Developmental Regulators of Kranz Anatomy



Olga Sedelnikova

Somerville College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2016

Developmental Regulators of Kranz Anatomy

Olga Sedelnikova, Somerville College

DPhil Thesis, Trinity 2016

Abstract

The C_4 biochemical pathway is the most efficient form of photosynthesis in warm environments and introducing this system into globally significant, but less efficient, C_3 photosynthetic crops could bring major yield increases. The most photosynthetically efficient C_4 grass species have a specialised leaf anatomy characterised by high vein density with two distinct cell types radially arranged around the vascular bundles (Kranz anatomy). Although this anatomy was first described in 1882, the genetic regulators controlling Kranz development are still not known. In recent years, transcriptomic analysis has allowed researchers to identify candidate Kranz regulator genes, and a model for Kranz development has been proposed, however, this model has not been experimentally validated.

This study used *in situ* hybridisation to visualise expression patterns for a set of candidate Kranz regulator genes in maize and the orthologous genes in rice. Further, morphological analysis of rice lines with constitutive expression of the candidate Kranz regulator genes *AINTEGUMENTA 1* and *DAG-LIKE 1* and *2* were used to characterise protein function during C_3 monocot leaf development. The function of maize AINTEGUMENTA in the C_3 eudicot arabidopsis was also investigated.

The results of the *in situ* hybridisation experiments led to the refinement of the Kranz model and identified potential roles for the candidate regulators during leaf development. The constitutive expression experiments highlighted the regulatory differences between eudicot and monocot leaf development and implicated the regulation of auxin-cytokinin homeostatis as a key factor in Kranz development. Ultimately, this work can be used to guide research into Kranz development and has direct implications for engineering C_4 photosynthesis into rice.

Acknowledgements

There were other joys to be found in their company which still more powerfully captivated my mind - the charms of talking and laughing together and kindly giving way to each other's wishes, reading elegantly written books together, sharing jokes and delighting to honor one another, disagreeing occasionally but without rancor . . . We would teach and learn from each other, sadly missing any who were absent and blithely welcoming them when they returned.

— St. Augustine*, The Confessions*[1]

I would like to thank Jane Langdale for all her guidance and support over the last four years, she's a truly inspirational scientist, and a brilliant supervisor. I'd also like to thank Steve Kelly for his advice and computational wizardry. Similarly, David Emms has been a fount of programming knowledge; thank you for putting up with my rudimentary coding questions and helping me with the bioinformatical side of this project.

Working in Jane's lab has also meant that I've had the privilege of being in the C_4 team. I'd like to particularly acknowledge past members Jim Fouracre, Mara Schuler and Peng Wang for all their help whilst I was taking my first steps into the world of rice research. I would also like to thank Dana Vlad for her extensive wet-lab advice and for the thought provoking developmental biology discussions.

The rest of the lab have been a pleasure to work with, and I particularly want to acknowledge the technical staff who've supported my work over the years; Sayuri Ando, Zara Lewis and Rob Pinches. I especially want to thank Julie Bull for being such an amazing lab manager.

[1] Augustine, Boulding, M., & Rotelle, J.E. (1997). *The Confessions*. London: Hodder & Stoughton.

I'd also like to thank my fellow PhD students for their friendship, support and breakfast banter; Emily Seward and Jessie Needham for the Uni Park walks, Eleanor Jaskowska for all the chats that kept me sane, and Tom Hughes for being the best PhD twin anyone could ask for (even if he's only pretending to be from Sheffield).

I'd like to thank my funding bodies; Somerville College, the Clarendon Fund and the Bill and Melinda Gates Foundation for enabling me to study at Oxford. Also, I would like to thank my advisory committee, Lee Sweetlove, Liam Dolan and Javi Augusti for the constructive feedback on my work, and Gem Toes for being an administrative superwoman.

I would like to thank my parents for introducing me into the world of science from an early age; the dinner time conversations about protein purification and days out to see pretty crystals under the microscope have probably shaped me more than I realise. I'm also thankful for having some great friends scattered around the world, but especially want to acknowledge my friends from undergrad for the decade of laughter, support and Christmas dinners. I'd also like to thank the Postgrads community at St Aldate's Church, especially Alison, Kezia, Kate, Eleanor and Daphne for their prayers and friendship.

Above all, I'm thankful to God for His guidance and for faithfully keeping His promises. I pray that this work would reveal something of His character, and would be glorifying to Him. Colossians 3:23-24.

Abbreviations

2,4-D	2,4-dichloro-phenoxyacetic acid
ABA	Abscisic acid
AGO	ARGONAUTE
AHL	AT-HOOK MOTIF NUCLEAR LOCALIZED
AHP6	HISTIDINE PHOSPHOTRANSFER PROTEIN 6
AIL	AINTEGUMENTA-LIKE
ANT	AINTEGUMENTA
AP2-EREBP	APETALA2-ETHYLENE RESPONSIVE ELEMENT BINDING PROTEIN
APL	ALTERED PHLOEM DEVELOPMENT
ARF	AUXIN RESPONSIVE FACTOR
ARR	ARABIDOPSIS RESPONSE REGULATOR
At	*Arabidopsis thaliana*
ATHB	ARABIDOPSIS THALIANA HOMEOBOX
ATL	ABERRANT TESTA SHAPE-LIKE
ATP	Adenosine triphosphate
ATS/KAN4	ABERRANT TESTA SHAPE/KANADI 4
AVB	ABNORMAL VASCULAR BUNDLES
BCIP	5-bromo-4-chloro-indolylphosphate
Bd	*Brachypodium distachyon*
BEP	Bambusoideae, Ehrhartoideae, Pooideae
BES1	BRI1-EMS-SUPPRESSOR 1
BFF1	ROLB DOMAIN B FACTOR 1
bHLH	Basic helix-loop-helix
BIN1	BRASSINOSTEROID INSENSITIVE 2
BLOSUM62	Blocks Substitution Matrix 62
BS	Bundle sheath

bsd	*bundle sheath defective*
bsd-pg	*bundle sheath defective pale green*
BZ	Brassinosteroid
BZR1	BRASSINAZOLE-RESISTANT 1
cDNA	complementary deoxyribonucleic acid
CLE	CLAVATA3/ESR-RELATED
Col-0	*Arabidopsis thaliana* cv. Columbia
CRL5	CROWN ROOTLESS 5
CTAB	Cetyltrimethylammonium bromide
CVL1	COTYLEDON VASCULAR PATTERN 2-LIKE 1
CVP2	COTYLEDON VASCULAR PATTERN 2
CYCD	CYCLIN D
CYP90D2	CYTOCHROME P450-90D2
DAG	DOF AFFECTING GERMINATION
DEPC	Diethylpyrocarbonate
DEX	Glucocorticoid dexamethasone
DIC	Differential interference contrast
DIG	Digoxigenin
DLK	DOF AFFECTING GERMINATION-LIKE
DNA	Deoxyribonucleic acid
DOF	DNA-BINDING WITH ONE ZINC FINGER
DOT	DEFECTIVELY ORGANISED TRIBUTARIES
EDTA	Ethylenediaminetetraacetic acid
EMS	Ethyl methanesulfonate
FCP	FON2-LIKE CLE PROTEIN 1
FP	Foliar plastochron
G2	GOLDEN 2
G3P	Glyceraldehyde-3-phosphate
GAMYB	GIBBERELLIN REGULATED MYB

GDC	Glycine decarboxylase complex
GFP	Green fluorescent protein
GLK1	GOLDEN2-LIKE 1
GO	Gene ontology
GUS	β-glucuronidase
HOX1	HOMEOBOX 1
HP	Husk plastochron
Hv	*Hordeum vulgare*
HYG	HYGROMYCIN RESISTANCE
IAA	Indole-3-acetic acid
ICK1	INTERACTS WITH CDC2 KINASE 1
IDD	INDETERMINATE DOMAIN
IP_3	Inositol 1,4,5-triphosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IQR	Inter quartile range
ITM	Intercellular trafficking motif
JA	Jasmonic acid
Jc	*Jatropha curcas*
KAN	KANADI
L2	Layer 2
LB	Luria-Bertani
LHW	LONELY HIGHWAY
lis1	*large interveinal space 1*
LOG	LONELY GUY
LRR-RLK	Leucine-rich repeat receptor-like kinase
M	Mesophyll
MAPK	MITOGEN ACTIVATED PROTEIN KINASE
MES	2-morpholinoethanesulfonic acid
MIL	MYB-RELATED PROTEIN INTERACTOR-LIKE

MOPS	3-morpholinopropane-1-sulfonic acid
MP/ARF5	MONOPTEROS/AUXIN RESPONSIVE FACTOR 5
mRNA	Messenger ribonucleic acid
MRPI	MYB-RELATED PROTEIN INTERACTOR
MS	Murashige and Skoog
NAA	1-naphthaleneacetic acid
NAC	NO APICAL MERISTEM/ATAF/CUP SHAPED COTYLEDON
NAD-ME	Nicotinamide adenine dinucleotide dependent malic enzyme
NADP-ME	Nicotinamide adenine dinucleotide phosphate dependent malic enzyme
NADPH	Nicotinamide adenine dinucleotide phosphate hydrogen
NAL	NARROW LEAF
NBT	4-nitro-blue-tetrazolium chloride
NKD1	NAKED 1
Nt	*Nicotiana tabacum*
O2	OPAQUE2
OBP2	OCTOPINE SYNTHASE ELEMENT BINDING FACTOR 2
OPS	OCTOPUS
ORF	Open reading frame
Os	*Oryza sativa*
PACMAD	Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, Danthonioideae
PBF	PROLAMIN BOX BINDING FACTOR
PCK	PHOSPHOENOLPYRUVATE CARBOXYKINASE
PEP	Phosphoenolpyruvate
PEPCK	Phosphoenolpyruvate carboxykinase
PHB	PHABULOSA
PHV	PHAVOLUTA
PIN	PIN-FORMED
PLT	PLETHORA
PMC	Procambium mother cell

