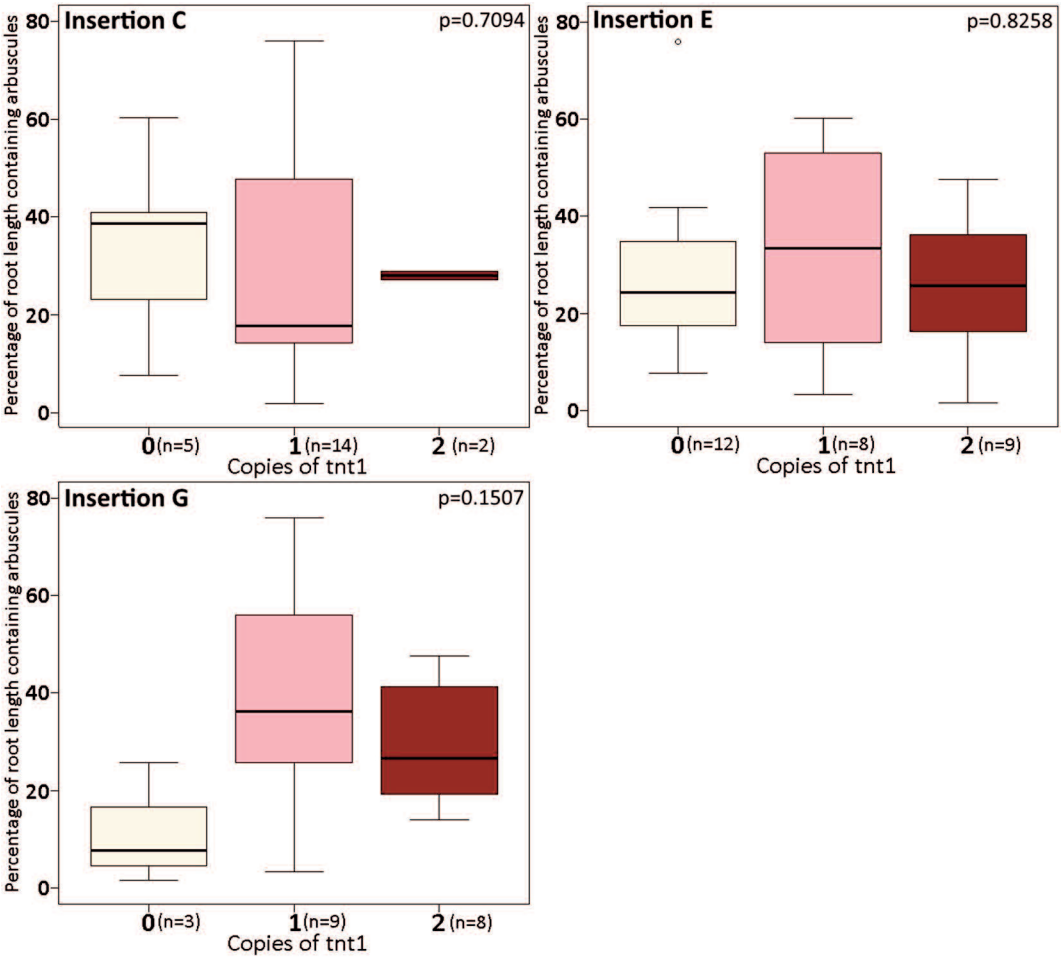
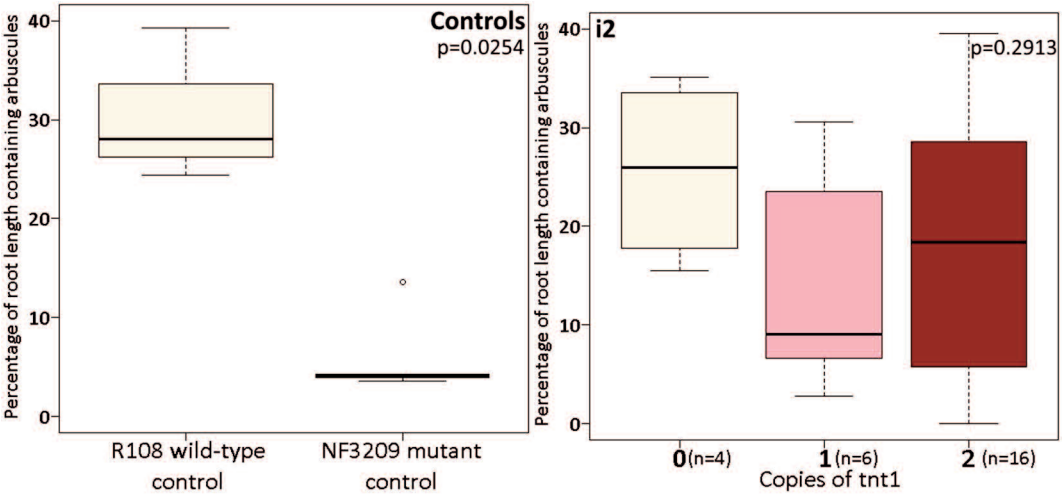
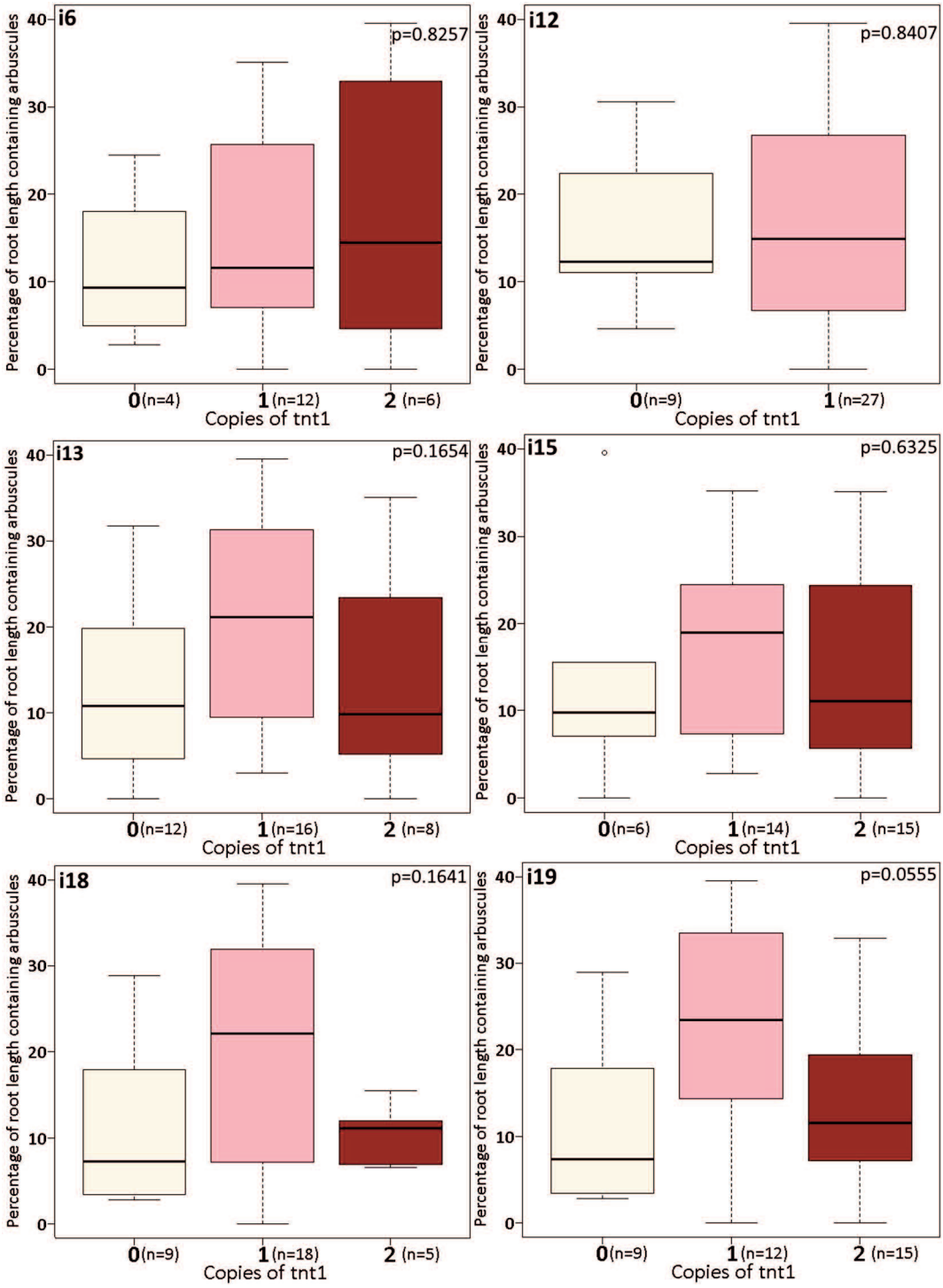
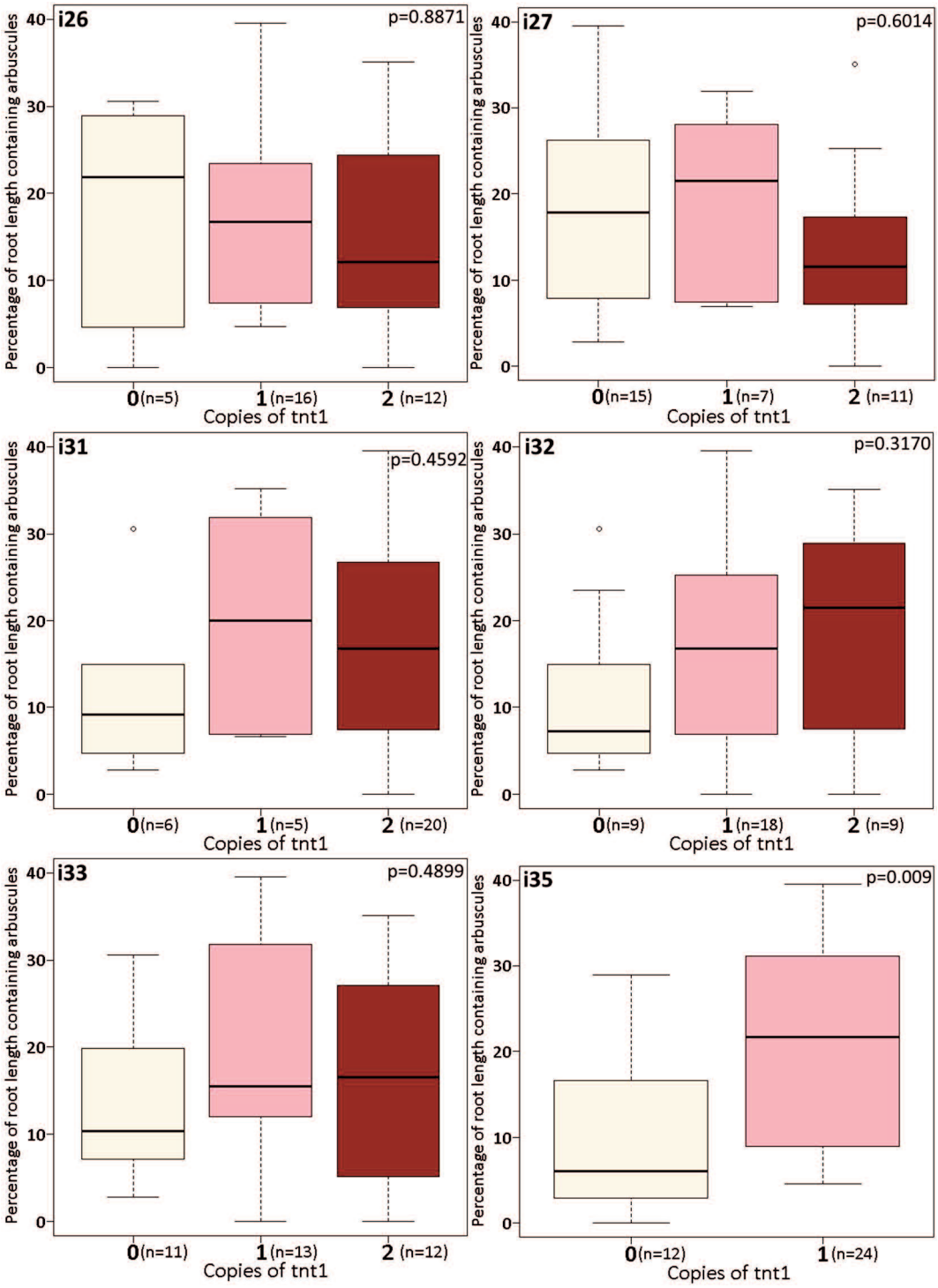


Figure 9.0.7 – NF3209 insertions that did not co-segregate with MYC phenotype (discovered by invPCR)

Insertions assessed in the July 2014 F_2 screen (**Section 5.4.1**). Box plot of the total AMF colonisation for plants containing 0 (cream), 1 (pink) or 2 (red) copies of each insertion (shown in the top left). All p values (shown in top right) obtained from Kruskal-Wallace testing.




i6
p=0.8257
Percentage of root length containing arbuscules
0 (n=4)
1 (n=12)
2 (n=6)
Copies of tnt1
i12
p=0.8407
0 (n=9)
1 (n=27)
Copies of tnt1
i13
p=0.1654
0 (n=12)
1 (n=16)
2 (n=8)
Copies of tnt1
i15
p=0.6325
0 (n=6)
1 (n=14)
2 (n=15)
Copies of tnt1
i18
p=0.1641
0 (n=9)
1 (n=18)
2 (n=5)
Copies of tnt1
i19
p=0.0555
0 (n=9)
1 (n=12)
2 (n=15)
Copies of tnt1


i26
p=0.8871
i27
p=0.6014
i31
p=0.4592
i32
p=0.3170
i33
p=0.4899
i35
p=0.009
Percentage of root length containing arbuscules
0 10 20 30 40
0 (n=5)
1 (n=16)
2 (n=12)
0 (n=15)
1 (n=7)
2 (n=11)
0 (n=6)
1 (n=5)
2 (n=20)
0 (n=9)
1 (n=18)
2 (n=9)
0 (n=11)
1 (n=13)
2 (n=12)
0 (n=12)
1 (n=24)
Copies of tnt1

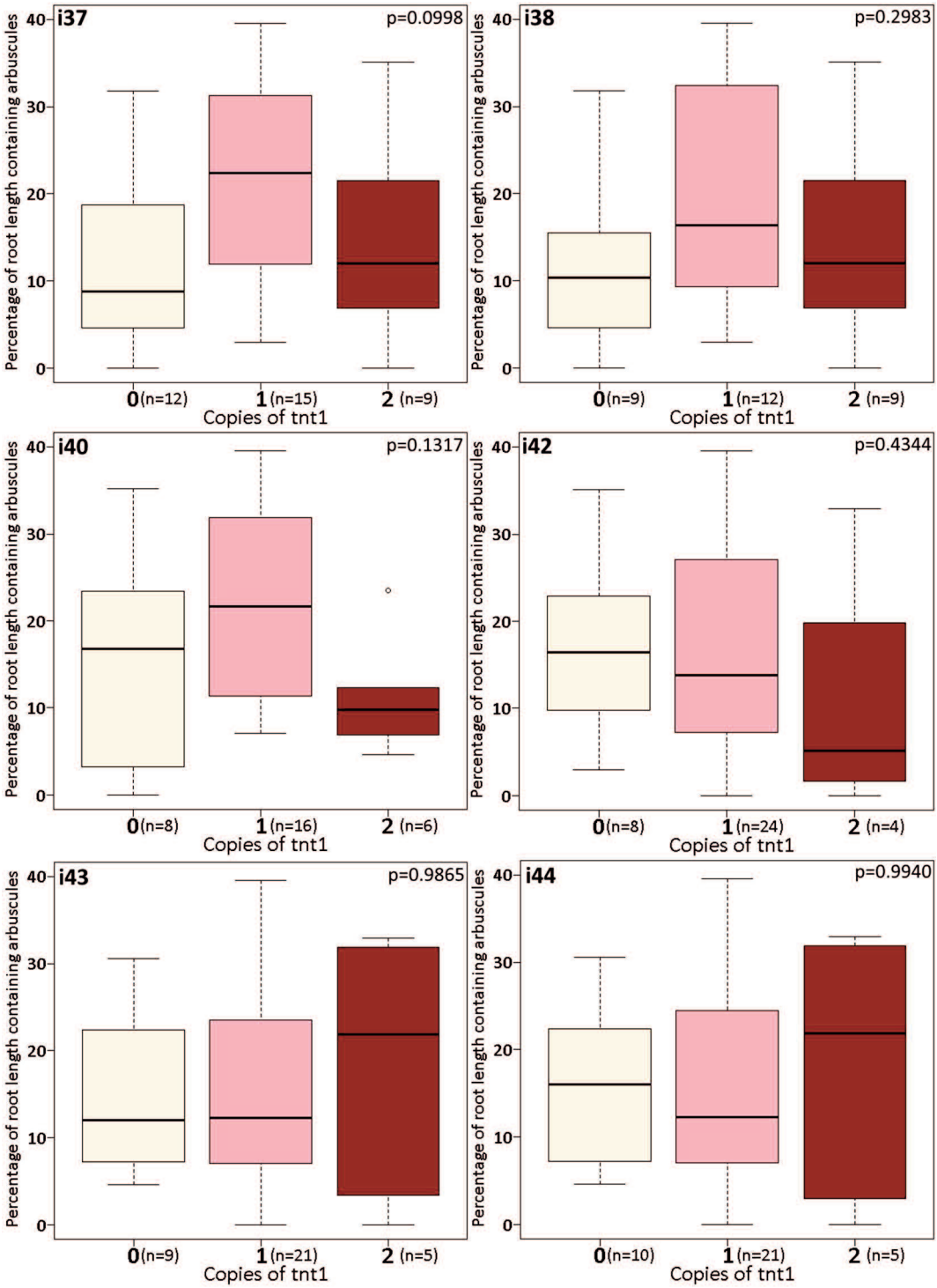


Figure 9.0.8 – NF3209 insertions that did not co-segregate with MYC phenotype (discovered by WGS) Insertions assessed in the June 2016 F_2 screen (**Section 5.4.3**). Box plot of the percentage of root length containing arbuscules for plants containing 0 (cream), 1 (pink) or 2 (red) copies of each insertion (shown in the top left). All p values (shown in top right) obtained from Kruskal-Wallace testing. The first box plot displays the colonisation of the R108 and NF3209 parental controls.

Appendix 4 – Media and Materials

A4.1 – Media, fertilisers and stock solutions

Table 9.12 lists solid growth media, **Table 9.13** lists fertilisers and their stock solutions, and **Table 9.14** lists stock chemical mixes used in the experiments described below.

Table 9.12 – Solid growth media

Media	Composition
Medicago Mix	50% F2 Compost (Everris, Suffolk, UK) 25% Vermiculite 25% Perlite
Sand:Terragreen Mix	80% Washed Terragreen (also called Biosorb or Turface) (Oil-Dri UK Ltd, UK) 20% Fine Sand
F Plates	2 ml F major salts 0.1 ml F minor salts 2.5 ml 20mM Fe-EDTA 15 g Agar Make up to 1 litre with diH_2O pH6.5 with KOH After autoclaving, add 1 ml of filter-sterilised 1M $CaCl_2$
LB Plates	10 g Bacto-tryptone 5 g Yeast extract 10 g NaCl 15 g Agar Make up to 1 litre with diH_2O pH7.5 with NaOH
M^{278} Plates	50 ml M major salts 10 ml M minor salts 3 g Phytagel Make up to 1 litre with diH_2O pH5.5 with KOH Add 10 g Sucrose for M+Suc
MW^{278} Plates	50 ml MW major salts 10 ml M minor salts 3 g Phytagel Make up to 1 litre with diH_2O pH5.5 with KOH Add 10 g Sucrose for MW+Suc
TY Plates	5 g Bacto-tryptone 3 g Yeast Extract 0.4 g $CaCl_2$ 15 g Agar Make up to 1 litre with diH_2O pH7.2 with NaOH
Water Agar Plates	10 g Agar Make up to 1 litre with diH_2O

Table 9.13 – Fertilisers, liquid media and stock solutions

Media	Composition	
LB Broth	10 g	Bacto-tryptone
	5 g	Yeast extract
	10 g	NaCl
		Make up to 1 litre with diH_2O
		pH7.5 with NaOH
		Autoclaved before using
R solution[278]	50 ml	R major salts
	10 ml	M minor salts
		Made up to 1 litre with diH_2O
		pH6.5 with KOH
R (no P) solution	50 ml	R (no P) major salts
	10 ml	M minor salts
		Made up to 1 litre with diH_2O
		pH6.5 with KOH
F major salts (500x)	14.4 g	KH_2PO_4
	14.2 g	Na_2HPO_4
	16.7 g	NH_4NO_3
	30.75 g	$MgSO_4.7H_2O$
		Made up to 250ml with diH_2O
F minor salts (10,000x)	112 mg	$MnSO_4.H_2O$
	156 mg	$CuSO_4.5H_2O$
	178 mg	$ZnSO_4.7H_2O$
	100 mg	H_2BO_3
	117 mg	$Na_2MoO_4.2H_2O$
		Made up to 100ml with diH_2O
		Stored at -20°C
Fe-EDTA (20mM)	1.4 g	$FeSO_4$
	1.85 g	Na_2EDTA
		Made up to 250ml with diH_2O & mixed at 50°C
		Stored at room temperature
M major salts (20x)	15 g	$MgSO_4.7H_2O$
	1.6 g	KNO_3
	1.3 g	KCl
	0.1 g	KH_2PO_4
	6 g	$Ca(NO_3)_2.4H_2O$
		Made up to 1 litre with diH_2O
MW major salts (20x)	15 g	$MgSO_4.7H_2O$
	9 g	$Na_2SO_4.10H_2O$
	1.6 g	KNO_3
	1.3 g	KCl
	0.4 g	KH_2PO_4
	6 g	$Ca(NO_3)_2.4H_2O$
		Made up to 1 litre with diH_2O
M minor salts (100x)	0.8 g	NaFeEDTA
	10 mg	KI
	40 mg	$MnCl_2.4H_2O$

	30 mg	$ZnSO_4.7H_2O$
	0.1 g	H_3BO_3
	10 mg	$CuSO_4.5H_2O$
	2 mg	$Na_2MoO_4.2H_2O$
		Made up to 1 litre with diH_2O
M Vitamins (200x)	60 mg	Glycine
	2 mg	Thiamine.HCl
	2 mg	Pyridoxin.HCl
	10 mg	Nicotinic acid
	1 g	Myo inositol
		Made up to 100 ml with diH_2O
R major salts (20x)	5 g	$MgSO_4.7H_2O$
	1.5 g	KCl
	0.3 g	KH_2PO_4
	9.5 g	$Ca(NO_3)_2.4H_2O$
		Made up to 1 litre with diH_2O
R (no P) major salts (20x)	5 g	$MgSO_4.7H_2O$
	1.67 g	KCl
	9.5 g	$Ca(NO_3)_2.4H_2O$
		Made up to 1 litre with diH_2O

All stock solutions are stored at 4°C unless otherwise noted.

Table 9.14 – Mixed solutions for experiments

Stock Solution	Composition	
FAA	100 ml	Ethanol
	10 ml	Acetic Acid
	20 ml	40% Formaldehyde solution
	70 ml	diH_2O
PBS	14.20 g	Na_2HPO_4
	2.45 g	KH_2PO_4
	80.07 g	NaCl
	2.01 g	KCl
		Made up to 1 litre with diH_2O

A4.2 – Genetic material

Plant lines used in the following experiments are listed in **Table 9.15**. Homozygous tnt1 lines were identified by PCR, or obtained from collaborators. Fungal and bacterial stocks are listed in **Table 9.16** and **9.17** respectively. Antibiotic concentrations used in growth media in each species are given in **Table 9.18**.

Table 9.15 – *M. truncatula* ecotypes and mutants used in this project

Line	Background	Description	Mutation	Source
R108	R108	Wild type	WT	Hoffman et al. (1997)[330]
A17	A17	Wild type	WT	Van den Bosch and Stacy (2003)[412]
NF443-3	R108	Reduced colonisation phenotype; *kin3*	tnt1	Samuel Roberts Noble Foundation[413], OK, USA
NF472-4	R108	No mycorrhizal phenotype	tnt1	SRNF
NF807	R108	Reduced colonisation phenotype; *ram1*	tnt1	SRNF
NF1436-1	R108	Reduced colonisation phenotype	tnt1	SRNF
NF3209-4	R108	Reduced colonisation phenotype	tnt1	SRNF
NF3438-7	R108	Reduced colonisation phenotype; *ram1*	tnt1	SRNF
NF905-3-3	R108	Reduced colonisation phenotype; *ha1*	tnt1	SRNF; backcrossed in Schultze Lab
NF905-3-5	R108	No mycorrhizal phenotype	tnt1	SRNF; backcrossed in Schultze Lab
EP10007	A17	Reduced colonisation phenotype; *dmi1*	T-DNA	Murray Lab
EP10013	A17	Reduced colonisation phenotype; *dmi2*	T-DNA	Murray Lab
EP10017	A17	Reduced colonisation phenotype; *dmi3*	T-DNA	Murray Lab
EP10058	A17	Reduced colonisation phenotype; *ram2*	T-DNA	Murray Lab
NF18637-10	R108	*Bfp1*	tnt1	SRNF; from Murray Lab
Daucus carota ssp. *sativus*		Hairy root culture		Helgason Lab
Allium ampeloprasum var. Albana		Mycorrhizae stock pot culture production		Suttons, Devon, UK
Allium schoenoprasum		Mycorrhizae stock pot culture production		B&Q, Renfrewshire, UK

Table 9.16 – Glomeromycotina lines used in this project

Name	Species	Strain	Description	Source
ROC Plates	*Rhizophagus irregularis*	DAOM-197198	Axenic culture grown with carrot hairy roots	Murry Lab
Stock Pot (*Rhizophagus*)	*Rhizophagus irregularis*	DAOM-197198	Grown with leek or chive	Bulked from Plate stocks
Stock Pot (four species mix)	*Claroideoglomus claroideum, Funneliformis mosseae, F. geosporum, Rhizophagus irregularis*	-	Grown with chive	Bulked from RootGrow Professional
RootGrow Professional	*Claroideoglomus claroideum, Funneliformis mosseae, F. geosporum, Rhizophagus irregularis* (all referred to as *Glomus* by the manufacturer)	-	Dried Terragreen and root matter.	Plantworks; commercial product

Table 9.17 – Bacterial strains used in this project

Strain	Species	Resistance	Description
DH5α	*Escherichia coli*	-	*E. coli* strain for plasmid amplification
ARqua1[414]	*Agrobacterium rhizogenes*	Streptomycin	For production of hairy root cultures and transformation of *M. truncatula* roots

Table 9.18 – Concentrations of antibiotics used in this project

Antibiotic	Concentration used		
	E. coli	***A. rhizogenes***	***M. truncatula* hairy roots**
Carbenicillin	100 µg/ml	-	-
Chloramphenicol	25 µg/ml	-	-
Kanamycin	50 µg/ml	-	25 µg/ml
Spectinomycin	50 µg/ml	150 µg/ml	-
Streptomycin	-	75 µg/ml	-

A4.3 – Growth Conditions

Table **9.19** lists the controlled growth conditions used in the various experiments.