PNH1	PINHEAD 1
PPDK	Pyruvate phosphate dikinase
Pt	*Populus trichocarpa*
PXY	PHLOEM INTERCALATED WITH XYLEM
RAL1	RADICLELESS
RBCS	RUBISCO SMALL SUBUNIT
RE	Regulatory element
REV	REVOLUTA
RISC	Ribonucleic acid induced silencing complex
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
RNAseq	Ribonucleic acid sequencing
rRNA	Ribosomal ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
RUBISCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose-1,5-bisphosphate
RVN1	RAVEN 1
SAD	SCUTELLUM AND ALEURONE EXPRESSED DOF
SAM	Shoot apical meristem
Sb	*Sorghum bicolor*
SCL23	SCARECROW-LIKE 23
SCR	SCARECROW
semiQ-RT PCR	Semi-quantitative reverse transcription polymerase chain reaction
SFC/VAN3	SCARFACE/VAN3
SHR	SHORTROOT
SHY2/IAA3	SHORT HYPOCOTYL 2/AUXIN-INDUCED AUX/IAA 3
Si	*Setaria italica*
SPCHL1	SPEECHLESS-LIKE 1
SSC	Saline sodium citrate

STY1	STYLISH1
T1, T2 etc.	Transcript 1, transcript 2 etc.
tan1	*tangled 1*
TE buffer	Tris-EDTA buffer
TF	Transcription factor
TMO5	TARGET OF MONOPTEROS
UBI	UBIQUITIN
UTR	Untranslated region
VB	Vascular bundle
WOX4	WUSCHEL-RELATED HOMEOBOX4
WT	Wild type
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
Zj	*Zoysia japonica*
Zm	*Zea mays*

Guide to Nomenclature

Gene and plasmid nomenclature

The gene nomenclature used in this thesis is as follows:

Gene: Italicised upper case e.g. *ANT*

Protein: Non-italicised upper case e.g. ANT

Recessive mutant: Italicised lower case e.g. *ant*

The plasmid nomenclature used in this thesis is as follows:

Promoter: Promoter name followed by subscript 'pro' e.g. $35S_{pro}$

Transcriptional unit: Italicised promoter and gene separated by '*::*'

e.g. $ZmUBI_{pro}$*::ZmANT1*

Truncation: 'Δ' e.g. Δ *ZmANT1*

Hairpin RNAi: Target gene followed by '*i*' e.g. *OsAIL5i*

Transgenic plant nomenclature

Rice

Plant regenerating from transformed callus: T_0

Progeny from a T_0 plant: T_1

Progeny from a T_1 plant: T_2

Arabidopsis

Plant used for seed dipping: T_0

Plants germinated from transformed seed: T_1

Progeny from a T_1 plant: T_2

Contents

Abstract i

Acknowledgements iii

Abbreviations v

Guide to Nomenclature xi

Contents xxii

List of Figures xxiii

List of Tables xxvii

1 Introduction 1

1.1 Utilising C_4 photosynthesis for crop improvement 1

1.1.1 Overview 1

1.1.2 The C_4 biochemical pathway 2

1.1.3 Kranz anatomy facilitates efficient C_4 photosynthesis 4

1.1.4 Evolution of C_4 photosynthesis 8

1.1.4.1 Convergent evolution of C_4 photosynthesis 8

1.1.4.2 Pre-conditioning of anatomical and biochemical traits . . . 10

1.1.4.3 Gene duplication and gene expression changes during C_4 evolution 11

1.2 Leaf development in C_3 and C_4 plants 13

1.2.1 Leaf tissue ontogeny in C_4 monocots 13

1.2.2 Regulators of leaf development 16

1.2.2.1 Procambium initiation 16

1.2.2.2 Vascular bundle development 20

1.2.2.3 Bundle sheath and mesophyll cell differentiation 26

1.3 Identification of Kranz regulators 29

1.3.1 Mutagenesis screens 29

1.3.2 Gene expression studies 30

1.3.3 Identifying candidate Kranz regulators 34

1.4 Thesis aims and objectives **35**

2 Materials and Methods **39**

2.1 General materials and methods **39**

2.2 Plant material and growth conditions **39**

2.2.1 Rice 39

2.2.2 Maize 40

2.2.3 Arabidopsis 40

2.3 Molecular cloning I: Gateway cloning **40**

2.3.1 DNA template 40

2.3.1.1 Genomic DNA 40

2.3.1.2 Complementary DNA 41

2.3.2 Polymerase chain reaction to amplify template DNA 42

2.3.3 DNA gel electrophoresis and gel extraction 42

2.3.4 Plasmid construction 43

2.3.5 Sanger sequencing 44

2.4 Molecular cloning II: Golden Gate assembly **44**

2.4.1 DNA template and level 0 plasmid synthesis 45

2.4.2 Level 1 plasmid ligation 45

2.4.3 Level 2 plasmid ligation 45

2.5 Bacterial transformation **46**

2.5.1 *Escherichia coli* 46

2.5.2 *Agrobacterium tumefaciens* 47

2.5.3 Colony PCR 47

2.6 Plant tranformation **48**

2.6.1 Rice callus transformation 48

2.6.1.1 Tissue culture media 48

2.6.1.2 Rice callus induction 49

2.6.1.3 *Agrobacterium* mediated callus transformation 49

2.6.1.4 Transformed rice callus selection 50

2.6.1.5 Transformed rice callus regeneration 50

2.6.2 Arabidopsis transformation using floral dipping 50

2.6.2.1 Arabidopsis plants for floral dipping 50

2.6.2.2 *Agrobacterium* mediated arabidopsis transformation 51

2.6.3 Identification of transformed plants 51

2.6.3.1 Hygromycin resistance assay to select arabidopsis transformants 51

2.6.3.2 SDS protocol for DNA extraction 51
2.6.3.3 PCR to identify trangenic plants 52
2.6.4 Transgene expression analysis 52
2.6.4.1 RNA extraction and cDNA synthesis 52
2.6.4.2 Reverse-transcription PCR 52
2.7 Morphological analysis 53
2.7.1 Rice leaf sectioning 53
2.7.2 Arabidopsis leaf clearing 53
2.7.3 Measurements and statistical analysis 53
2.8 *In situ* hybridisation 54
2.8.1 RNA probes 54
2.8.1.1 RNA probe design 54
2.8.1.2 Digoxigenin-labelled RNA probe synthesis 54
2.8.1.3 RNA probe DIG-incorporation and hydrolysis efficiency . . 55
2.8.2 Slide preparation and pre-treatment 56
2.8.2.1 Tissue fixation 56
2.8.2.2 Sectioning 57
2.8.2.3 Tissue pre-treatment 57
2.8.3 Hybridisation and post-hybridisation treatments 57
2.8.3.1 Hybridisation 57
2.8.3.2 Post-hybridisation washes 58
2.8.3.3 Antibody staining and fluorescence counter-staining 58
2.8.3.4 Image analysis 58
2.9 *In silico* analyses 59
2.9.1 Phylogenetic inference 59
2.9.2 RNAseq analysis 59
2.9.3 *In silico* promoter region analysis 62
2.9.3.1 Regulatory element identification 62
2.9.4 Gene Ontology term analysis 62

3 Using *in situ* hybridisation to refine models of Kranz development 63
3.1 Introduction 63
3.1.1 Maize transcriptome analysis identified candidate regulators of Kranz anatomy 63
3.1.2 Spatio-temporal control of transcription factor activity is essential for tissue patterning 65

3.1.3 *In situ* hybridisation can resolve the spatio-temporal expression patterns of developmental regulators 65
3.2 Results 66
3.2.1 Transcript profiles of candidate transcription factors distinguish three distinct stages of Kranz development 66
3.2.2 *ZmMIL2* transcript accumulation marks the leaf cell layer that gives rise to vascular tissue 68
3.2.3 Procambial centres are initiated in a single cell in ground tissue . . 68
3.2.3.1 *ZmSHR2*, *ZmANT1* and *ZmMIL1* mRNA accumulates in the procambium mother cell 68
3.2.3.2 Inhibition of procambium initiation correlates with the presence of *ZmDOT1* mRNA 68
3.2.4 Transcripts encoding distinct combinations of transcription factors accumulate in mesophyll versus bundle sheath cells during specification 69
3.2.4.1 *SHORTROOT* transcripts are differentially expressed in bundle sheath cells 69
3.2.4.2 The *INDETERMINATE DOMAIN* genes *ZmNKD1* and *ZmRVN1* are differentially expressed in mesophyll and bundle sheath cells 80
3.2.4.3 *ZmSCR* transcript accumulation patterns change during Kranz cell differentiation 80
3.2.4.4 Expression of a cytokinin-responsive gene in differentiated mesophyll cells 81
3.2.5 Candidate regulator ortholog expression reveals distinct differences between rice and maize leaf development 81
3.2.5.1 Candidate regulator ortholog expression is significantly lower during rice leaf development compared to maize foliar leaves 81
3.2.5.2 *MIL2* transcripts do not localise to undifferentiated ground tissue in rice leaves 88
3.2.5.3 Procambial initiation in rice is confined to regions adjacent to mature veins 88
3.2.5.4 *OsDOT1* transcripts are detected later in rice development than *ZmDOT1* in maize 89
3.2.5.5 Gene expression in procambium differs between maize and rice 89