Table 9.19 – Growth conditions used in this project

Location	Conditions
University Greenhouses	Heated to at least 18°C (day) or 14°C (night) Sequential cooling (vents open at 20°C, fans switch on at 22°C and chillers at 30°C) Supplementary lighting between 5 am and 9 pm if outside light level below 150 watts/m^2. Shades deploy if outside light level exceeds 850 watts/m^2 No humidity control (normally ~55%)
Growth Cabinet	16 hour days of 24°C and 300 µmol photons/m^2/second; 8 hour nights at 20°C
Growth Room	16 hour days of 21.5°C and 60-85 µmol photons/m^2/second; 8 hour nights at 19°C

Appendix 5 – Data Analysis

A5.1 – Linux scripts for analysis of the Illumina WGS dataset

Script for locating the tnt1 insertions in bulk Illumina read data

```
#!/bin/csh
#$ -S /bin/csh
bwa index tnt1.fasta
bwa mem -M -t 16 –T 20 –B 3 tnt1.fasta Illumina_Reads.fastq >
tnt1_Alignment.sam
samtools view -hSb -o tnt1_Alignment.bam tnt1_Alignment.sam
samtools view -hb -F 4 -o tnt1_Alignment_mapped.bam
tnt1_Alignment.bam
samtools sort -o tnt1_Alignment_sorted.bam
tnt1_Alignment_mapped.bam
samtools index tnt1_Alignment_sorted.bam
samtools view -h -o tnt1_Alignment_mapped.sam
tnt1_Alignment_mapped.bam
```

Script for tnt1 zygosity calling:

```
#!/bin/csh
#$ -S /bin/csh
bwa index Genome_Seq.fasta
bwa mem -M -t 16 –T 40 –O 10 –B 8 Genome_Seq.fasta
Illumina_Reads.fastq > Genome_Seq_Alignment.sam
samtools view -hSb -o Genome_Seq_Alignment.bam
Genome_Seq_Alignment.sam
samtools view -hb -F 4 -o Genome_Seq_Alignment_mapped.bam
Genome_Seq_Alignment.bam
samtools sort -o Genome_Seq_Alignment_sorted.bam
Genome_Seq_Alignment_mapped.bam
samtools index Genome_Seq_Alignment_sorted.bam
```

A5.2 – Optimisation of *bwa mem* functional parameters for accurate read mapping

bwa mem parameters can be varied by adding '–letter integer' to the code before the .fasta input. For defaults, do not attach the parameter. We used –t 16 and the default –r for all our alignments, but these may need to be changed depending on hardware. Similarly, –d would be important for long read lengths, but for Illumina data can be safely ignored. Adding '-H' (hard clipping) be used to reduce redundant output if this is desired. The most important parameters for optimising the output are those that control 'scoring'. This process gives a positive score for each matching base between the read and the provided sequence (–a, default 1), a negative score for each of the various sorts of mismatches, and only aligns those reads that reach a threshold (–T, default 30).

The mismatch parameters with their default score penalty. If any given non-matching region qualifies under multiple definitions, then it

–B 4 a non-matching base (a SNP relative to the provided sequence)

–U 9 a non-paired base (an insertion or deletion relative to the provided sequence)

–O 6 a gap where the read jumps over a region of the sequence. Add 1 to the penalty (–E) for every base in the gap, and the maximum gap size is limited by –w (default 100). As with –d, this latter parameter is not partially relevant to short Illumina reads.

For the initial alignment to locate FST by alignment to tnt1, a relaxed threshold is generally necessary to ensure that reads contain only small amounts of tnt1 sequence (e.g. 20 bp) are found. Parts of sequences mapping erroneously to tnt1 internal sequence are unimportant, as we will filter these out by selecting only reads that overlap the 3' or 5' ends of tnt1 from the output file. For alignments to examine zygosity and/or look for insertions in a gene of interest, we found that tightening up alignment stringency was often required, as the default parameters lead to related sequence (e.g. from regions of short repeats or closely related gene families) being aligned to many sequences. The exact combinations for the desired use will depend on the specific sequence and the data being aligned to it. If short alignments are needed, then –T should be kept low and the mismatch parameters increased, whereas full length alignment to a less related sequence (e.g. a gene homolog from another species) can be forced with a high –T and reduced mismatch parameters.

Appendix 6 – DNA Sequences

Sequence of tnt1 and genes used in **Chapter 4** to *in silico* test for already known mycorrhizal-related genes. All sequences are derived from *M. truncatula* unless otherwise marked (*Glycine max Gm*; *Pisum sativum Ps*).

Table 9.20 – DNA sequences used in this project

Sequence	NCBI reference
AMT1:4	XM_003590875.2
*Gm*AMT4:1	XM_003533638.2
LIN	EU926660.1
PsCRY	DQ845340.1
DMI2	EU306659.1
DMI3	XM_003628076.2
HA1	AJ132891.1
KPI106	XM_003628452.1
miR396a	TGCTTTTCCACAGCTTTCTTGAACTTCTTTCGTATCTTAAATCTGTTTTCAAGATTAAAAGTCCTAGAAGCTCAAGAAAGCTGTGGGAGAATA
miR396b	TATCTTCCCACAGCTTTATTGAACCGCAAAACAATGGCCATGGCAATGCAAAGGAACTCAATTTGGATGCAGTTCAAGAAAGCTGTGGAAGAATA
*Ps*LA	DQ848351.1
PT4	AY116211.1
RAM1	XM_003622047.1
RAM2	JN572682.1
SbtM1	XM_003611147.2
SbtM3	XM_003611142.2
SCP1	XM_003628010.1
STR	FJ659114.1
STR2	FJ659116.1
tnt1	X13777.1

Abbreviations

[?]	Concentration of substance '?'
2OH-FA	2-hydrozyl fatty acid
A17	An ecotype of *Medicago truncatula*
aa	Amino acid
ABA	Abscisic acid; plant hormone
ABC	ATP-binding cassette family transporters
AM	Arbuscular mycorrhiza
AMF	Arbuscular mycorrhizal fungi
AMS	Arbuscular mycorrhizal symbiosis
At	*Arabidopsis thaliana* (mouse ear cress)
BLAST	Basic local alignment search tool
C#	A fatty acid with a carbon chain length of #
C16:0	Saturated sixteen carbon fatty acid
C16:1ω5	Mono-unsaturated sixteen carbon fatty acid with the double bond on the 5^{th} carbon; a fatty acid configuration only found in fungi
CAZymes	Carbohydrate active enzyme
Chr	Chromosome
Chr#p	Short arm of chromosome #
Chr#q	Long arm of chromosome #
CMN	Common mycorrhizal network
CMV	Cauliflower mosaic virus
CO	Chitin oligomer
CO4	Chitin oligomer (chain length 3-5)
CO8	Chitin oligomer (chain length 6-9)
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats; a gene editing technique based on bacterial viral defence
CRRK	Cysteine rich receptor kinase
CSP	Common symbiosis pathway (sometimes called the CSSP; common symbiosis signalling pathway)
CTAB	Cetyltrimethylammonium bromide
DELLA	Family of DNA binding proteins characterised by a conserved DELLA amino acid motif that targets them for degradation by the GA-binding E3 Ub ligase GID1
diH_2O	Deionised water
dNTP	Deoxyribose nucleoside triphosphate; an equimolar mixture of the four bases (A/C/G/T)
dpi	Days post infection/inoculation
E3 Ub ligase	Family of enzymes that target molecules for addition of ubiquitin degradation tags by an E2 complex
EcMF	Ectomycorrhizal fungi
EcMS	Ectomycorrhizal symbiosis
EMS	Ethyl methanesulfonate
ER	Endoplasmic reticulum
ERM	Extraradical mycelium
ErMF	Ericoid mycorrhizal fungi
ErMS	Ericoid mycorrhizal symbiosis
F_0	Parental mutant plant for a backcrossed population

F_1	Intermediate step in a backcrossed population; a plant produced by crossing a mutant F_0 to the R108 wild type
F_2	Output plants from a backcrossed population; the offspring of the selfed F_1
FA	Fatty acid
FAA	Formalin–acetic acid–alcohol; sample fixation solution
FSP	Fungal-only signalling pathway
FST	Flanking sequence tag
GA	Gibberellic acid; plant hormone
gDNA	Genomic DNA
GDSL	Family of serine esterases/lipases characterised by a conserved GDSL amino acid motif
GlcNAc	N-acetylglucosamine; monomer of chitin
Gm	*Glycine max* (soybean)
GRAS	Gibberellic-acid insensitive/Repressor of GAI/Scarecrow family transcription factor
GSE	Germinated spore extract
HIGS	Host-induced gene silencing; a technique to knock down AMF genes via a RNAi construct expressed in the plant host and taken up by the symbiont
i#	tnt1 insertion #
IAA	Auxin; plant hormone
INDEL	Short insertion or deletion; a category of DNA mutations
invPCR	Inverse polymerase chain reaction
IRM	Intraradical mycelium
JA	Jasmonic acid; plant hormone
KO	Knock out
LCM	Laser capture microdissection; a technique to obtain nucleic acids from a single specific cell
LCO	Lipochitooligosaccharide
Lj	*Lotus japonicus* (birds foot trefoil)
LRR-RK	Leucine rich repeat receptor kinase
M	Minimal medium
MAMP	Microbe associated molecular pattern
MAPK	Mitogen-activated protein kinase; cytoplasmic signalling proteins involved in transmitting messages from the cell surface to the nucleus
MAPKK	A cytoplasmic kinase that phosphorylates a MAPK
MGR	Mycorrhizal growth response
MIR	Mycorrhizal induced resistance
miR#	Micro RNA #
Mt	*Medicago truncatula* (barrel medic)
MW	Modified White's medium
MYA	Million years ago
MYC/Myc	Relating to the mycorrhizal symbiosis
MYR	Mycorrhizal yield response
NF#	Noble Foundation *Medicago truncatula* tnt1 insertion mutant line
NHEJ	Non homologous end joining
NM	Non-mycorrhizal
NOD/Nod	Relating to the legume/*rhizobium* symbiosis

OMF	Orchid mycorrhizal fungi
OMS	Orchid mycorrhizal symbiosis
ORF	Open reading frame
OTU	Operational taxonomic unit
PAM	Periarbuscular matrix
PAMP	Pathogen associated molecular pattern
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PGPR	Plant growth promoting rhizobacteria
Pi	Phosphate ion (PO_4^{3-})
Ps	*Pisum sativum* (garden pea)
qPCR	Quantitative real time polymerase chain reaction
R	Rorison's nutrient solution
R108	An ecotype of *Medicago truncatula*
rDNA	DNA encoding ribosomal RNAs
RGR	Relative growth rate
Ri	*Rhizophagus irregularis*
RK	Receptor kinase
RNAseq	RNA sequencing
RNS	Root nodule symbiosis
SA	Salicylic acid; plant hormone
SL	Strigolactone; plant hormone
SNP	Single nucleotide polymorphism
T-DNA	*Agrobacterium* Transfer DNA; often used as a type of random insertion mutagenesis
TE	Transposable element (aka a transposon)
TF	Transcription factor
TMV	Tobacco mosaic virus
tnt1	Mobile retroviral-like transposable element from tobacco; often used as a type of random insertion mutagenesis
t-SNARE	Target soluble NSF attachment protein receptor; plasma membrane protein involved in vesicle fusion
VIGS	Virus-induced gene silencing; a technique to knock down genes by introducing a virus containing a fragment of a host gene to trigger host anti-virus defence mechanisms to degrade transcripts of the target gene
WGA	Wheat germ agglutinin (fungal cell wall binding compound, conjugated to fluorescent dye Alexa Fluor 488 nm)
WGS	Whole genome sequencing
WI	Wild arbuscular mycorrhizal fungal inoculum
wpi	Weeks post infection/inoculation
WT	Wild type
β-MAG	β-monoacylglycerol

Bibliography

1. Spatafora, J. W. *et al.* A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108,** 1028–1046 (2016).
2. Wang, B. & Qiu, Y.-L. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16,** 299–363 (2006).
3. Heckman, D. S. *et al.* Molecular evidence for the early colonization of land by fungi and plants. *Science* **293,** 1129–1133 (2001).
4. Chang, Y. *et al.* Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. *Genome Biol. Evol.* **7,** 1590–1601 (2015).
5. An, Z.-Q. & Hendrix, J. W. Determining viability of endogonaceous spores with a vital stain. *Mycologia* **80,** 259–261 (1988).
6. Kiers, T. E. *et al.* Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* **333,** 880–882 (2011).
7. Walder, F. & van der Heijden, M. Regulation of resource exchange in the arbuscular mycorrhizal symbiosis. *Nat. Plants* **1,** 15159 (2015).
8. Fitter, A. H. What is the link between carbon and phosphorus fluxes in arbuscular mycorrhizas? A null hypothesis for symbiotic function. *New Phytol.* **172,** 3–6 (2006).
9. Zhang, X. *et al.* The receptor kinase CERK1 has dual functions in symbiosis and immunity signalling. *Plant J.* **81,** 258–267 (2015).
10. Lauressergues, D. *et al.* The microRNA miR171h modulates arbuscular mycorrhizal colonization of *Medicago truncatula* by targeting NSP2. *Plant J.* **72,** 512–522 (2012).
11. Fellbaum, C. R. *et al.* Fungal nutrient allocation in common mycorrhizal networks is regulated by the carbon source strength of individual host plants. *New Phytol.* **203,** 646–656 (2014).
12. Kloppholz, S. *et al.* A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Curr. Biol.* **21,** 1204–1209 (2011).
13. Van der Heijden, M. G. *et al.* Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytol.* **205,** 1406–1423 (2015).
14. Lenoir, I. *et al.* Arbuscular mycorrhizal fungal responses to abiotic stresses: A review. *Phytochemistry* **123,** 4–15 (2016).
15. Schouteden, N. *et al.* Arbuscular mycorrhizal fungi for the biocontrol of plant-parasitic nematodes: a review of the mechanisms involved. *Front. Microbiol.* **6,** 1280 (2015).
16. Pozo, M. J. & Azcon-Aguilar, C. Unraveling mycorrhiza-induced resistance. *Curr. Opin. Plant Biol.* **10,** 393–398 (2007).
17. Garcia, K. & Zimmermann, S. D. The role of mycorrhizal associations in plant potassium nutrition. *Front. Plant Sci.* **5,** 1–9 (2014).
18. Bücking, H. & Kafle, A. Role of arbuscular mycorrhizal fungi in the nitrogen uptake of plants: current knowledge and research gaps. *Agronomy* **5,** 587–612 (2015).
19. Babikova, Z. *et al.* Underground allies: How and why do mycelial networks help plants defend themselves?: What are the fitness, regulatory, and practical implications of defence-related signaling between plants via common mycelial networks? *BioEssays* **36,** 21–26 (2014).
20. Hodge, A. *et al.* Nutritional ecology of arbuscular mycorrhizal fungi. *Fungal Ecol.*