3.3 Discussion **98**
3.3.1 Model of Kranz anatomy development in maize 98
3.3.1.1 Competence of ground tissue to initiate procambium 98
3.3.1.2 SHORTROOT-SCARECROW pathway in mesophyll and bundle sheath cell patterning 98
3.3.2 Understanding differences between maize and rice gene expression may enable Kranz anatomy engineering in a C_3 context 106
3.3.2.1 Ground tissue proliferation is restricted in rice leaf development 106
3.3.2.2 Key developmental regulators are present during procambium development in maize 106
3.3.2.3 Inhibition of procambium initiation differs in maize and rice 107
3.4 Future Perspectives **107**
3.4.1 MIL protein interaction partners could be key Kranz regulators . . . 107
3.4.2 Downregulation of *DOT1* expression could increase vein density in rice 108

4 The role of *AINTEGUMENTA-like* genes in Kranz development 109
4.1 Introduction **109**
4.1.1 AINTEGUMENTA-like proteins identified as putative Kranz anatomy regulators 109
4.1.2 AIL proteins have developmental functions that are consistent with a role in Kranz development 110
4.1.3 AIL proteins interact with auxin and cytokinin signalling networks . . 112
4.1.4 Functional domains in AIL proteins 113
4.1.5 Kranz regulators are predicted to change rice vein patterning 114
4.2 Results **116**
4.2.1 AtANT orthologs are putative positive Kranz regulators that underwent gene duplication in monocots 116
4.2.2 Verification and correction of monocot *ANT* gene models using RNAseq data and RT-PCR 119
4.2.3 Functional domain differences are seen between C_3 and C_4 monocot ANT1 protein sequences 122
4.2.4 *ZmANT1* transcripts encode both a full-length and a truncated protein with possible regulatory functions 127
4.2.5 Developmental consequences of *ZmANT1* overexpression in rice . 130

4.2.5.1 Transformation of *ZmANT1* into rice led to low rates of regeneration and a high frequency of gene silencing 130
4.2.5.2 Transformation of $\triangle$*ZmANT1* into rice led to successful regeneration and high levels of transgene expression 131
4.2.6 ZmANT1 mediated suppression of leaf blade expansion in rice is dependent on the N' terminal activation domain 134
4.2.6.1 ZmANT1 decreases leaf width without any effect on vein initiation when the N' terminal transcriptional activation domain is present 134
4.2.6.2 ZmANT1 increases bundle sheath cell width and vein size 135
4.2.7 AINTEGUMENTA-like target gene roles include cell growth regulation and RNA modification 143
4.2.8 ZmANT1 is antagonistic to cytokinin signalling in rice leaf development 151
4.2.8.1 ZmANT1 represses kinetin-mediated leaf blade expansion in rice 151
4.2.8.2 ZmANT1 and kinetin are antagonistic in rice vein development 154
4.2.9 The predicted AIL protein targets in arabidopsis leaf development share little orthology with maize targets 156
4.2.10 Preliminary study shows ZmANT1 effects on leaf development in arabidopsis are opposite to those in rice 162
4.2.10.1 *ZmANT1*$_{pro}$*::ZmANT1* transformation into arabidopsis produced multiple transgenic lines and good seed set 162
4.2.10.2 ZmANT1 activity in arabidopsis increases leaf area 165
4.2.10.3 ZmANT1 activity increases vein initiation in arabidopsis . . 165
4.2.10.4 ZmANT1 activity increases vein and bundle sheath width in arabidopsis 168
4.3 Discussion 171
4.3.1 The effect of ZmANT1 in leaf development is not conserved between arabidopsis and rice 171
4.3.1.1 ANT protein role in leaf lamina growth and vein initiation is not conserved between arabidopsis and rice 171
4.3.1.2 ZmANT1 interactions with auxin and cytokinin regulatory networks are not conserved between arabidopsis and rice . 172
4.4 Future perspectives 174

4.4.0.1 ZmANT1 has potential for Kranz anatomy engineering in eudicots . . . 174
4.4.0.2 Role of other *AINTEGUMENTA-like* genes in Kranz development . . . 174
4.4.0.3 ZmANT1 interaction with cytokinin in maize . . . 175

5 The role of *DOF AFFECTING GERMINATION-like* genes in Kranz development 177
5.1 Introduction . . . 177
5.1.1 DAG-like proteins are identified as putative Kranz regulators in monocots and eudicots . . . 177
5.1.2 DAG-like proteins are mobile transcription factors that can act in protein complexes as transcriptional repressors . . . 178
5.1.3 DAG-like proteins are implicated in auxin responses and leaf development . . . 179
5.2 Results . . . 181
5.2.1 The *DAG-like* clade underwent gene duplications in both monocots and eudicots . . . 181
5.2.2 Verification and correction of *ZmDLK1* and *ZmDLK1H* gene models using RNAseq data . . . 182
5.2.3 *ZmDLK1* and *ZmDLK1H* may be regulated by non-coding RNA . . 186
5.2.4 Non-coding RNA is not expressed from the rice *OsDLK1* locus . . . 186
5.2.5 Conserved C_3 versus C_4 differences between monocot DLK1 protein sequences are prevalent in the C' region . . . 186
5.2.6 Developmental consequences of *ZmDLK1* overexpression in rice . 188
5.2.6.1 Transformation of *ZmDLK1* into rice led to low rates of regeneration and poor seed set . . . 188
5.2.6.2 *ZmDLK1* expression retards plant growth but does not perturb vein patterning in rice . . . 190
5.2.7 Developmental consequences of *ZmDLK1H* overexpression in rice 191
5.2.7.1 Transformation of *ZmDLK1H* into rice was not successful . 191
5.2.8 Gene silencing of *ZmDLK1H* in the presence of truncated *GRMZM2G-171855* . . . 191
5.2.8.1 A maize genomic DNA fragment was used to co-transform *ZmDLK1H* and $\triangle$*GRMZM2G171855* . . . 191

5.2.8.2 Co-transformation of *ZmDLK1H* and △*GRMZM2G171855* into rice led to the regeneration of plants through *ZmDLK1H* gene silencing . . . 197
5.2.9 Hormone and light responsive regulatory elements identified in *ZmDLK* putative promoter regions . . . 197
5.2.10 Sub-functionalisation of *DLK1* and *DLK1H* . . . 201
5.3 Discussion . . . 205
5.3.1 ZmDLK1 is a negative regulator of plant growth in rice . . . 205
5.3.2 *ZmDLK1* and *ZmDLK1H* transcript levels regulated by anti-sense RNA . . . 205
5.3.3 Amino acid differences between ZmDLK1 and ZmDLK1H modulate effects on target genes during leaf development . . . 206
5.3.4 *ZmDLK1* and *ZmDLK1H* are expressed at different times during leaf development . . . 206
5.3.5 Hormone signalling regulates *ZmDLK* transcript levels . . . 207
5.3.6 Mis-calculation of transcript levels for *ZmDLK* genes during leaf development due to anti-sense transcripts . . . 208
5.4 Future perspectives . . . 208
5.4.1 The effect of ZmDLK1 and ZmDLK1H on rice bundle sheath development . . . 208
5.4.2 Conserved DLK protein role in eudicots and monocots during Kranz development . . . 209

6 General Discussion 211
6.1 Overview . . . 211
6.2 *In situ* hybridisation and constitutive expression results have progressed understanding of monocot leaf development . . . 213
6.2.1 Regulatory mechanisms underpinning leaf development in eudicots are not conserved in C_3 monocots . . . 213
6.2.2 Differences exist between C_3 and C_4 ground tissue proliferation . . 214
6.2.3 DOT1 could be a key regulator of vein density in monocots . . . 215
6.2.4 SHR and SCR are not sufficient for bundle sheath and mesophyll cell differentiation in C_4 tissues . . . 215
6.3 Further experimental approaches to investigate Kranz development . 216
6.3.1 Transgenic approaches when ectopic expression is detrimental to plant fitness . . . 216

6.3.2 The effect of transcriptional complexes on Kranz development in C_4 species . . . 217
6.3.3 Identifying Kranz regulator developmental networks . . . 218
6.3.4 Auxin and cytokinin homeostasis during C_3 and C_4 leaf development . . . 218
6.4 Research implications for engineering Kranz anatomy in C_3 species . 219
6.4.1 Modulating procambium initiation competency . . . 220
6.4.2 Gene stacking for transcriptional complexes . . . 220

Appendices

A Phylogenetic inference of rice orthologs to candidate Kranz regulators 223
A.1 ANT1 phylogeny . . . 224
A.2 CRF1 phylogeny . . . 225
A.3 DOT1 phylogeny . . . 226
A.4 NKD1 phylogeny . . . 227
A.5 MIL phylogeny . . . 228
A.6 RVN1 phylogeny . . . 229
A.7 SCR1 phylogeny . . . 230
A.8 SHR phylogeny . . . 231

B AINTEGUMENTA-like protein targets and supplementary experiments 233
B.1 AIL protein target genes in maize and arabidopsis . . . 234
B.2 GO term analysis of Arabidopsis AIL protein target genes . . . 234
B.3 Preliminary results: OsAIL5 promotes vein initiation in rice . . . 257
B.3.0.1 *$ZmUBI_{pro}$::OsAIL5i* transformation gave multiple regenerants but low survival rates and seed set . . . 257
B.3.0.2 Surviving *UBI_{pro}::OsAIL5i* lines showed no or low levels of gene silencing . . . 257
B.3.0.3 Suppression of OsAIL5 activity results in a trend for lower vein density . . . 258
B.3.0.4 *OsAIL5* summary . . . 260

C Digital appendix legend 263
C.1 Gene cloning for plant transformation . . . 264
C.1.1 Gene and primer details for Gateway gene cloning . . . 264
C.1.2 Gateway and Golden Gate plasmid details and resistance markers 264

C.1.3 Gateway plasmid Sanger sequencing primers 264
C.1.4 Golden Gate plasmids 264
C.1.5 Colony PCR, genotyping and expression analysis primer pairs . . . 264
C.2 In situ hybridisation 264
C.2.1 *In situ* probe synthesis PCR primers, RNA probe hydrolysis time and SSC wash concentration 264
C.2.2 Tissue-Tek VIP 5 vacuum infiltration processor cycle details 264

References 265

List of Figures

1.1 C_3 photosynthetic reactions 3
1.2 RUBISCO oxygenase activity leads to photorespiration 5
1.3 C_4 photosynthetic pathways increase CO_2 concentration around RUBISCO 6
1.4 Comparison of C_3 and C_4 leaf anatomy 7
1.5 Leaf blade development in maize. 14
1.6 *SHORTROOT-SCARECROW* pathway regulates xylem and endoderm cell differentiation in arabidopsis roots 24
1.7 *SHR-SCR* model for Kranz cell differentiation. 38

2.1 Calcofluor counter-staining to catagorise key stages of vein development . 60

3.1 Summary of transcript localisation patterns for 11 candidate Kranz regulators. 67
3.2 *ZmMIL2* mRNA localisation during maize foliar development 70
3.3 *ZmSHR2*, *ZmANT1* and *ZmMIL1* transcripts localise to the procambium mother cell during Kranz development 72
3.4 Presence of *ZmDOT1* transcript is negatively correlated with procambium initiation in adjacent cells. 74
3.5 *ZmDOT1* transcript accumulates in procambium after an asymmetric cell division near to a mature vein. 76
3.6 *ZmSHR1* mRNA is present in an additional cell layer compared to *ZmSHR2* during Kranz cell differentiation. 78
3.7 *INDETERMINATE DOMAIN* genes *ZmNKD1* and *ZmRVN1* show distinct transcript localisation patterns during Kranz cell differentiation. 82
3.8 mRNA localisation of the homeologs *SCR1* and *SCR1H* differs during Kranz development. 84
3.9 *ZmCRF1* mRNA localised to mesophyll cells during Kranz cell differentiation 86
3.10 *OsMIL2* mRNA localisation during rice leaf development. 90
3.11 Cell divisions in the ground tissue of rice leaves occur predominanty next to mature veins. 92
3.12 *OsDOT1* mRNA accumulation in developing rice leaves continues from early procambium to late stages of procambium development. 94

3.13 *OsANT*, *OsMIL1* and *OsSHR2* mRNA accumulates in vasculature. 96
3.14 Summary of mRNA accumulation of candidate genes during Kranz development in maize 100
3.15 Model for vein initiation during Kranz anatomy development 102
3.16 Model for bundle sheath and mesophyll cell specification through the SHORTROOT-SCARECROW developmental pathway 104

4.1 Functional domains in the AtANT protein. 115
4.2 AINTEGUMENTA-like protein phylogenetic tree showing putative maize Kranz regulators. 117
4.3 Pruned AINTEGUMENTA-like protein phylogenetic tree to highlight AINTEGUMENTA clade. 118
4.4 Mis-annotation of *ZmANT1* gene models in Phytozome. 120
4.5 Confirmation of *ZmANT1* unidentified transcript variant using RT-PCR. . . 121
4.6 Mis-annotation of *OsANT1* gene model in Phytozome. 123
4.7 Functional domain annotation of ZmANT1 using AtANT1 amino acid sequence alignment. 125
4.8 ANT1 amino acid sequence differences between C_3 and C_4 monocots are present in functional domains. 126
4.9 CRL5 amino acid sequence differences between C_3 and C_4 monocots are outside the annotated functional domains. 128
4.10 *ZmANT1* transcript 3 encodes a truncated protein variant (ΔZmANT1) that lacks a transcriptional activation domain. 129
4.11 Composite of gel images showing *ZmUBI$_{pro}$::ZmANT1* genotyping and expression analysis. 132
4.12 Composite of gel images showing *ZmUBI$_{pro}$::ΔZmANT1* genotyping and expression analysis. 133
4.13 Increased vein density in *ZmUBI$_{pro}$::ZmANT1* plants is due to narrower leaves and is dependent on the N' terminal domain. 136
4.14 *ZmUBI$_{pro}$::ZmANT1* plants show a N' terminal dependent decrease in interveinal distance due to fewer mesophyll cells. 138
4.15 *ZmUBI$_{pro}$::ZmANT1* plants show increased bundle sheath width and vein size. 140
4.16 *ZmUBI$_{pro}$::ΔZmANT1* plants do not display bundle sheath cell or vein size differences. 142
4.17 Biological process gene ontology of putative ANT target genes expressed during maize leaf development. 146

4.18 Molecular function gene ontology of putative ANT target genes expressed during maize leaf development. . . . 148
4.19 ZmANT1 inhibits kinetin effects on vein density and interveinal distance in rice . . . 152
4.20 Kinetin-induced mesophyll cell proliferation and bundle sheath size increases are not seen in *ZmANT1* lines. . . . 155
4.21 Kinetin and ZmANT1 are antagonistic in rice vein development. . . . 157
4.22 Composite of gel images showing $ZmANT1_{pro}$:*ZmANT1* arabidopsis genotyping and expression analysis. . . . 164
4.23 Arabidopsis leaf was measured using Leaf Image Interface (LIMANI) software. . . . 166
4.24 $ZmANT1_{pro}$::*ZmANT1* arabidopsis plants have larger leaves compared to wild-type. . . . 167
4.25 $ZmANT1_{pro}$::*ZmANT1* transgenic arabidopsis lines have more vasculature and a higher vein density compared to wild-type. . . . 169
4.26 $ZmANT1_{pro}$::*ZmANT1* transgenic arabidopsis lines show increased vein width. . . . 170

5.1 Functional domains in DAG-like proteins. . . . 180
5.2 Phylogenetic tree of DAG-like proteins . . . 183
5.3 *ZmDLK1* gene model is correctly annotated in Phytozome . . . 184
5.4 Mis-annotation of *ZmDLK1H* gene model in Phytozome. . . . 185
5.5 Mis-annotation of *OsDLK1* gene model in Phytozome . . . 187
5.6 DLK1 amino acid sequence differences between C_3 and C_4 monocots are present in the putative transcriptional activation domain. . . . 189
5.7 Composite of gel images showing $ZmUBI_{pro}$::*ZmDLK1* genotyping and expression analysis. . . . 192
5.8 Constitutive *ZmDLK1* expression retards rice growth. . . . 193
5.9 *ZmDLK1* transgenic rice plant does not display vein patterning purturbations in leaf primordia. . . . 194
5.10 Anti-sense RNA transcription start site is predicted in the *ZmDLK1H* exon genomic region. . . . 196
5.11 Construct to test the effect of *GRMZM2G171855* antisense region on *ZmDLK1H* expression . . . 198
5.12 Composite of gel images showing $ZmUBI_{pro}$::*ZmDLK1H* genotyping and expression analysis. . . . 200

A.1 Phylogenetic tree of RVN1 proteins 224
A.2 Phylogenetic tree of CRF1 proteins 225
A.3 Phylogenetic tree of DOT1 proteins 226
A.4 Phylogenetic tree of NKD1 proteins 227
A.5 Phylogenetic tree of MIL proteins 228
A.6 Phylogenetic tree of RVN1 proteins 229
A.7 Phylogenetic tree of SCR proteins 230
A.8 Phylogenetic tree of SHR proteins 231

B.1 Biological process gene ontology of putative AIL protein target genes expressed during maize and arabidopsis leaf development. 253
B.2 Molecular function gene ontology of putative AIL protein target genes expressed during maize and arabidopsis leaf development. 254
B.3 Biological process gene ontology of arabidopsis orthologs to maize AIL target genes, grouped by conservation of expression and regulation in leaf development. 255
B.4 Molecular function gene ontology of arabidopsis orthologs to maize AIL target genes, grouped by conservation of expression and regulation in leaf development. 256
B.5 Composite of gel images showing *ZmUBI*$_{pro}$*::OsAIL5i* genotyping and expression analysis. 259
B.6 Reduced vein density trend in the *OsAIL5i-3* RNAi rice line. 261
B.7 No significant vascular bundle differences in the *OsAIL5i-3* RNAi transgenic rice line. 262