3, 267–273 (2010).
21. Bücking, H. *et al.* Common mycorrhizal networks and their effect on the bargaining power of the fungal partner in the arbuscular mycorrhizal symbiosis. *Commun. Integr. Biol.* **9,** 1–4 (2016).
22. Helgason, T. *et al.* Ploughing up the wood-wide web? *Nature* **394,** 431 (1998).
23. Cooper, J. *et al.* The future distribution and production of global phosphate rock reserves. *Resour. Conserv. Recycl.* **57,** 78–86 (2011).
24. Van Kauwenbergh, S. J. *et al.* World reserves of phosphate rock… a dynamic and unfolding story. *Better Crop.* **97,** 18–20 (2013).
25. Herrera-Estrella, L. & López-Arredondo, D. Phosphorus: the underrated element for feeding the world. *Trends Plant Sci.* **21,** 461–463 (2016).
26. Hijri, M. Analysis of a large dataset of mycorrhiza inoculation field trials on potato shows highly significant increases in yield. *Mycorrhiza* **26,** 209–214 (2016).
27. Remy, W. *et al.* Four hundred-million-year-old vesicular arbuscular mycorrhizae. *PNAS* **91,** 11841–11843 (1994).
28. Taylor, T. N. *et al.* Fossil arbuscular mycorrhizae from the early Devonian. *Mycologia* **87,** 560–573 (1995).
29. Strullu-Derrien, C. *et al.* Fungal associations in *Horneophyton ligneri* from the Rhynie Chert (c. 407 million year old) closely resemble those in extant lower land plants: Novel insights into ancestral plant-fungus symbioses. *New Phytol.* **203,** 964–979 (2014).
30. Bortolazzo, A. *et al.* Mycorrhizal signaling in *Physcomitrella patens*: what can we learn about mycorrhizal associations from non-mycorrhizal lineages? in *IMMM3* (2017).
31. Kamel, L. *et al.* Biology and evolution of arbuscular mycorrhizal symbiosis in the light of genomics. *New Phytol.* **213,** 531–536 (2017).
32. Field, K. J. *et al.* Symbiotic options for the conquest of land. *Trends Ecol. Evol.* **30,** 477–486 (2015).
33. Garcia, K. *et al.* Molecular signals required for the establishment and maintenance of ectomycorrhizal symbioses. *New Phytol.* **208,** 79–87 (2015).
34. Hocher, V. *et al.* Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. *Plant Physiol.* **156,** 700–711 (2011).
35. Banhara, A. *et al.* Colonization of root cells and plant growth promotion by *Piriformospora indica* occurs independently of plant common symbiosis genes. *Front. Plant Sci.* **6,** 667 (2015).
36. Galibert, F. *et al.* The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* **293,** 668–672 (2001).
37. Tisserant, E. *et al.* Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *PNAS* **110,** 20117–20122 (2013).
38. Oláh, B. *et al.* Nod factors and a diffusible factor from arbuscular mycorrhizal fungi stimulate lateral root formation in *Medicago truncatula* via the DMI1/DMI2 signalling pathway. *Plant J.* **44,** 195–207 (2005).
39. Maillet, F. *et al.* Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature* **469,** 58–63 (2011).
40. Genre, A. *et al.* Short-chain chitin oligomers from arbuscular mycorrhizal fungi trigger nuclear Ca^{2+} spiking in *Medicago truncatula* roots and their production is enhanced by strigolactone. *New Phytol.* **198,** 179–189 (2013).

41. Feng, F. *et al.* The crosstalk between chitin-induced defences and symbiosis. in *MPMI16* (2014).
42. D'Haeze, W. & Holsters, M. Nod factor structures, responses, and perception during initiation of nodule development. *Glycobiology* **12,** 79–105 (2002).
43. Arrighi, J.-F. *et al.* The *Medicago truncatula* lysin motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. *Plant Physiol.* **142,** 265 (2006).
44. Czaja, L. F. *et al.* Transcriptional responses toward diffusible signals from symbiotic microbes reveal MtNFP- and MtDMI3-dependent reprogramming of host gene expression by arbuscular mycorrhizal fungal lipochitooligosaccharides. *Plant Physiol.* **159,** 1671–1685 (2012).
45. Fliegmann, J. *et al.* LYR3, a high-affinity LCO-binding protein of *Medicago truncatula*, interacts with LYK3, a key symbiotic receptor. *FEBS Lett.* **590,** 1477–1487 (2016).
46. Antolín-Llovera, M. *et al.* Knowing your friends and foes - plant receptor-like kinases as initiators of symbiosis or defence. *New Phytol.* **204,** 791–802 (2014).
47. Op den Camp, R. *et al.* LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. *Science* **331,** 909–912 (2011).
48. Limpens, E. *et al.* LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. *Science* **302,** 630–633 (2003).
49. Roberts, N. J. *et al.* Rhizobial and mycorrhizal symbioses in *Lotus japonicus* require lectin nucleotide phosphohydrolase, which acts upstream of calcium signaling. *Plant Physiol.* **161,** 556–567 (2013).
50. Popp, C. PhD Thesis: Analysis of the membrane binding mechanism of Remorins and their role in beneficial endosymbioses. (2017).
51. Lefebvre, B. *et al.* A remorin protein interacts with symbiotic receptors and regulates bacterial infection. *PNAS* **107,** 2343–2348 (2010).
52. Kevei, Z. *et al.* 3-hydroxy-3-methylglutaryl coenzyme a reductase 1 interacts with NORK and is crucial for nodulation in *Medicago truncatula*. *Plant Cell* **19,** 3974–3989 (2007).
53. Chen, T. *et al.* A MAP kinase kinase interacts with SymRK and regulates nodule organogenesis in Lotus japonicus. *Plant Cell* **24,** 823–838 (2012).
54. Charpentier, M. *et al.* Nuclear-localized cyclic nucleotide-gated channels mediate symbiotic calcium oscillations. *Science* **352,** 1102–1105 (2016).
55. Ané, J.-M. *et al. Medicago truncatula* DMI1 required for bacterial and fungal symbioses in legumes. *Science* **303,** 1364–1367 (2004).
56. Capoen, W. *et al.* Nuclear membranes control symbiotic calcium signaling of legumes. *PNAS* **108,** 14348–14353 (2011).
57. Kodavali, P. K. *et al.* Structural and functional characterization of annexin 1 from *Medicago truncatula*. *Plant Physiol. Biochem.* **73,** 56–62 (2013).
58. Groth, M. *et al.* NENA, a Lotus japonicus homolog of Sec13, is required for rhizodermal infection by arbuscular mycorrhiza fungi and rhizobia but dispensable for cortical endosymbiotic development. *Plant Cell* **22,** 2509–2526 (2010).
59. Saito, K. *et al.* NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in *Lotus japonicus*. *Plant Cell* **19,** 610–624 (2007).
60. Kanamori, N. *et al.* A nucleoporin is required for induction of Ca^{2+} spiking in legume nodule development and essential for rhizobial and fungal symbiosis.

PNAS **103,** 359–364 (2006).
61. Messinese, E. *et al.* A novel nuclear protein interacts with the symbiotic DMI3 calcium- and calmodulin-dependent protein kinase of *Medicago truncatula*. *Mol. Plant-Microbe Interact.* **20,** 912–921 (2007).
62. Oldroyd, G. E. D. & Downie, J. A. Nuclear calcium changes at the core of symbiosis signalling. *Curr. Opin. Plant Biol.* **9,** 351–357 (2006).
63. Hirsch, S. *et al.* GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in *Medicago truncatula*. *Plant Cell* **21,** 545–557 (2009).
64. Pimprikar, P. *et al.* A CCaMK-CYCLOPS-DELLA complex activates transcription of RAM1 to regulate arbuscule branching. *Curr. Biol.* **26,** 987–998 (2016).
65. Sun, J. *et al.* Activation of symbiosis signaling by arbuscular mycorrhizal fungi in legumes and rice. *Plant Cell* **27,** 823–838 (2015).
66. Gutjahr, C. *et al.* Rice perception of symbiotic arbuscular mycorrhizal fungi requires the karrikin receptor complex. *Science* **350,** 1521–1524 (2015).
67. Nadal, M. *et al.* An N-acetylglucosamine transporter required for arbuscular mycorrhizal symbioses in rice and maize. *Nat. Plants* **3,** 17073 (2017).
68. Delaux, P. *et al.* Evolution of the plant – microbe symbiotic ' toolkit '. *Trends Plant Sci.* **18,** 298–304 (2013).
69. Delaux, P. & Guillaume, B. NSP1 is a component of the Myc signaling pathway. *New Phytol.* **199,** 59–65 (2013).
70. Hohnjec, N. *et al.* Pre-announcement of symbiotic guests: transcriptional reprogramming by mycorrhizal lipochitooligosaccharides shows a strict co-dependency on the GRAS transcription factors NSP1 and RAM1. *BMC Genomics* **16,** 994 (2015).
71. Floss, D. S. *et al.* DELLA proteins regulate expression of a subset of AM symbiosis-induced genes in *Medicago truncatula*. *Plant Signal. Behav.* **11,** e1162369 (2016).
72. Gobbato, E. *et al.* A GRAS-type transcription factor with a specific function in mycorrhizal signaling. *Curr. Biol.* **22,** 2236–2241 (2012).
73. Floss, D. S. *et al.* A transcriptional program for arbuscule degeneration during AM symbiosis is regulated by MYB1. *Curr. Biol.* **27,** 1206–1212 (2017).
74. Bucher, M. *et al.* Evolutionary conservation of mycorrhiza-specific phosphate transporter gene regulation in *Lotus japonicus* by CTTC MOTIF-BINDING TRANSCRIPTION FACTOR1 (CBX1). in *ICOM9* (2017).
75. Oldroyd, G. E. D. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature* **11,** 252–263 (2013).
76. Daniels, B. A. & Trappe, J. M. Factors affecting spore germination of the vesicular-arbuscular mycorrhizal fungus, *Glomus epigaeus*. *Mycologia* **72,** 457–471 (1980).
77. Azcón, R. Germination and hyphal growth of *Glomus mosseae in vitro*: Effects of rhizophere bacteria and cell-free culture media. *Soil Biol. Biochem.* **19,** 417–419 (1987).
78. Azcon-Aguilar, C. *et al.* Effect of soil micro-organisms on spore germination and growth of the vesicular-arbuscular mycorrhizal fungus *Glomus mosseae*. *Trans. Br. Mycol. Soc.* **86,** 337–340 (1986).
79. Becard, G. *et al.* Extensive *in vitro* hyphal growth of vesicular-arbuscular mycorrhizal fungi in the presence of CO_2 and flavonols. *Appl. Environ. Microbiol.* **58,** 821–825 (1992).

80. Akiyama, K. *et al.* Isolation and identification of a phosphate deficiency-induced C-glycosylflavonoid that stimulates arbuscular mycorrhiza formation in melon roots. *MPMI* **15,** 334–340 (2002).
81. Hassan, S. & Mathesius, U. The role of flavonoids in root–rhizosphere signalling: opportunities and challenges for improving plant–microbe interactions. *J. Exp. Bot.* **63,** 3429–3444 (2012).
82. Becard, G. *et al.* Flavonoids are not necessary plant signal compounds in arbuscular mycorrhizal symbioses. *MPMI* **8,** 252–258 (1995).
83. Nagahashi, G. & Douds, D. D. The effects of hydroxy fatty acids on the hyphal branching of germinated spores of AM fungi. *Fungal Biol.* **115,** 351–358 (2011).
84. Trépanier, M. *et al.* Dependence of arbuscular-mycorrhizal fungi on their plant host for palmitic acid synthesis. *Appl. Environ. Microbiol.* **71,** 5341–5347 (2005).
85. Jiang, Y. *et al.* Plants transfer lipids to sustain colonization by mutualistic mycorrhizal and parasitic fungi. *Science* **356,** 1172–1175 (2017).
86. Luginbuehl, L. H. *et al.* Fatty acids in arbuscular mycorrhizal fungi are synthesized by the host plant. *Science* **356,** 1175–1178 (2017).
87. Torrecillas, E. *et al.* Host preferences of arbuscular mycorrhizal fungi colonizing annual herbaceous plant species in semiarid mediterranean prairies. *Appl. Environ. Microbiol.* **78,** 6180–6186 (2012).
88. Lopez-Obando, M. *et al.* Strigolactone biosynthesis and signaling in plant development. *Development* **142,** 3615–3619 (2015).
89. Kretzschmar, T. *et al.* A petunia ABC protein controls strigolactone-dependent symbiotic signalling and branching. *Nature* **483,** 341–344 (2012).
90. López-Ráez, J. A. *et al.* Tomato strigolactones are derived from carotenoids and their biosynthesis is promoted by phosphate starvation. *New Phytol.* **178,** 863–874 (2008).
91. Koltai, H. Review: Implications of non-specific strigolactone signaling in the rhizosphere. *Plant Sci.* **225,** 9–14 (2014).
92. Akiyama, K. *et al.* Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* **435,** 824–827 (2005).
93. Wang, E. *et al.* A common signaling process that promotes mycorrhizal and oomycete colonization of plants. *Curr. Biol.* **22,** 2242–2246 (2012).
94. Shinya, T. *et al.* Chitin-mediated plant–fungal interactions: catching, hiding and handshaking. *Curr. Opin. Plant Biol.* **26,** 64–71 (2015).
95. Stracke, S. *et al.* A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature* **417,** 959–962 (2002).
96. Bras, P. *et al.* A receptor kinase gene regulating symbiotic nodule development. *Nature* **417,** 962–966 (2002).
97. Antolín-Llovera, M. *et al.* Cleavage of the symbiosis receptor-like kinase ectodomain promotes complex formation with nod factor receptor 5. *Curr. Biol.* **24,** 422–427 (2014).
98. Madsen, E. B. *et al.* Autophosphorylation is essential for the *in vivo* function of the *Lotus japonicus* Nod factor receptor 1 and receptor-mediated signalling in cooperation with Nod factor receptor 5. *Plant J.* **65,** 404–417 (2011).
99. Fliegmann, J. *et al.* Lipo-chitooligosaccharidic symbiotic signals are recognized by LysM receptor-like kinase LYR3 in the legume *Medicago truncatula*. *ACS Chem. Biol.* **8,** 1900–1906 (2013).
100. Gaude, N. *et al.* Arbuscule-containing and non-colonized cortical cells of mycorrhizal roots undergo extensive and specific reprogramming during