List of Tables

1.1 Kranz regulator genes proposed to function in procambium development and cell differentiation. 37

2.1 Phusion enzyme PCR conditions 42
2.2 BigDye Terminator v3.1 sequencing reaction PCR conditions 44
2.3 Golden Gate digestion/ligation thermal cycle 46
2.4 GoTaq enzyme PCR conditions 48

3.1 Putative Kranz regulators analysed by *in situ* hybridisation. 64
3.2 Putative Kranz regulator rice orthologs. 87

4.1 *AINTEGUMENTA*-like genes identified as putative Kranz regulators in maize111
4.2 $ZmUBI_{pro}$::*ZmANT1* rice transformation regeneration summary 131
4.3 $ZmUBI_{pro}$::Δ*ZmANT1* transformation regeneration summary 131
4.4 *AINTEGUMENTA-like* protein target genes encode putative Kranz regulators. 144
4.5 *AINTEGUMENTA-like* protein target genes expressed during maize leaf development that encode auxin and cytokinin responsive proteins. 150
4.6 Effects of ZmANT1 and kinetin on rice leaf development 156
4.7 *AINTEGUMENTA-like* protein target genes expressed during arabidopsis leaf development that encode auxin and cytokinin responsive proteins. . . 159
4.8 AIL regulation of auxin and cytokinin responsive genes in maize leaf development is not lkely to be conserved in arabidopsis leaf development . . 161
4.9 $ZmANT1_{pro}$::*ZmANT1* arabidopsis transformation summary 163

5.1 *DOF AFFECTING GERMINATION*-like genes identified as putative Kranz regulators in maize 178
5.2 $ZmUBI_{pro}$::*ZmDLK1* transformation regeneration summary 188
5.3 $ZmUBI_{pro}$::*ZmDLK1H* transformation regeneration summary 199
5.4 Regulatory elements present in *ZmDLK1* and *ZmDLK1H* putative promoter regions. 202
5.5 Regulatory elements present in *ZmDLK1* and *ZmDLK1H* anti-sense transcript putative promoter regions. 203

5.6 Regulatory elements present in *OsDLK1* putative promoter region. 204

B.1 Maize AIL target genes 235
B.2 Arabidopsis AIL target genes 243
B.3 *ZmUBI$_{pro}$::OsAIL5i* transformation regeneration summary 258

1 Introduction

1.1 Utilising C_4 photosynthesis for crop improvement

1.1.1 Overview

Life on earth is underpinned by the conversion of our star's electromagnetic energy into chemical energy by photosynthetic organisms. Humans have harnessed this biochemical reaction through the domestication of plants into crop species (reviewed in Hancock, 2012). However, due to the geographically constrained nature of these domestication events, economically important crops are not necessarily the most efficient in terms of energy capture.

The need for highly efficient crops is rising; global population is predicted to exceed 9 billion by 2050 and the demand for food is predicted to increase by at least 50%. This poses a life-threatening problem as current crop yield increases will not be sufficient to meet the change in demand, therefore a step change in agricultural production is required (FAO, 2009). Crop improvement research is attempting to contribute to such a change by, for example, investigating how to

introduce the most efficient photosynthetic systems into globally significant crops to bring major yield increases (Sage & Zhu, 2011; Ort *et al.*, 2015).

The C_4 biochemical pathway is a highly efficient form of photosynthesis in warm environments, and is the focus of a large collaborative project which aims to engineer the pathway into rice (a crop with the less efficient C_3 photosynthetic pathway). This has the potential to increase rice yields by 50%, however, for the project to be successful, both the biochemical and anatomical features of C_4 plants must be engineered into the crop (Covshoff & Hibberd, 2012).

This introduction will briefly discuss the biochemistry and evolution of the C_4 photosynthetic pathway, and then highlight how specialised leaf morphology (Kranz anatomy) facilitates efficient C_4 photosynthesis. Current hypotheses about the mechanisms involved in Kranz patterning will then be reviewed in the broader context of leaf development. This chapter will go on to consider the approaches used to investigate how C_4 anatomy develops and what is required to engineer this morphology in rice. It will conclude by outlining the experimental strategies used in this thesis to identify regulators of Kranz anatomy and will summarise the specific research aims.

1.1.2 The C_4 biochemical pathway

Several different photosynthetic systems are found in nature; the most common is termed C_3 photosynthesis (fig. 1.1; p.3). In this pathway, RUBISCO (Ribulose-1,5-bisphosphate carboxylase/oxygenase) enzyme catalyses the fixation of CO_2 into organic carbon (through carboxylation) in the light-independent reactions. However, the enzyme also displays oxygenase activity, this catalysis leads to a reaction cycle termed photorespiration that is costly in terms of energy (fig. 1.2; p.5) (reviewed in Leegood *et al.*, 1995). RUBISCO oxygenase activity increases at higher temperatures and O_2:CO_2 ratios (Sharkey, 1988).

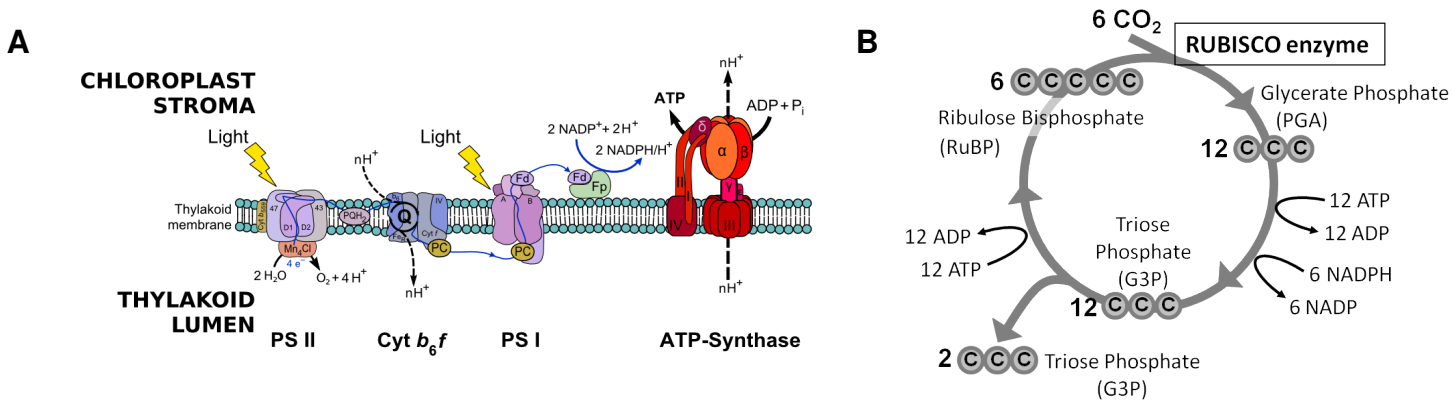


Figure 1.1: C_3 photosynthetic reactions. In C_3 plants, leaf mesophyll cells are enriched with photosynthetic chloroplasts. C_3 photosynthesis occurs in these organelles through a series of reactions. **A)** Light dependent reactions act to convert light energy into chemical energy using proteins embedded in the thylakoid membrane. Photons are absorbed by chlorophyll molecules in the photosystem II (PSII) and photosystem I (PSI) protein complexes. This causes electron excitation and transport along a series of acceptor molecules (blue path). The electron flow causes protons to be transported across the thylakoid membrane and leads to NADPH and ATP synthesis. Electrons in PSII are replenished through the oxidisation of water. For further detail see Taiz *et al.* (2015). **B)** Light independent reactions in chloroplast stroma are used to fix CO_2 into organic carbon. CO_2 diffuses into the chloroplast stroma where the RUBISCO enzyme catalyses a carboxylation reaction between CO_2 and a five-carbon molecule (RuBP). The NADPH and ATP produced by the light-dependent reactions (A) are used to convert the resultant six-carbon molecule into glycerate phosphate (PGA) and then triose phosphate (G3P). For each twelve G3P made, two are converted into sucrose in the cytosol of the cell, whilst the others are used to regenerate RuBP in the chloroplast. (Taiz *et al.*, 2015). Images adapted from Taiz *et al.* (2015) and Leegood *et al.* (1995).

In the C_4 photosynthetic pathway, photorespiration is greatly reduced because the O_2:CO_2 ratio is modulated to increase CO_2 levels around RUBISCO (fig. 1.3; p.6). This change in ratio is due to the spatial separation of CO_2 capture from RUBISCO activity. These systems are characterised by two morphologically distinct chloroplast types that have different functions; light dependent reactions occur in one chloroplast type (that does not contain RUBISCO), meanwhile the other chloroplast type does contain RUBISCO so can perform light independent reactions (reviewed in Laetsch, 1974). The most efficient C_4 plants use a two cell-type tissue pattern termed Kranz anatomy. In this system, the CO_2 is carboxylated into a four-carbon sugar in a mesophyll (M) cell before being transported to a bundle sheath (BS) cell (where RUBISCO containing chloroplasts are localised) and decarboxylated back into CO_2. Therefore, the O_2:CO_2 ratio around the RUBISCO enzyme is significantly decreased and photorespiration is minimised.

A range of biochemical pathways for C_4 have been documented. For example, three decarboxylating enzymes are known; PEPCK (phosphoenolpyruvate carboxykinase) , NAD-ME (NAD-malic enzyme) and NADP-ME (NADP-malic enzyme) (Berry *et al.*, 1970; Björkman, 1968; Downton, 1970), and the four-carbon sugar can be malate or aspartate (Downton, 1970). Although these biochemical processes require energy (3 ATP and 6 NADPH, per G3P produced), the cost of the C_4 pathway is favourable compared to the cost of photorespiration in temperatures above 25°C (reviewed in Kajala *et al.*, 2011).

1.1.3 Kranz anatomy facilitates efficient C_4 photosynthesis

The interplay between metabolism and leaf anatomy in C_4 photosynthesis was first clearly elucidated by Hatch (1987). However, Kranz anatomy (characterised by two cell types radially arrange around a vein; fig.1.4, p.7) has been documented since the 1880s (Haberlandt, 1882). More recently, a single-cell C_4 pathway was

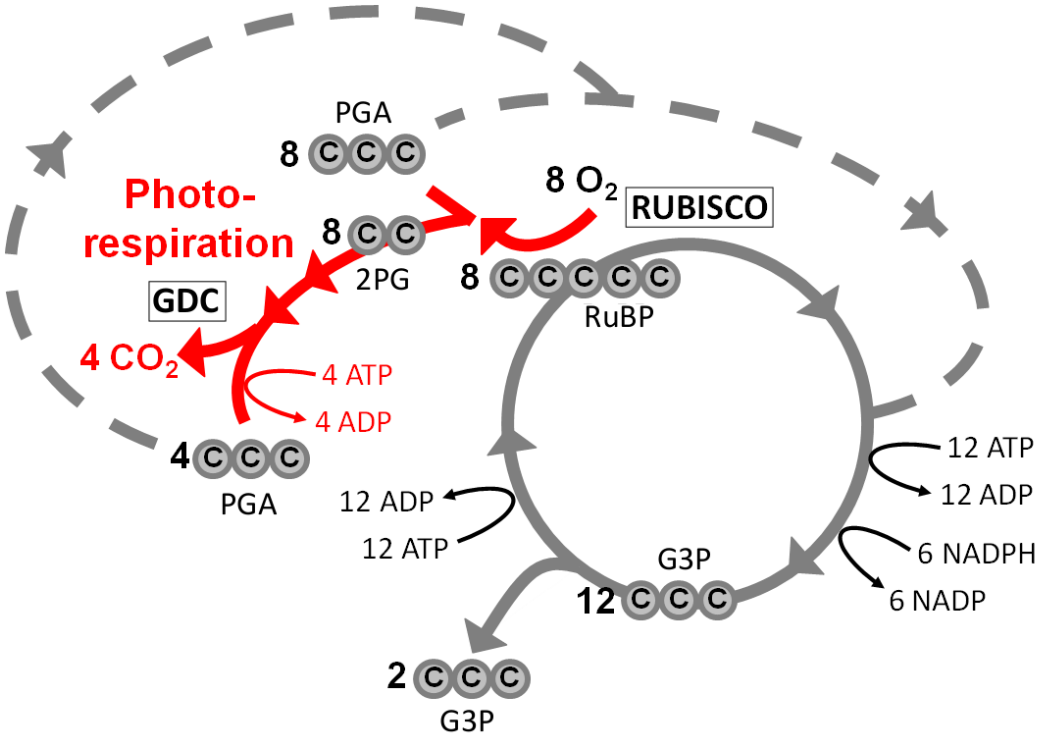


Figure 1.2: RUBISCO oxygenase activity leads to photorespiration. The RUBISCO enzyme can carboxylate RuBP leading to the light independent reactions and the production of sucrose (solid grey lines). However, it is also capable of oxygenase activity where an oxygen is added to RuBP, leading to photorespiration (solid red lines). In this process a glycerate phosphate (PGA) molecule is produced that is fed into the light-independent reactions (dashed grey line), but in addition, a two-carbon molecule (2-phosphoglycolate [2PG]) is made. 2PG is converted into PGA through an energy intensive process that involves glycine decarboxylation by the glycine decarboxylase complex (GDC) and the release of previously fixed carbon in the form of CO_2. Image adapted from Leegood *et al.* (1995).

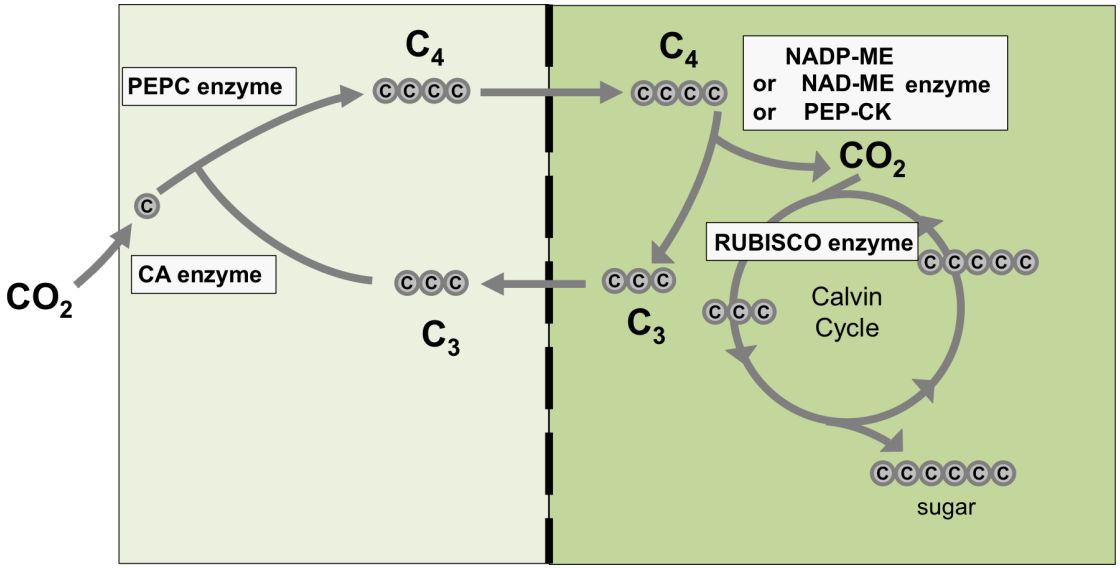


Figure 1.3: C_4 photosynthetic pathways increase CO_2 concentration around RUBISCO. Photorespiration can be minimised by concentrating CO_2 around RUBISCO. In C_4 pathways, this is facilitated by spatial separation of RUBISCO from ambient conditions and a CO_2 transport system in the form of a four-carbon sugar. Ambient CO_2 is converted to hydrogen carbonate by carbonic anhydrase (CA), and this molecule is incorporated with the three-carbon (C_3) molecule PEP (phosphoenolpyruvate) into a four-carbon (C_4) molecule by phosphoenolpyruvate carboxylase (PEPC). The C_4 molecule is transported to an area where RUBISCO is localised, and is decarboxylated to release CO_2 and a C_3 molecule by phosphoenolpyruvate carboxykinase (PEPCK), NAD-malic enzyme (NAD-ME) or NADP-malic enzyme (NADP-ME). The CO_2 released is used in the light-independent reactions initially through the carboxylase activity of RUBISCO. PEP is regenerated through the enzymatic activity of PPDK (pyruvate phosphate dikinase) which is co-localised with PEPC. This spatial separation of CO_2 capture (light green) and RUBSICO activity (dark green) can occur within one cell or between different cell types (reviewed in Edwards & Voznesenskaya, 2011). Image adapted from Edwards & Voznesenskaya (2011).