arbuscular mycorrhizal development. *Plant J.* **69,** 510–28 (2012).
101. Gutjahr, C. & Paszkowski, U. Multiple control levels of root system remodeling in arbuscular mycorrhizal symbiosis. *Front. Plant Sci.* **4,** 204 (2013).
102. Carotenuto, G. *et al.* The rice LysM receptor-like kinase OsCERK1 is required for the perception of short-chain chitin oligomers in arbuscular mycorrhizal signaling. *New Phytol.* **214,** 1440–1446 (2017).
103. Nadal, M. & Paszkowski, U. Polyphony in the rhizosphere: presymbiotic communication in arbuscular mycorrhizal symbiosis. *Curr. Opin. Plant Biol.* **16,** 473–479 (2013).
104. Gao, C. *et al.* Arbuscular mycorrhizal fungal abundance and composition shift over life cycle of *Sorghum*, differ among root, rhizosphere and soil, and were affected by drought. in *ICOM9* (2017).
105. Dumbrell, A. J. *et al.* Distinct seasonal assemblages of arbuscular mycorrhizal fungi revealed by massively parallel pyrosequencing. *New Phytol.* **190,** 794–804 (2011).
106. Sergeevich, S. A. *et al.* in *Plants Futur.* (Bocchiaro, P. & Zamperini, A.) 135–160 (INTECH, 2016). doi:10.5772/711
107. Hogg, B. V *et al.* The DMI1 and DMI2 early symbiotic genes of *Medicago truncatula* are required for a high-affinity nodulation factor-binding site associated to a particulate fraction of roots. *Plant Physiol.* **140,** 365–373 (2006).
108. Genre, A. *et al.* Arbuscular mycorrhizal fungi elicit a novel intracellular apparatus in *Medicago truncatula* root epidermal cells before infection. *Plant Cell* **17,** 3489–3499 (2005).
109. Zhang, C. *et al.* Plant vacuole morphology and vacuolar trafficking. *Front. Plant Sci.* **5,** 476 (2014).
110. Genre, A. *et al.* Multiple exocytotic markers accumulate at the sites of perifungal membrane biogenesis in arbuscular mycorrhizas. *Plant Cell Physiol.* **53,** 244–255 (2012).
111. Pumplin, N. *et al. Medicago truncatula* Vapyrin is a novel protein required for arbuscular mycorrhizal symbiosis. *Plant J.* **61,** 482–494 (2010).
112. Huisman, R. *et al.* A symbiosis-dedicated SYNTAXIN OF PLANTS 13II isoform controls the formation of a stable host-microbe interface in symbiosis. *New Phytol.* **211,** 1338–1351 (2016).
113. Ivanov, S. *et al.* Rhizobium – legume symbiosis shares an exocytotic pathway required for arbuscule formation. *PNAS* **109,** 8316–8321 (2012).
114. Balestrini, R. & Bonfante, P. The interface compartment in arbuscular mycorrhizae: A special type of plant cell wall? *Plant Biosyst.* **139,** 8–15 (2005).
115. Zhang, F. *et al.* Purification and characterization of a symbiosis-induced endocellulase from the ectomycorrhizal symbiont *Laccaria bicolor*. in *IMMM3* (2017).
116. Veneault-Fourrey, C. *et al.* Genomic and transcriptomic analysis of *Laccaria bicolor* CAZyme reveals insights into polysaccharides remodelling during symbiosis establishment. *Fungal Genet. Biol.* **72,** 168–181 (2014).
117. Martin, F. M. Evolutionary biology of mycorrhizal symbioses in the age of genomics. in *IMMM3* (2017).
118. Kamel, L. *et al.* The comparison of expressed candidate secreted proteins from two arbuscular mycorrhizal fungi unravels common and specific molecular tools to invade different host plants. *Front. Plant Sci.* **8,** 124 (2017).
119. Le Marquer, M. *et al.* Study of the role of secreted peptides by *Rhizophagus*

irregularis in the establishment of the arbuscular endomycorrhizal symbiosis. in *IMMM3* (2017).
120. Takeda, N. *et al.* Apoplastic plant subtilases support arbuscular mycorrhiza development in *Lotus japonicus*. *Plant J.* **58,** 766–777 (2009).
121. Kiirika, L. M. *et al.* Silencing of the Rac1 GTPase MtROP9 in *Medicago truncatula* stimulates early mycorrhizal and oomycete root colonizations but negatively affects rhizobial infection. *Plant Physiol.* **159,** 501–516 (2012).
122. Arthikala, M.-K. *et al.* RbohB, a *Phaseolus vulgaris* NADPH oxidase gene, enhances symbiosome number, bacteroid size, and nitrogen fixation in nodules and impairs mycorrhizal colonization. *New Phytol.* **202,** 886–900 (2014).
123. Takeda, N. *et al.* CERBERUS and NSP1 of *Lotus japonicus* are common symbiosis genes that modulate arbuscular mycorrhiza development. *Plant Cell Physiol.* **54,** 1711–1723 (2013).
124. Rech, S. S. *et al.* A tandem Kunitz protease inhibitor (KPI106)-serine carboxypeptidase (SCP1) controls mycorrhiza establishment and arbuscule development in *Medicago truncatula*. *Plant J.* **75,** 711–725 (2013).
125. Harrison, M. J. A sugar transporter from *Medicago truncatula*: altered expression pattern in roots during vesicular-arbuscular (VA) mycorrhizal associations. *Plant J.* **9,** 491–503 (1996).
126. Wewer, V. *et al.* Fatty acid synthesis and lipid metabolism in the obligate biotrophic fungus *Rhizophagus irregularis* during mycorrhization of *Lotus japonicus*. *Plant J.* **79,** 398–412 (2014).
127. Konecny, J. *et al.* Root transportome in mycorrhizal *Medicago truncatula* – the sweet part. in *ICOM9* (2017).
128. Keymer, A. *et al.* Lipid transfer from plants to arbuscular mycorrhiza fungi. *Elife* **6,** e29107 (2017).
129. Bravo, A. *et al.* Arbuscular mycorrhiza-specific enzymes FatM and RAM2 fine-tune lipid biosynthesis to promote development of arbuscular mycorrhiza. *New Phytol.* **214,** 1631–1645 (2017).
130. Zhang, Q. *et al.* Two *Medicago truncatula* half-ABC transporters are essential for arbuscule development in arbuscular mycorrhizal symbiosis. *Plant Cell* **22,** 1483–1497 (2010).
131. Gutjahr, C. *et al.* The half-size ABC transporters STR1 and STR2 are indispensable for mycorrhizal arbuscule formation in rice. *Plant J.* **69,** 906–920 (2012).
132. Olsson, P. A. *et al.* The use of phospholipid and neutral lipid fatty acids to estimate biomass of arbuscular mycorrhizal fungi in soil. *Mycol. Res.* **99,** 623–629 (1995).
133. Helber, N. *et al.* A versatile monosaccharide transporter that operates in the arbuscular mycorrhizal fungus *Glomus* sp is crucial for the symbiotic relationship with plants. *Plant Cell* **23,** 3812–3823 (2011).
134. Allen, M. in *Mycorrhizal Funct. an Integr. plant-fungal Process* 137–140 (Springer Science & Business Media, 1992).
135. Garcia, K. *et al.* Take a trip through the plant and fungal transportome of mycorrhiza. *Trends Plant Sci.* **21,** 937–950 (2016).
136. Maruyama, H. *et al.* PHO1-type fungal transporter is responsible for phosphate export to the host in arbuscular mycorrhizal symbiosis. in *IMMM3* (2017).
137. Harrison, M. J. *et al.* A phosphate transporter from *Medicago truncatula* involved in the acquisition of phosphate released by arbuscular mycorrhizal fungi. *Plant Cell* **14,** 2413–2429 (2002).

138. Wang, E. *et al.* A H^+-ATPase that energizes nutrient uptake during mycorrhizal symbioses in rice and *Medicago truncatula*. *Plant Cell* **26,** 1818–1830 (2014).
139. Krajinski, F. *et al.* The H^+-ATPase HA1 of *Medicago truncatula* Is essential for phosphate transport and plant growth during arbuscular mycorrhizal symbiosis. *Plant Cell* **26,** 1808–1817 (2014).
140. Sieh, D. *et al.* The arbuscular mycorrhizal symbiosis influences sulfur starvation responses of *Medicago truncatula*. *New Phytol.* **197,** 606–616 (2013).
141. Cruz, C. *et al.* Enzymatic evidence for the key role of arginine in nitrogen translocation by arbuscular mycorrhizal fungi. *Plant Physiol.* **144,** 782–792 (2007).
142. Breuillin-Sessoms, F. *et al.* Suppression of arbuscule degeneration in *Medicago truncatula* phosphate transporter mutants is dependent on the ammonium transporter 2 family protein AMT2;3. *Plant Cell* **27,** tpc.114.131144 (2015).
143. Guether, M. *et al.* A mycorrhizal-specific ammonium transporter from *Lotus japonicus* acquires nitrogen released by arbuscular mycorrhizal fungi. *Plant Physiol.* **150,** 73–83 (2009).
144. Whiteside, M. D. *et al.* Amino acid uptake in arbuscular mycorrhizal plants. *PLoS One* **7,** 8–11 (2012).
145. Porcel, R. *et al.* PIP aquaporin gene expression in arbuscular mycorrhizal *Glycine max* and *Lactuca sativa* plants in relation to drought stress tolerance. *Plant Mol. Biol.* **60,** 389–404 (2006).
146. Krajinski, F. *et al.* Arbuscular mycorrhiza development regulates the mRNA abundance of Mtaqp1 encoding a mercury-insensitive aquaporin of *Medicago truncatula*. *Planta* **211,** 85–90 (2000).
147. Li, T. *et al.* Aquaporin genes GintAQPF1 and GintAQPF2 from *Glomus intraradices* contribute to plant drought tolerance. *Plant Signal. Behav.* **8,** e24030 (2013).
148. Kikuchi, Y. *et al.* Aquaporin-mediated long-distance polyphosphate translocation directed towards the host in arbuscular mycorrhizal symbiosis: application of virus-induced gene silencing. *New Phytol.* **211,** 1202–1208 (2016).
149. Casieri, L. *et al.* Biotrophic transportome in mutualistic plant-fungal interactions. *Mycorrhiza* **23,** 597–625 (2013).
150. Allen, J. W. & Shachar-hill, Y. Sulfur transfer through an arbuscular mycorrhiza. *Plant Physiol.* **149,** 549–560 (2009).
151. Garcia, K. *et al.* Physiological responses and gene co-expression network of mycorrhizal roots under K^+ deprivation. *Plant Physiol.* **173,** pp.01959.2016 (2017).
152. Ellerbeck, M. *et al.* Characterization of three ammonium transporters of the Glomeromycotan fungus *Geosiphon pyriformis*. *Eukaryot. Cell* **12,** 1554–1562 (2013).
153. Belmondo, S. *et al.* A dipeptide transporter from the arbuscular mycorrhizal fungus *Rhizophagus irregularis* is upregulated in the intraradical phase. *Front. Plant Sci.* **5,** 436 (2014).
154. Russo, G. *et al.* TPLATE expression and localization reveal the involvement of cell divisionand endocytosis-related processes in arbuscular mycorrhizal colonization. in *IMMM3* (2017).
155. Gobbato, E. *et al.* RAM1 and RAM2 function and expression during arbuscular mycorrhizal symbiosis and *Aphanomyces euteiches* colonization. *Plant Signal. Behav.* **8,** e26049 (2013).

156. Yu, N. *et al.* A DELLA protein complex controls the arbuscular mycorrhizal symbiosis in plants. *Cell Res.* **24,** 130–133 (2014).
157. Heck, C. *et al.* Symbiotic fungi control plant root cortex development through the novel GRAS transcription factor MIG1. *Curr. Biol.* **26,** 2770–2778 (2016).
158. Xue, L. *et al.* Network of GRAS transcription factors involved in the control of arbuscule development in *Lotus japonicus*. *Plant Physiol.* **167,** 854–871 (2015).
159. Park, H.-J. *et al.* Hyphal branching during arbuscule development requires RAM1. *Plant Physiol.* **169,** 2774–2788 (2015).
160. Delaux, P.-M. *et al.* Algal ancestor of land plants was preadapted for symbiosis. *PNAS* **112,** 13390–13395 (2015).
161. Cerri, M. R. *et al. Medicago truncatula* ERN transcription factors: regulatory interplay with NSP1/NSP2 GRAS factors and expression dynamics throughout rhizobial infection. *Plant Physiol.* **160,** 2155–2172 (2012).
162. Volpe, V. *et al.* The phosphate transporters LjPT4 and MtPT4 mediate early root responses to phosphate status in non mycorrhizal roots. *Plant. Cell Environ.* **39,** 660–671 (2015).
163. Gutjahr, C. & Parniske, M. Cell biology: control of partner lifetime in a plant–fungus relationship. *Curr. Biol.* **27,** 420–423 (2017).
164. Wang, Z. *et al.* Rice SPX1 and SPX2 inhibit phosphate starvation responses through interacting with PHR2 in a phosphate-dependent manner. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 14953–14958 (2014).
165. Fonouni-Farde, C. *et al.* Root development and endosymbioses: DELLAs lead the orchestra. *Trends Plant Sci.* **21,** 898–900 (2016).
166. Floss, D. S. *et al.* DELLA proteins regulate arbuscule formation in arbuscular mycorrhizal symbiosis. *PNAS* **110,** 1–10 (2013).
167. Takeda, N. *et al.* Gibberellins interfere with symbiosis signaling and gene expression and alter colonization by arbuscular mycorrhizal fungi in *Lotus japonicus*. *Plant Physiol.* **167,** 545–557 (2015).
168. Wellman, C. H. *et al.* Fragments of the earliest land plants. *Nature* **425,** 282–285 (2003).
169. Redecker, D. *et al.* Glomalean fungi from the Ordovician. *Science* **289,** 1920–1921 (2000).
170. Ruhfel, B. R. *et al.* From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol. Biol.* **14,** 23 (2014).
171. Edwards, D. *et al.* Could land-based early photosynthesizing ecosystems have bioengineered the planet in mid-Palaeozoic times? *Palaeontology* **58,** 803–837 (2015).
172. Chen, L. *et al.* Cell differentiation and germ–soma separation in Ediacaran animal embryo-like fossils. *Nature* **516,** 238–241 (2014).
173. Lipnicki, L. I. The role of symbiosis in the transition of some eukaryotes from aquatic to terrestrial environments. *Symbiosis* **65,** 39–53 (2015).
174. Martin, F. M. *et al.* Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science* **356,** (2017).
175. Bidartondo, M. I. *et al.* The dawn of symbiosis between plants and fungi. *Biol. Lett.* **7,** 574–577 (2011).
176. Lepage, B. A. *et al.* Fossil ectomycorrhizae from the middle Eocene. *Am. J. Bot.* **84,** 410–412 (1997).
177. Dressler, R. & Dressler, K. *Phylogeny and Classification of the Orchid Family*. (Cambridge University Press, 1993).