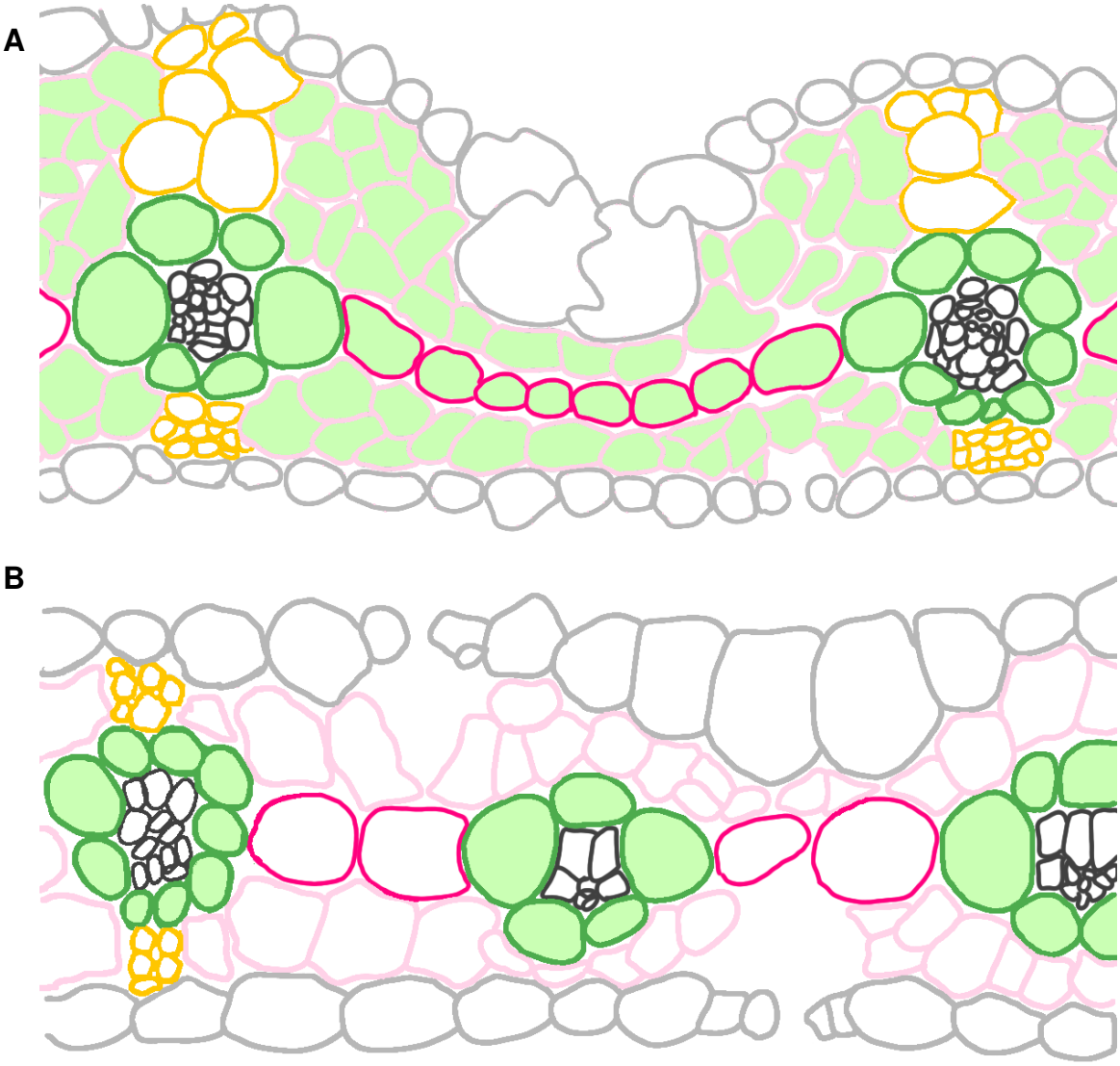


Figure 1.4: Comparison of C_3 and C_4 leaf anatomy. A-B) Diagrams showing transverse cross-sections of a mature rice (C_3) (A) and maize (C_4) (B) leaf. Cells are outlined - upper and lower epidermis (grey), sclerenchyma (yellow), vasculature (black), bundle sheath (BS) (dark green) and mesophyll (M) (light and dark pink). The middle layer of M cells (dark pink) highlights the difference in cell number between veins in C_3 and C_4 species. There are also differences in RUBISCO enzyme localisation (light green). In C_4 plants, BS cells can also have suberised cell walls (Furbank *et al.*, 1989). Image adapted from Wang *et al.* (2014b).

discovered in *Borszczowia aralocaspica* (Voznesenskaya *et al.*, 2001). In this system RUBISCO is compartmentalised to a distinct area in the cell so that the PEPC (phosphoenolpyruvate carboxylase) enzyme is spatially separated from NAD-ME. Further, two different types of chloroplasts are seen within one cell. This is in contrast to Kranz anatomy, where the enzyme separation and chloroplast dimorphism occur between two cell types (Brown, 1958). Although the one-cell system shows that the C_4 pathway can occur within a single cell type, highly specialised cell morphology is required to give correct compartmentalisation and an efficient C_4 pathway (Voznesenskaya *et al.*, 2001). For example, the system requires long cells to optimise RUBISCO sequestration and this kind of physiological constraint may limit the use of the one-cell system for engineering C_4 into C_3 crops. In contrast, there appear to be fewer physiological constraints on Kranz anatomy because many anatomical variations of the two-cell type system have been reported in both eudicots and monocots (reviewed in Edwards & Voznesenskaya, 2011). More specifically, C_4 grasses display variations in the two-cell type radial patterning around veins (Brown, 1975), as well as different BS and M cell sizes, for example, enlarged BS cells are common, but not universal. The level of anatomical diversity between C_4 species, taken together with the diversity in biochemical pathways, is indicative of convergent evolution and multiple C_4 evolutionary origins (reviewed in Nelson, 2011).

1.1.4 Evolution of C_4 photosynthesis

1.1.4.1 Convergent evolution of C_4 photosynthesis

The evolution of C_4 photosynthesis has been extensively studied; the earliest isotope evidence for C_4 has been dated to $\sim$30 million years ago (Urban *et al.*, 2010), which corroborates with phylogenetic analysis (Christin *et al.*, 2008; Vicentini *et al.*, 2008). Phylogenetic analysis has also identified many independent

instances of C_4 evolution from C_3 ancestors; recent analysis estimates 66 distinct C_4 lineages, with 26 of these lineages in the monocots (Grass Phylogeny Working Group II, 2012; reviewed in Sage *et al.*, 2012). In monocots, Kranz anatomy is found in the Poaceae and Cyperaceae families, with the majority of these events in the Poaceae (Brown, 1975; reviewed in Edwards & Voznesenskaya, 2011). Although, it is difficult to ascertain whether all the lineages are truly independent, the high number of inferred events indicates that C_4 traits are an example of convergent evolution (Christin *et al.*, 2009). Lineages displaying intermediate C_3 - C_4 photosynthetic traits have also been documented, however only two grass clades have been identified (Kennedy & Laetsch, 1974; reviewed in Sage & Zhu, 2011). The clustering of C_4 evolution events to a relatively small number of clades suggests that the C_3 to C_4 transition may require certain genetic characteristics to be present in the C_3 ancestor to enable C_4 evolution (Sage & Zhu, 2011).

C_3-C_4 intermediates have been used to infer the key changes that occur during C_4 evolution. Given the number of biochemical and morphological differences between C_3 and C_4 species, it has long been proposed that the evolution of C_4 occurred through a series of intermediate states (Laetsch, 1971). However, it was not known whether the identified intermediate species were on the evolutionary trajectory towards C_4 photosynthesis or were evolutionary 'dead-ends' (Sage *et al.*, 2012). In recent years, several groups have modelled the evolutionary trajectory from C_3 to C_4 photosynthesis. For example, Heckmann *et al.* (2013) calculated the fitness landscape between C_3 and C_4 photosynthetic states using a biochemical model based on carbon fixation. Notably, they found that the biochemical characteristics of the known C_3 - C_4 intermediates map to the predicted evolutionary trajectories towards C_4 photosynthesis. A different modelling approach was used by Williams *et al.* (2013), this group conducted a meta-analysis to quantify C_4 characteristics in C_3 - C_4 intermediates and model C_4 evolution. Significantly, this model predicted similar evolutionary trajectories to the biochemical model.

These studies suggest that known C_3 - C_4 intermediates are on the evolutionary trajectory towards C_4 photosynthesis and confirm the validity of using these species to infer the evolution of C_4 photosynthesis.

1.1.4.2 Pre-conditioning of anatomical and biochemical traits

Research on C_3-C_4 intermediates, alongside the modelling studies, suggests that a series of changes to the C_3 ancestral state are necessary for C_4 evolution. A key biochemical pre-adaptation for C_4 appears to be the evolution of C_2 photosynthesis, a system that involves carbon transport (in the form of glycine) from M to BS cells. In species that show C_2 photosynthesis, the glycine decarboxylase complex (GDC) is localised to BS cells. Glycine is transported from M to BS cells where GDC-mediated decarboxylation causes the release of CO_2 (see fig. 1.2; p.5). The decreased O_2:CO_2 ratio in the BS cell leads to decreased photorespiration rates and increased photosynthetic efficiency (Rawsthorne *et al.*, 1988; Vogan *et al.*, 2007; reviewed in Schulze *et al.*, 2016). Importantly, there is evidence that C_2 photosynthesis may enable the evolution of C_4 photosynthesis through a direct mechanistic process. In plant leaves, as well as decarboxylation, GDC also catalyses deamination of glycine, releasing NH_3 which is re-fixed in chloroplasts. This means that in C_2 plants there is potential for a nitrogen imbalance between M and BS cells and therefore selective pressure for the evolution of an efficient nitrogen shuttle from BS and M cells. Metabolic modelling showed that a pathway including reactions found in C_4 photosynthesis could give one of the most efficient shuttling mechanisms. Significantly, in *Flavaria* C_3-C_4 intermediates, the expression levels of non-C_4 genes that were predicted to have a role in nitrogen shuttling correlated with increased C_4 traits (Mallmann *et al.*, 2014). These findings indicate that changes in the regulation of BS and M cell gene expression are a key early step in C_4 evolution.