178. Schwery, O. *et al.* As old as the mountains: The radiations of the Ericaceae. *New Phytol.* **207,** 355–367 (2015).
179. Ribas, I. The Sun and stars as the primary energy input in planetary atmospheres. in *Sol. Stellar Var. Impact Earth Planets* 1–16 (2009).
180. Li, T. *et al.* Relative importance of an arbuscular mycorrhizal fungus (*Rhizophagus intraradices*) and root hairs in plant drought tolerance. *Mycorrhiza* **24,** 595–602 (2014).
181. Quirk, J. *et al.* Constraining the role of early land plants in Palaeozoic weathering and global cooling. *Proc. Biol. Sci.* **282,** 20151115 (2015).
182. Hibbett, D. S. *et al.* Evolutionary instability of ectomycorrhizal symbioses in basidiomycetes. *Nature* **407,** 506–8 (2000).
183. Modjo, H. S. & Hendrix, J. W. The mycorrhizal fungus *Glomus macrocarpum* as a cause of tobacco stunt disease. *Phytopathology* **76,** 688–691 (1986).
184. Field, K. J. *et al.* From mycoheterotrophy to mutualism: Mycorrhizal specificity and functioning in *Ophioglossum vulgatum* sporophytes. *New Phytol.* **205,** 1492–1502 (2015).
185. Veiga, R. S. L. *et al.* Can arbuscular mycorrhizal fungi reduce the growth of agricultural weeds? *PLoS One* **6,** 1–10 (2011).
186. Jin, L. *et al.* Mycorrhizal-induced growth depression in plants. *Symbiosis* **72,** 81–88 (2016).
187. Walder, F. *et al.* Mycorrhizal networks: common goods of plants shared under unequal terms of trade. *Plant Physiol.* **159,** 789–797 (2012).
188. Zhang, H. *et al.* Plant carbon limitation does not reduce nitrogen transfer from arbuscular mycorrhizal fungi to *Plantago lanceolata*. *Plant Soil* **396,** 369–380 (2015).
189. Jacquemyn, H. *et al.* Mycorrhizal networks and coexistence in species-rich orchid communities. *New Phytol.* **206,** 1127–1134 (2015).
190. Zelmer, C. D. *et al.* Fungi associated with terrestrial orchid mycorrhizas, seeds and protocorms. *Mycoscience* **37,** 439–448 (1996).
191. Simard, S. W. & Durall, D. M. Mycorrhizal networks: a review of their extent, function, and importance. *Can. J. Bot.* **82,** 1140–1165 (2004).
192. Van der Heijden, M. G. A. *et al.* Different arbuscular mycorrhizal fungal species are potential determinants of plant community structure. *Ecology* **79,** 2082–2091 (1998).
193. Fernandez, I. *et al. Arabidopsis*-mycorrhiza, an ambiguous relationship. in *ICOM9* (2017).
194. Bidartondo, M. I. *et al.* Epiparasitic plants specialized on arbuscular mycorrhizal fungi. *Nature* **419,** 389–392 (2002).
195. Merckx, V. S. F. T. *et al.* Mycoheterotrophic interactions are not limited to a narrow phylogenetic range of arbuscular mycorrhizal fungi. *Mol. Ecol.* **21,** 1524–1532 (2012).
196. Bidartondo, M. I. The evolutionary ecology of mycoheterotrophy. *New Phytol.* **167,** 335–352 (2005).
197. Robinson, D. & Fitter, A. The magnitude and control of carbon transfer between plants linked by a common mycorrhizal network. *J. Exp. Bot.* **50,** 9–13 (1999).
198. Fochi, V. *et al.* Fungal and plant gene expression in the *Tulasnella calospora–Serapias vomeracea* symbiosis provides clues about nitrogen pathways in orchid mycorrhizas. *New Phytol.* **213,** 365–379 (2017).
199. Wang, D. *et al.* Symbiosis specificity in the legume - rhizobial mutualism. *Cell.*

Microbiol. **14,** 334–342 (2012).
200. Ortiz, N. *et al.* Contribution of arbuscular mycorrhizal fungi and/or bacteria to enhancing plant drought tolerance under natural soil conditions: Effectiveness of autochthonous or allochthonous strains. *J. Plant Physiol.* **174,** 87–96 (2015).
201. Ceballos, I. *et al.* The *in vitro* mass-produced model mycorrhizal fungus, *Rhizophagus irregularis*, significantly increases yields of the globally important food security crop cassava. *PLoS One* **8,** e70633 (2013).
202. Augé, R. M. *et al.* Arbuscular mycorrhizal symbiosis and osmotic adjustment in response to NaCl stress: a meta-analysis. *Front. Plant Sci.* **5,** 1–14 (2014).
203. Gosling, P. *et al.* Evidence for functional redundancy in arbuscular mycorrhizal fungi and implications for agroecosystem management. *Mycorrhiza* **26,** 77–83 (2016).
204. Fellbaum, C. R. *et al.* Carbon availability triggers fungal nitrogen uptake and transport in arbuscular mycorrhizal symbiosis. *Proc. Natl. Acad. Sci.* **109,** 2666–2671 (2012).
205. Thirkell, T. J. *et al.* Resolving the 'nitrogen paradox' of arbuscular mycorrhizas: fertilization with organic matter brings considerable benefits for plant nutrition and growth. *Plant, Cell Environ.* **39,** 1683–1690 (2016).
206. Corrêa, A. *et al.* Nitrogen and carbon/nitrogen dynamics in arbuscular mycorrhiza: the great unknown. *Mycorrhiza* **25,** 499–515 (2015).
207. Johnson, N. C. *et al.* Mycorrhizal phenotypes and the Law of the Minimum. *New Phytol.* **205,** 1473–1484 (2015).
208. Hodge, A. & Fitter, A. H. Substantial nitrogen acquisition by arbuscular mycorrhizal fungi from organic material has implications for N cycling. *PNAS* **107,** 13754–13759 (2010).
209. Li, J. F. *et al.* Arbuscular mycorrhizal fungi increase growth and phenolics synthesis in *Poncirus trifoliata* under iron deficiency. *Sci. Hortic. (Amsterdam).* **183,** 87–92 (2015).
210. Giri, B. & Mukerji, K. G. Mycorrhizal inoculant alleviates salt stress in *Sesbania aegyptiaca* and *Sesbania grandiflora* under field conditions: Evidence for reduced sodium and improved magnesium uptake. *Mycorrhiza* **14,** 307–312 (2004).
211. Jamal, A. *et al.* Arbuscular mycorrhizal fungi enhance zinc and nickel uptake from contaminated soil by soybean and lentil. *Int. J. Phytoremediation* **4,** 205–221 (2002).
212. Tamayo, E. *et al.* Genome-wide analysis of copper, iron and zinc transporters in the arbuscular mycorrhizal fungus *Rhizophagus irregularis*. *Front. Plant Sci.* **5,** 547 (2014).
213. Shah, F. *et al.* Ectomycorrhizal fungi decompose soil organic matter using oxidative mechanisms adapted from saprotrophic ancestors. *New Phytol.* **209,** 1705–1719 (2016).
214. Lin, G. *et al.* Arbuscular mycorrhizal fungal effects on plant competition and community structure. *J. Ecol.* **103,** 1224–1232 (2015).
215. Doubková, P. *et al.* Arbuscular mycorrhizal symbiosis alleviates drought stress imposed on *Knautia arvensis* plants in serpentine soil. *Plant Soil* **370,** 149–161 (2013).
216. Chitarra, W. *et al.* Arbuscular mycorrhizal symbiosis-mediated tomato tolerance to drought. *Plant Signal. Behav.* **11,** e1197468 (2016).
217. Audenaert, K. *et al.* Abscisic acid determines basal susceptibility of tomato to

Botrytis cinerea and suppresses salicylic acid-dependent signaling mechanisms. *Plant Physiol.* **128,** 491–501 (2002).
218. Rillig, M. C. *et al.* The role of arbuscular mycorrhizal fungi and glomalin in soil aggregation: comparing effects of five plant species. *Plant Soil* **238,** 325–333 (2002).
219. Göhre, V. & Paszkowski, U. Contribution of the arbuscular mycorrhizal symbiosis to heavy metal phytoremediation. *Planta* **223,** 1115–1122 (2006).
220. Li, H. *et al.* Do arbuscular mycorrhizal fungi affect cadmium uptake kinetics, subcellular distribution and chemical forms in rice? *Sci. Total Environ.* **571,** 1183–1190 (2016).
221. Seguel, A. *et al.* The role of arbuscular mycorrhizas in decreasing aluminium phytotoxicity in acidic soils: a review. *Mycorrhiza* **23,** 167–183 (2013).
222. Ruscitti, M. *et al.* Improvement of copper stress tolerance in pepper plants (*Capsicum annuum L.*) by inoculation with arbuscular mycorrhizal fungi. *Theor. Exp. Plant Physiol.* **29,** 37–49 (2017).
223. Merlos, M. A. *et al.* The arbuscular mycorrhizal fungus *Rhizophagus irregularis* differentially regulates the copper response of two maize cultivars differing in copper tolerance. *Plant Sci.* **253,** 68–76 (2016).
224. Gonzalez-Chavez, C. *et al.* Arbuscular mycorrhizal fungi confer enhanced arsenate resistance on *Holcus lanatus*. *New Phytol.* **155,** 163–171 (2002).
225. Firmin, S. *et al.* Arbuscular mycorrhizal fungal inoculation protects *Miscanthus x giganteus* against trace element toxicity in a highly metal-contaminated site. *Sci. Total Environ.* **527–528,** 91–99 (2015).
226. Doubkova, P. & Sudova, R. Limited impact of arbuscular mycorrhizal fungi on clones of *Agrostis capillaris* with different heavy metal tolerance. *Appl. Soil Ecol.* **99,** 78–88 (2016).
227. Cameron, D. D. *et al.* Mycorrhiza-induced resistance: More than the sum of its parts? *Trends Plant Sci.* **18,** 539–545 (2013).
228. Babikova, Z. *et al.* Arbuscular mycorrhizal fungi and aphids interact by changing host plant quality and volatile emission. *Funct. Ecol.* **28,** 375–385 (2014).
229. Shrivastava, G. *et al.* Colonization by arbuscular mycorrhizal and endophytic fungi enhanced terpene production in tomato plants and their defense against a herbivorous insect. *Symbiosis* **65,** 65–74 (2015).
230. Liu, J. *et al.* Arbuscular mycorrhizal symbiosis is accompanied by local and systemic alterations in gene expression and an increase in disease resistance in the shoots. *Plant J.* **50,** 529–544 (2007).
231. Boyacıoglu, T. U. & Uyanöz, R. Effects of mycorrhizal fungi on tolerance capability of corn grown under salt stress condition. *J. Plant Nutr.* **37,** 107–122 (2014).
232. Song, Y. *et al.* Enhanced tomato disease resistance primed by arbuscular mycorrhizal fungus. *Front. Plant Sci.* **6,** 786 (2015).
233. Minton, M. M. *et al.* Effects of arbuscular mycorrhizal fungi on herbivory defense in two Solanum (Solanaceae) species. *Plant Ecol. Evol.* **149,** 157–164 (2016).
234. Mora-Romero, G. A. *et al.* Mycorrhiza-induced protection against pathogens is both genotype-specific and graft-transmissible. *Symbiosis* **66,** 55–64 (2015).
235. Cordier, C. *et al.* Cell defense responses associated with localized and systemic resistance to *Phytophthora parasitica* induced in tomato by an arbuscular mycorrhizal fungus. *Mol. Plant-Microbe Interact.* **11,** 1017–1028 (1998).

236. Pozo, M. J. *et al.* Localized versus systemic effect of arbuscular mycorrhizal fungi on defence responses to *Phytophthora infection* in tomato plants. *J. Exp. Bot.* **53,** 525–534 (2002).
237. Lopez-Raez, J. A. *et al.* Hormonal and transcriptional profiles highlight common and differential host responses to arbuscular mycorrhizal fungi and the regulation of the oxylipin pathway. *J. Exp. Bot.* **61,** 2589–2601 (2010).
238. Jung, S. C. *et al.* Mycorrhiza-induced resistance and priming of plant defenses. *J. Chem. Ecol.* **38,** 651–664 (2012).
239. Mauch-Mani, B. *et al.* Defense priming: an adaptive part of induced resistance. *Annu. Rev. Plant Biol.* **68,** 485–512 (2017).
240. Van Wees, S. C. *et al.* Plant immune responses triggered by beneficial microbes. *Curr. Opin. Plant Biol.* **11,** 443–448 (2008).
241. Barto, E. K. *et al.* The fungal fast lane: Common mycorrhizal networks extend bioactive zones of allelochemicals in soils. *PLoS One* **6,** e27195 (2011).
242. Stinson, K. A. *et al.* Invasive plant suppresses the growth of native tree seedlings by disrupting belowground mutualisms. *PLoS Biol.* **4,** 727–731 (2006).
243. Booth, M. G. Mycorrhizal networks mediate overstorey-understorey competition in a temperate forest. *Ecol. Lett.* **7,** 538–546 (2004).
244. Harley, J. L. The significance of mycorrhiza. *Mycol. Res.* **92,** 129–139 (1989).
245. Bago, B. *et al.* Branched absorbing structures (BAS): A feature of the extraradical mycelium of symbiotic arbuscular mycorrhizal fungi. *New Phytol.* **139,** 375–388 (1998).
246. Giovannetti, M. *et al.* Microchambers and video-enhanced light microscopy for monitoring cellular events in living hyphae of arbuscular mycorrhizal fungi. *Plant Soil* **226,** 153–159 (2000).
247. Kothamasi, D. *et al.* Arbuscular mycorrhizae and phosphate solubilising bacteria of the rhizosphere of the mangrove ecosystem of Great Nicobar island, India. *Biol. Fertil. Soils* **42,** 358–361 (2006).
248. Matusova, R. *et al.* The strigolactone germination stimulants of the plant-parasitic *Striga* and *Orobanche* spp. are derived from the carotenoid pathway. *Plant Physiol.* **139,** 920–934 (2005).
249. Foo, E. *et al.* Plant hormones in arbuscular mycorrhizal symbioses: an emerging role for gibberellins. *Ann. Bot.* **111,** 769–779 (2013).
250. Fich, E. A. *et al.* The plant polyester cutin: biosynthesis, structure, and biological roles. *Annu. Rev. Plant Biol.* **67,** 207–233 (2016).
251. García-Parisi, P. A. *et al.* Three symbionts involved in interspecific plant-soil feedback: epichloid endophytes and mycorrhizal fungi affect the performance of rhizobia-legume symbiosis. *Plant Soil* **412,** 151–162 (2017).
252. Martínez-Hidalgo, P. *et al.* in *Biol. Nitrogen Fixat. Benef. Plant-Microbe Interact.* (González-Andrés, F. & James, E.) 123–130 (Springer International Publishing, 2016). doi:10.1007/978-3-319-32528-6_11
253. Lumini, E. *et al.* Presymbiotic growth and sporal morphology are affected in the arbuscular mycorrhizal fungus *Gigaspora margarita* cured of its endobacteria. *Cell. Microbiol.* **9,** 1716–1729 (2007).
254. Mirabal Alonso, L. *et al.* Spores of the mycorrhizal fungus *Glomus mosseae* host yeasts that solubilize phosphate and accumulate polyphosphates. *Mycorrhiza* **18,** 197–204 (2008).
255. Mohandas, S. *et al.* Guava (*Psidium guajava L.*) rhizosphere *Glomus mosseae* spores harbor actinomycetes with growth promoting and antifungal attributes.

Sci. Hortic. (Amsterdam). **150,** 371–376 (2013).
256. Naumann, M. *et al.* The obligate endobacteria of arbuscular mycorrhizal fungi are ancient heritable components related to the Mollicutes. *ISME J.* **4,** 862–871 (2010).
257. van Buuren, M. *et al.* Construction and characterization of genomic libraries of two endomycorrhizal fungi: *Glomus versiforme* and *Gigaspora margarita*. *Mycol. Res.* **103,** 955–960 (1999).
258. Naito, M. *et al.* '*Candidatus Moeniiplasma glomeromycotorum*', an endobacterium of arbuscular mycorrhizal fungi. *Int. J. Syst. Evol. Microbiol.* **67,** 1177–1184 (2017).
259. Salvioli, A. *et al.* Symbiosis with an endobacterium increases the fitness of a mycorrhizal fungus, raising its bioenergetic potential. *ISME J.* **10,** 130–144 (2015).
260. Ghignone, S. *et al.* The genome of the obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions. *ISME J.* **6,** 136–145 (2012).
261. Ikeda, Y. *et al.* A novel virus-like double-stranded RNA in an obligate biotroph arbuscular mycorrhizal fungus: a hidden player in mycorrhizal symbiosis. *Mol. Plant-Microbe Interact.* **25,** 1005–1012 (2012).
262. Kitahara, R. *et al.* A unique mitovirus from Glomeromycota, the phylum of arbuscular mycorrhizal fungi. *Arch. Virol.* **159,** 2157–2160 (2014).
263. Turina, M. *et al.* The virome of the arbuscular mycorrhizal fungus *Gigaspora margarita*. in *ICOM9* (2017).
264. Frey-Klett, P. *et al.* The mycorrhiza helper bacteria revisited. *New Phytol.* **176,** 22–36 (2007).
265. Halary, S. *et al.* Mating type gene homologues and putative sex pheromone-sensing pathway in arbuscular mycorrhizal fungi, a presumably asexual plant root symbiont. *PLoS One* **8,** e80729 (2013).
266. Ropars, J. *et al.* Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. *Nat. Microbiol.* **1,** 16033 (2016).
267. Bago, B. *et al. In vivo* studies on the nuclear behavior of the arbuscular mycorrhizal fungus *Gigaspora rosea* grown under axenic conditions. *Protoplasma* **203,** 1–15 (1998).
268. Williams, J. G. Dictyostelium finds new roles to model. *Genetics* **185,** 717–726 (2010).
269. Nishii, I. & Miller, S. M. *Volvox*: Simple steps to developmental complexity? *Curr. Opin. Plant Biol.* **13,** 646–653 (2010).
270. Helber, N. & Requena, N. Expression of the fluorescence markers DsRed and GFP fused to a nuclear localization signal in the arbuscular mycorrhizal fungus Glomus intraradices. *New Phytol.* **177,** 537–548 (2008).
271. Samuel Roberts Noble Foundation. *Medicago truncatula* Handbook. (2006). at <https://www.noble.org/medicagohandbook/>
272. Tadege, M. *et al.* Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume *Medicago truncatula*. *Plant J.* **54,** 335–47 (2008).
273. Grandbastien, M.-A. *et al.* Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenet. Genome Res.* **110,** 229–241 (2005).
274. Grandbastien, M.-A. *et al.* Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* **337,** 376–380 (1989).

275. D'Erfurth, I. *et al.* Efficient transposition of the Tnt1 tobacco retrotransposon in the model legume *Medicago truncatula*. *Plant J.* **34,** 95–106 (2003).
276. Ratet, P. *et al.* in *Medicago truncatula Handb.* 1–12 (2006).
277. *Medicago truncatula* Mutant Database. *Samuel Roberts Noble Found.* at <http://medicago-mutant.noble.org/mutant/database.php>
278. Schultze, M. in *Plant Miner. Nutr. - Methods Protoc.* (Maathuis, F. J. M.) 47–59 (Springer Science & Buisness Media, 2013). doi:10.1007/978-1-62703-152-3_3
279. Booth, R. *et al.* in *Methods Comp. plant Ecol. a Lab. Man.* (Hendry, G. & Grime, J.) 19–23 (Chapmn Hall, 1993).
280. Rogers, S. O. & Bendich, A. J. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Mol. Biol.* **5,** 69–76 (1985).
281. Schultze, M. Personal Communication.
282. Stirnberg, P. Personal Communication.
283. IDT OligoAnalzer 3.1. *Intergrated DNA Technol.* at <https://eu.idtdna.com/calc/analyzer>
284. Vierheilig, H. *et al.* An overview of methods for the detection and observation of arbuscular mycorrhizal fungi in roots. *Physiol. Plant.* **125,** 393–404 (2005).
285. Lum, M. R. *et al.* Investigation of four classes of non-nondulating white sweetclover (*Melilotus alba Annua Desr*.) mutants and their responses to arbuscular-mycorrhizal fungi. *Integr. Comp. Biol.* **42,** 295–303 (2002).
286. Center for information biology. Gene Expression Atlas. (2014). at <http://cibex.nig.ac.jp/data/index.html>
287. Sharma, M. P. & Buyer, J. S. Comparison of biochemical and microscopic methods for quantification of arbuscular mycorrhizal fungi in soil and roots. *Appl. Soil Ecol.* **95,** 86–89 (2015).
288. Johnson, N. C. *et al.* Functioning of mycorrhizal associations along the mutualism-parasitism continuum. *New Phytol.* **135,** 575–586 (1997).
289. Smith, F. A. & Smith, S. E. How harmonious are arbuscular mycorrhizal symbioses? Inconsistent concepts reflect different mindsets as well as results. *New Phytol.* **205,** 1381–1384 (2015).
290. Kahiluoto, H. & Vestberg, M. The effect of arbuscular mycorrhiza on biomass production and phosphorus uptake from sparingly soluble sources by leek (*Allium porrum L.*) in Finnish field soils. *Biol. Agric. Hortic.* **16,** 65–85 (1998).
291. Groten, K. *et al.* Silencing a key gene of the common symbiosis pathway in Nicotiana attenuata specifically impairs arbuscular mycorrhizal infection without influencing the root-associated microbiome or plant growth. *Plant. Cell Environ.* **38,** 2398–2416 (2015).
292. Guillotin, B. *et al.* NIN is involved in the regulation of arbuscular mycorrhizal symbiosis. *Front. Plant Sci.* **7,** 1704 (2016).
293. Kakar, K. *et al.* A community resource for high-throughput quantitative RT-PCR analysis of transcription factor gene expression in *Medicago truncatula*. *Plant Methods* **4,** 18 (2008).
294. Charpentier, M. *et al.* ABA promotion of arbuscular mycorrhizal colonization requires a component of the PP2A Protein Phosphatase Complex. *Plant Physiol.* **166,** 2077–2090 (2014).
295. Tang, H. *et al.* An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC Genomics* **15,** 312 (2014).
296. Goodwin, S. *et al.* Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17,** 333–351 (2016).

297. Mudge, J. *et al. Medicago truncatula* R108 Genome. *Medicago truncatula HapMap Proj.* (2015). at <http://www.medicagohapmap.org/downloads/r108>
298. Ziya, S. Identifing genes in *Medicago* mycorrhizal mutants. (2011).
299. Betts, L. Identifing genes in *Medicago* mycorrhizal mutants. (2011).
300. Harborough, S. R. Identifing genes in *Medicago* mycorrhizal mutants. (2012).
301. Keith, L. Identifing genes in *Medicago* mycorrhizal mutants. (2012).
302. Bouvet, L. Identifing genes in *Medicago* mycorrhizal mutants. (2013).
303. Samuel Roberts Noble Foundation. *Medicago truncatula* Genome Browser IMGAG V4.1. at <https://gb.noble.org/cgi-bin/gb2/gbrowse/medicago_4/>
304. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **25,** 1754–1760 (2009).
305. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
306. Veerappan, V. *et al.* Rapid identification of causative insertions underlying *Medicago truncatula* tnt1 mutants defective in symbiotic nitrogen fixation from a forward genetic screen by whole genome sequencing. *BMC Genomics* **17,** 141 (2016).
307. Milne, I. *et al.* Using tablet for visual exploration of second-generation sequencing data. *Brief. Bioinform.* **14,** 193–202 (2013).
308. Medicago truncatula R108 BLAST tool. *Medicago truncatula HAPMAP Proj.* at <http://www.medicagohapmap.org/tools/r108_blastform>
309. Medicago truncatula A17 BLAST tool. *JCVI* at <http://blast.jcvi.org/er-blast/index.cgi?project=mtbe>
310. NCBI BLAST Tool. *NCBI* at <https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome>
311. Handa, Y. *et al.* RNA-seq transcriptional profiling of an arbuscular mycorrhiza provides insights into regulated and coordinated gene expression in *Lotus japonicus* and *Rhizophagus irregularis*. *Plant Cell Physiol.* **56,** 1490–1511 (2015).
312. Giovannetti, M. *et al.* Early *Lotus japonicus* root transcriptomic responses to symbiotic and pathogenic fungal exudates. *Front. Plant Sci.* **6,** 480 (2015).
313. Samuel Roberts Noble Foundation. Noble Foundation *Medicago* Expression Atlas. at <https://mtgea.noble.org/v3/blast_search_form.php>
314. Kobae, Y. *et al.* Localized expression of arbuscular mycorrhiza-inducible ammonium transporters in soybean. *Plant Cell Physiol.* **51,** 1411–1415 (2010).
315. Weston, D. E. *et al.* The Pea DELLA proteins LA and CRY are important regulators of gibberellin synthesis and root growth. *Plant Physiol.* **147,** 199–205 (2008).
316. Lévy, J. *et al.* A putative Ca^{2+} and calmodulin- dependent protein kinase required for bacterial and fungal symbioses. *Science* **303,** 1361–1364 (2004).
317. Lelandais-Briere, C. *et al.* Noncoding RNAs, emerging regulators in root endosymbioses. *Mol. Plant-Microbe Interact.* **29,** 170–180 (2015).
318. Bazin, J. *et al.* miR396 affects mycorrhization and root meristem activity in the legume *Medicago truncatula*. *Plant J.* **74,** 920–934 (2013).
319. Javot, H. *et al.* A *Medicago truncatula* phosphate transporter indispensable for the arbuscular mycorrhizal symbiosis. *PNAS* **104,** 1720–1725 (2007).
320. Kamphuis, L. G. *et al.* The *Medicago truncatula* reference accession A17 has an aberrant chromosomal configuration. *New Phytol.* **174,** 299–303 (2007).
321. Sampson, J. *et al.* Efficient study design for next generation sequencing. *Genet. Epidemiol.* **35,** 269–277 (2011).
322. Li, D. *et al.* MEGAHIT: An ultra-fast single-node solution for large and complex

metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31,** 1674–1676 (2015).
323. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).
324. Pouteau, S. *et al.* Specific expression of the tobacco tnt1 in protoplasts. *EMBO J.* **10,** 1911–1918 (1991).
325. Gust, A. A. *et al.* Plant LysM proteins: modules mediating symbiosis and immunity. *Trends Plant Sci.* **17,** 495–502 (2012).
326. de Torres-Zabala, M. *et al.* Pseudomonas syringae pv. tomato hijacks the *Arabidopsis* abscisic acid signalling pathway to cause disease. *EMBO J.* **26,** 1434–1343 (2007).
327. Li, G. *et al. Medicago truncatula* ecotypes A17 and R108 differed in their response to iron deficiency. *J. Plant Physiol.* **171,** 639–647 (2014).
328. Foroozanfar, M. PhD Thesis: Genetic control of tolerance to salinity in *Medicago truncatula*. (2013).
329. Gaige, A. R. *et al. Medicago truncatula* ecotypes A17 and R108 show variations in jasmonic acid/ethylene induced resistance to *Macrophomina phaseolina*. *Can. J. Plant Pathol.* **34,** 98–103 (2012).
330. Hoffmann, B. *et al.* A new *Medicago truncatula* line with superior *in vitro* regeneration, transformation, and symbiotic properties isolated through cell culture selection. *MPMI* **10,** 307–315 (1997).
331. Lee, M. & Phillips, R. L. The chromosomal basis of somaclonal variation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **39,** 413–437 (1988).
332. Wang, T.-Z. *et al.* Genome variations account for different response to three mineral elements between *Medicago truncatula* ecotypes Jemalong A17 and R108. *BMC Plant Biol.* **14,** 122 (2014).
333. Ronfort, J. *et al.* Microsatellite diversity and broad scale geographic structure in a model legume: building a set of nested core collection for studying naturally occurring variation in *Medicago truncatula*. *BMC Plant Biol.* **6,** 28 (2006).
334. Songstad, D. D. *et al.* Genome editing of plants. *CRC. Crit. Rev. Plant Sci.* **36,** (2017).
335. Engler, C. *et al.* A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3,** e3647 (2008).
336. Tadege, M. *et al.* Insertional mutagenesis: A Swiss Army knife for functional genomics of *Medicago truncatula*. *Trends Plant Sci.* **10,** 229–235 (2005).
337. Iantcheva, A. *et al.* Tnt1 retrotransposon as an efficient tool for development of an insertional mutant collection of *Lotus japonicus*. *Vitr. Cell. Dev. Biol. - Plant* **52,** 338–347 (2016).
338. Mhiri, C. *et al.* The promoter of the tobacco tnt1 retrotransposon is induced by wounding and by abiotic stress. *Plant Mol. Biol.* **33,** 257–266 (1997).
339. Williams, M. E. Alternative mutagens for maize (Zea mays L.). *Maize Genomics Genet.* **7,** 1–8 (2016).
340. Vandenbussche, M. *et al.* Petunia, your next supermodel? *Front. Plant Sci.* **7,** 72 (2016).
341. Bombarely, A. *et al.* Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. *Nat. Plants* **2,** 16074 (2016).
342. Miyao, A. *et al.* Molecular spectrum of somaclonal variation in regenerated rice revealed by whole-genome sequencing. *Plant Cell Physiol.* **53,** 256–264 (2012).
343. Veerappan, V. *et al.* Keel petal incision: a simple and efficient method for

genetic crossing in *Medicago truncatula*. *Plant Methods* **10,** 11 (2014).
344. Mun, J. H. *et al.* Distribution of microsatellites in the genome of *Medicago truncatula*: A resource of genetic markers that integrate genetic and physical maps. *Genetics* **172,** 2541–2555 (2006).
345. Enattah, N. S. *et al.* Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am. J. Hum. Genet.* **82,** 57–72 (2008).
346. Akoh, C. C. *et al.* GDSL family of serine esterases/lipases. *Prog. Lipid Res.* **43,** 534–552 (2004).
347. Miguel, C. & Marum, L. An epigenetic view of plant cells cultured *in vitro*: Somaclonal variation and beyond. *J. Exp. Bot.* **62,** 3713–3725 (2011).
348. Rakocevic, A. *et al.* MERE1, a low-copy-number copia-type retroelement in *Medicago truncatula* active during tissue culture. *Plant Physiol.* **151,** 1250–1263 (2009).
349. Couzigou, J. M. *et al.* Positive gene regulation by a natural protective miRNA enables arbuscular mycorrhizal symbiosis. *Cell Host Microbe* **21,** 106–112 (2017).
350. Mielczarek, M. & Szyda, J. Review of alignment and SNP calling algorithms for next-generation sequencing data. *J. Appl. Genet.* **57,** 71–79 (2016).
351. Narasimhan, V. *et al.* BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* **32,** 1749–1751 (2016).
352. Small, I. RNAi for revealing and engineering plant gene functions. *Curr. Opin. Biotechnol.* **18,** 148–153 (2007).
353. Scholte, M. *et al.* T-DNA tagging in the model legume *Medicago truncatula* allows efficient gene discovery. *Mol. Breed.* **10,** 203–215 (2002).
354. Carelli, M. *et al.* in *Legum. Genomics Methods Protoc.* (Rose, R. J.) 101–118 (Humana Press, 2013). doi:10.1007/978-1-62703-613-9_9
355. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32,** 347–355 (2014).
356. Li, J.-F. *et al.* Multiplex and homologous recombination–mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9 - Supplimentary Infomation. *Nat. Biotechnol.* **31,** 688–691 (2013).
357. NEB. New England Biolabs Tm Calculator. at <http://tmcalculator.neb.com/#!/>
358. Patron, N. J. Type 2S Assemlby Protocol. at <http://synbio.tsl.ac.uk/golden-gate-assembly-protocol/>
359. Weber, E. *et al.* A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6,** e16765 (2011).
360. Engler, C. *et al.* A Golden Gate modular cloning toolbox for plants. *ACS Synth. Biol.* **3,** 839–843 (2014).
361. Sambrook, J. & Russel, D. W. *Molecular Cloning*. (Cold Spring Harbor Lab Press, 2001).
362. Hofgen, R. & Willmitzer, L. Storage of competent cells for Agrobacterium transformation. *Nucleic Acids Res.* **16,** 9877 (1988).
363. Carvalho, R. Personal Communication.
364. JCVI. *Medicago truncatula* Genome Database. at <http://www.medicagogenome.org/home>
365. Aung, K. *et al.* Pho2, a phosphate overaccumulator, is caused by a nonsense mutation in a microRNA399 target gene. *Plant Physiol.* **141,** 1000–1011 (2006).
366. Gu, M. *et al.* Complex regulation of plant phosphate transporters and the gap

Between molecular mechanisms and practical application: what are missing? *Mol. Plant* **9,** 396–416 (2015).
367. Jackson, K. J. PhD Thesis: Characterisation and expression of receptor-like cytoplasmic kinases in *Medicago truncatula* during rhizobial and arbuscular mycorrhizal symbioses. (2015).
368. Favre, P. *et al.* A novel bioinformatics pipeline to discover genes related to arbuscular mycorrhizal symbiosis based on their evolutionary conservation pattern among higher plants. *BMC Plant Biol.* **14,** 1–20 (2014).
369. Delaux, P. M. *et al.* Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *PLoS Genet.* **10,** e1004487 (2014).
370. Bravo, A. *et al.* Genes conserved for arbuscular mycorrhizal symbiosis identified through phylogenomics. *Nat. Plants* **2,** 15208 (2016).
371. Gutjahr, C. *et al.* Arbuscular mycorrhiza-specific signaling in rice transcends the common symbiosis signaling pathway. *Plant Cell* **20,** 2989–3005 (2008).
372. Roth, R. *et al.* Arbuscule-specific Receptor Kinase1, ARK1, is required to sustain fungal symbiosis in rice roots. in *IMMM3* (2017).
373. Aloui, A. *et al.* The plasma membrane proteome of *Medicago truncatula* roots as modified by arbuscular mycorrhizal symbiosis. *Mycorrhiza* 1–16 (2017). doi:10.1007/s00572-017-0789-5
374. Finn, R. D. *et al.* InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45,** 190–199 (2017).
375. Kelley, L. A. *et al.* The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10,** 845–858 (2015).
376. Käll, L. *et al.* Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Res.* **35,** 429–432 (2007).
377. Bordoli, M. R. *et al.* A secreted tyrosine kinase acts in the extracellular environment. *Cell* **158,** 1033–1044 (2014).
378. Chivasa, S. *et al.* Proteomic analysis of the *Arabidopsis thaliana* cell wall. *Electrophoresis* **23,** 1754–1765 (2002).
379. González, M. *et al.* High abundance of Serine/Threonine-rich regions predicted to be hyper-O-glycosylated in the secretory proteins coded by eight fungal genomes. *BMC Microbiol.* **12,** 213 (2012).
380. Rathore, A. S. & Gupta, R. D. Chitinases from bacteria to human: Properties, applications, and future perspectives. *Enzyme Res.* **2015,** 1–8 (2015).
381. Rich, M. K. *et al.* Role of the GRAS transcription factor ATA/RAM1 in the transcriptional reprogramming of arbuscular mycorrhiza in *Petunia hybrida.* *BMC Genomics* **18,** 589 (2017).
382. Frenzel, A. *et al.* Combined transcriptome profiling reveals a novel family of arbuscular mycorrhizal-specific *Medicago truncatula* lectin genes. *Mol. Plant. Microbe. Interact.* **18,** 771–782 (2005).
383. Yano, K. *et al.* CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 20540–20545 (2008).
384. Miller, J. B. *et al.* Calcium/calmodulin-dependent protein kinase is negatively and positively regulated by calcium, providing a mechanism for decoding calcium responses during symbiosis signaling. *Plant Cell* **12,** 5053–5066 (2013).
385. Leborgne-Castel, N. & Bouhidel, K. Plasma membrane protein trafficking in plant – microbe interactions: a plant cell point of view. *Front. Plant Sci.* **5,** 1–13 (2014).

386. Verbruggen, E. *et al.* Mycorrhizal fungal establishment in agricultural soils : factors determining inoculation success. *New Phytol.* **197,** 1104–1109 (2013).
387. Orr, H. A. Fitness and its role in evolutionary genetics. *Natl. Rev. Genet.* **10,** 531–539 (2009).
388. Gao, L.-L. *et al.* Colonization patterns in a mycorrhiza-defective mutant tomato vary with different arbuscular-mycorrhizal fungi. *New Phytol.* **151,** 477–491 (2001).
389. Corradi, N. Origin and evolution of genetic diversity within and between the AMF mycelia. in *ICOM9* (2017).
390. Murray, J. Personal Communication.
391. Lee, J. *et al.* Improved PCR primers for the detection and identification of arbuscular mycorrhizal fungi. *FEMS Microbiol. Ecol.* **65,** 339–349 (2008).
392. Tillott, P. M. in *A Hist. Cty. York City York* 498–506 (1961).
393. Adolfsson, L. *et al.* Mycorrhiza symbiosis increases the surface for sunlight capture in *Medicago truncatula* for better photosynthetic production. *PLoS One* **10,** e0115314 (2015).
394. Sulieman, S. *et al.* Growth and nodulation of symbiotic *Medicago truncatula* at different levels of phosphorus availability. *J. Exp. Bot.* **64,** 2701–2712 (2013).
395. Depuydt, S. Arguments for and against self and non-self root recognition in plants. *Front. Plant Sci.* **5,** 614 (2014).
396. Rose, R. J. *Medicago truncatula* as a model for understanding plant interactions with other organisms, plant development and stress biology: Past, present and future. *Funct. Plant Biol.* **35,** 253–264 (2008).
397. Bradbury, S. M. *et al.* Interactions between three alfalfa nodulation genotypes and two *Glomus* species. *New Phytol.* **119,** 115–120 (1991).
398. Schornack, S. Assessing candidate effector activities in *Nicotiana benthamiana*. in *IMMM3* (2017).
399. Koch, A. M. *et al.* Evolutionary asymmetry in the arbuscular mycorrhizal symbiosis: conservatism in fungal morphology does not predict host plant growth. *New Phytol.* **214,** 1330–1337 (2017).
400. MacLean, A. M. *et al.* Plant signaling and metabolic pathways enabling arbuscular mycorrhizal symbiosis. *Plant Cell* (2017). doi:10.1105/tpc.17.00555
401. Paszkowski, U. *et al.* Maize mutants affected at distinct stages of the arbuscular mycorrhizal symbiosis. *Plant J.* **47,** 165–173 (2006).
402. Sędzielewska Toro, K. & Brachmann, A. The effector candidate repertoire of the arbuscular mycorrhizal fungus *Rhizophagus clarus*. *BMC Genomics* **53,** 1689–1699 (2016).
403. Klironomos, J. N. *et al.* Species of arbuscular mycorrhizal fungi affect mycorrhizal responses to simulated herbivory. *Appl. Soil Ecol.* **26,** 133–141 (2004).
404. Ruiz Lozano, J. M. *et al.* Effects of arbuscular-mycorrhizal *Glomus* species on drough tolerance: physiology and nutritional plant responses. *Appl Env. Microbiol* **61,** 456–460 (1995).
405. Burleigh, S. H. *et al.* Functional diversity of arbuscular mycorrhizas extends to the expression of plant genes involved in P nutrition. *J. Exp. Bot.* **53,** 1593–1601 (2002).
406. Yang, S.-Y. *et al.* Nonredundant regulation of rice arbuscular mycorrhizal symbiosis by two members of the phosphate transporter1 gene family. *Plant Cell* **24,** 4236–4251 (2012).
407. Jayaraman, D. *et al.* Identification of the phosphorylation targets of symbiotic

receptor-like kinases using a high-throughput multiplexed assay for kinase specificity. *Plant J.* **90,** 1196–1207 (2017).
408. Dai, M. *et al.* Impact of land use on arbuscular mycorrhizal fungal communities in rural Canada. *Appl. Environ. Microbiol.* **79,** 6719–6729 (2013).
409. Wyss, T. *et al.* Population genomics reveals that within-fungus polymorphism is common and maintained in populations of the mycorrhizal fungus *Rhizophagus irregularis*. *ISME J.* **10,** 2514–2536 (2016).
410. Davison, J. *et al.* Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science* **439,** 970–973 (2015).
411. Watts-Williams, S. J. *et al.* Local and distal effects of arbuscular mycorrhizal colonization on direct pathway Pi uptake and root growth in *Medicago truncatula*. *J. Exp. Bot.* **66,** 4061–4073 (2015).
412. Van den Bosch, K. A. & Stacey, G. Summaries of legume genomics projects from around the globe - community resources for crops and models. **131,** 840–865 (2016).
413. Samuel Roberts Noble Foundation. *Medicago truncatula* Mutant Database. at <http://medicago-mutant.noble.org/mutant/>
414. Quandt, H. J. *et al.* Transgenic root nodules of Vicia hirsuta: a fast and efficient system for the study of gene expression in indeterminate-type nodules. *MPMI* **6,** 699–703 (1993).
415. Cheng, X. *et al.* An efficient reverse genetics platform in the model legume Medicago truncatula. *New Phytol.* **201,** 1065–1076 (2014).
416. Wolfram|Alpha - Birthday Problem. *Wolfram Alpha LLC* at <https://m.wolframalpha.com/input/?i=birthday+paradox+50+people&lk=3>