Notably, it has been hypothesised that pre-conditioning of leaf anatomy may enable the evolution of C_2, and therefore C_4 photosynthesis. For example, relatively larger BS cells, fewer M cells and decreased interveinal distance have been correlated with the transition to C_4 photosynthesis in grasses (Brown *et al.*, 1983; Christin *et al.*, 2013). It is postulated that this anatomical pre-conditioning may enable BS-specific localisation of the GDC (Christin *et al.*, 2013). Interestingly, these anatomical traits are commonly found in the PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, Danthonioideae) clade, whereas members of the BEP (Bambusoideae, Ehrhartoideae, Pooideae) clade (including globally important crops such as rice and wheat) usually show the opposite traits (decreased BS size and increased M cell number) (Christin *et al.*, 2013). The monocot studies correlate with research on eudicots, for example, high vein density is a common feature in the C_3 relatives of C_4 species in many genera, including *Cleome* (a genus closely related to arabidopsis) (Marshall *et al.*, 2007; Voznesenskaya *et al.*, 2007). Similarly, C_3 species displaying enlarged 'activated' BS cells with centripetally arranged chloroplasts and GDC-containing mitochondria which are localised near to vascular bundles (key features of C_2 and C_4 BS cells) are found in several genera, including *Flaveria* and *Euphorbia* (Sage *et al.*, 2013, 2011b). These anatomical features appear to be important transition states along the evolutionary trajectory towards C_4 photosynthesis (reviewed in Sage *et al.*, 2014), and understanding the genetic changes that cause these pre-adaptations may enable the engineering of C_4 traits in C_3 crops.

1.1.4.3 Gene duplication and gene expression changes during C_4 evolution

It has been hypothesised that the duplication of genes important for C_4 biochemistry and morphology is a further key pre-adaptation that enables C_4 evolution

(Monson, 2003). Ohno (1970) proposed that novel traits may arise through neo-functionalisation of protein function following gene duplication whereby one of the duplicated genes, free of selective constraints, accumulates mutations. Most of the gene duplicates become silenced (Lynch & Conery, 2000), but rarely, a mutation may confer a new protein function that becomes fixed through selection or genetic drift. Subfunctionalisation of protein function has also been proposed (Hughes, 1994). In this model, the functions of the original protein are divided between the two duplicated genes. It is also possible that duplicated genes can retain the ancestral function, but have altered expression patterns, leading to novel traits (reviewed in Moore & Purugganan, 2005). There is some evidence that duplications of genes encoding catalytic enzymes and transporters are associated with C_4 evolution (Monson, 2003; Christin *et al.*, 2015; Emms *et al.*, 2016) and it is possible that gene duplication is also a key pre-adaptation for Kranz anatomy development.

In contrast to this hypothesis, it has been postulated that regulatory changes to genes in the C_3 ancestor are more significant to the evolution of C_4 than gene duplication. Comparisons of C_3 and C_4 Cleomaceae species found that the transcriptional regulation of C_4-implicated genes was strongly correlated to C_4 traits, but that the copy number of these genes was not a significant factor (van den Bergh *et al.*, 2014). These results corroborate other studies on expression patterns in Cleomaceae species, where C_4-implicated genes were found to be expressed in C_3 species, but with altered expression domains (Bräutigam *et al.*, 2011; Külahoglu *et al.*, 2014). These studies highlight the importance of regulatory changes during the evolution of C_4 photosynthesis, however, the hypothesis that regulatory changes are more significant than gene duplications is only based on research into one eudicot group, and it is possible that gene duplication played a more significant role in enabling C_4 evolution in other clades.

Overall, research into C_4 evolution highlights that anatomical pre-adaptations

correlate with the evolutionary trajectory towards C_4 photosynthesis. These morphological changes may enable the evolution of C_2 photosynthesis which is a key intermediate state in the evolution of C_4 photosynthesis. The development of these anatomical features may be due to changes in gene copy number and/or transcriptional regulation. This therefore implies that understanding the genetic control of leaf development in C_3 and C_4 plants could increase our understanding of Kranz anatomy evolution and C_4 photosynthesis.

1.2 Leaf development in C_3 and C_4 plants

1.2.1 Leaf tissue ontogeny in C_4 monocots

The key features of Kranz anatomy (for example, high vein density, increased BS:M cell ratio and chloroplast dimorphism between BS and M cells) are all patterned during the early stages of leaf development. The ontogeny of early maize foliar development has been well characterised from the earliest stages of leaf development. These studies show that the non-epidermal foliar cells derive mostly from cell divisions in the subepidermal (L2) region of the shoot apical meristem (SAM) (Poethig & Szymkowiak, 1995). These cells then divide to give three layers of ground tissue from which the Kranz anatomy differentiates (Langdale *et al.*, 1989). Veins develops from cells in the middle layer of the developing leaf in distinct stages (fig.1.5; p.14). The first stage of cellular differentiation is termed procambium initiation (vascular procambium is a partially differentiated tissue that develops from undifferentiated ground tissue) (Smith, 1934; Sharman, 1942; Esau, 1943). Procambium then differentiates into the various vascular tissues (Esau, 1943). In maize and other monocots that have a single layer of BS cells around veins, the BS cells also differentiate from the procambium tissue (Dengler *et al.*, 1985; Sakaguchi & Fukuda, 2008).

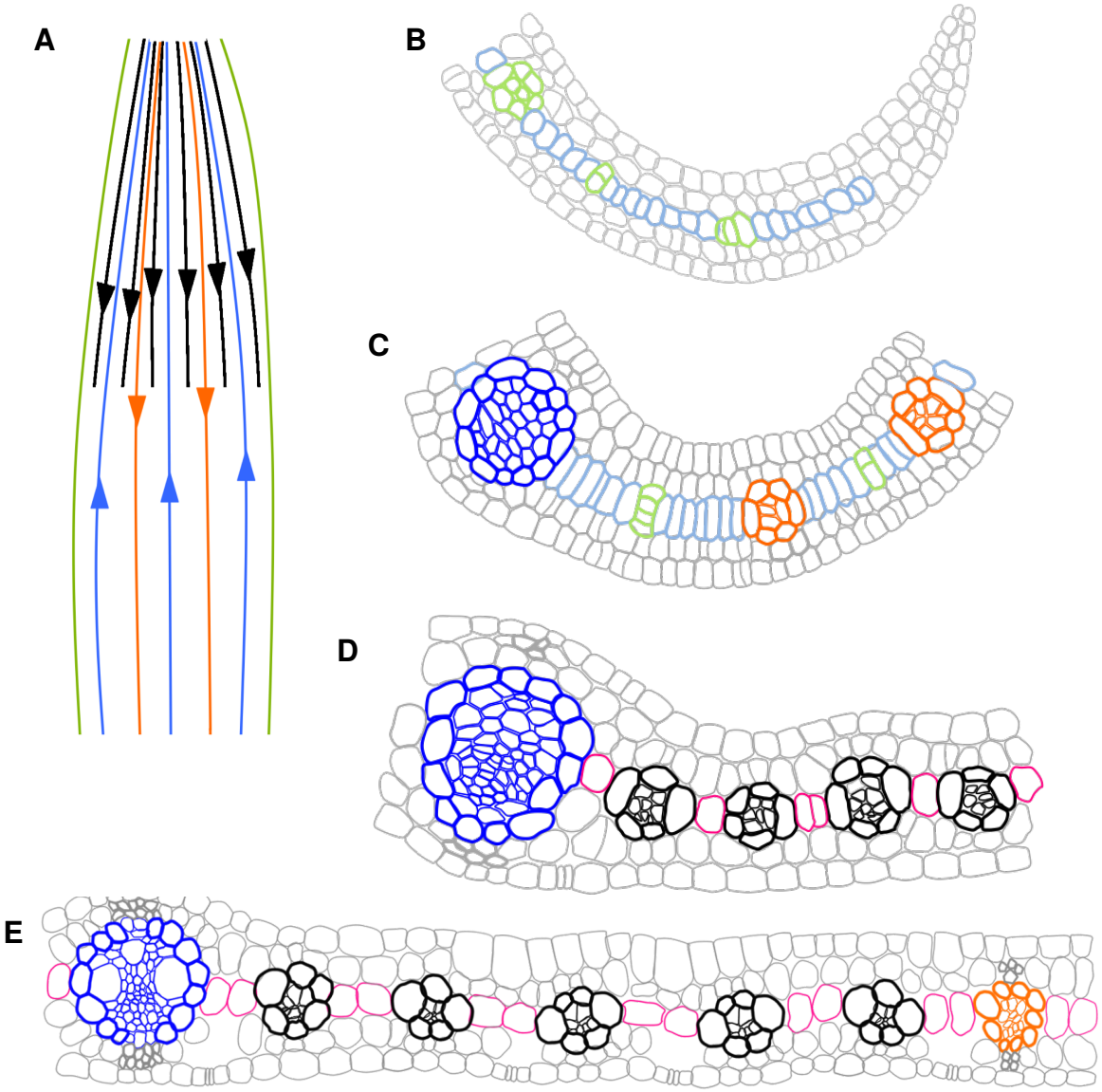


Figure 1.5: Leaf blade development in maize. Figure legend is on p.15.

Figure 1.5: Leaf blade development in maize. A) Direction of vein growth in developing maize blade (diagram has leaf base towards the bottom of the page)- acropetal growth of mid-vein and lateral veins (blue), basipetal growth of rank 1 (orange) and rank 2 (black) intermediate veins. **B-D)** Diagrams of transverse leaf cross-sections showing developmental stages of maize blade tissue patterning. Procambium initiation (green) occurs from undifferentiated cells in the middle layer of ground tissue (light blue) (A), procambium tissue differentiates into various cell types to give vascular bundles, including bundle sheath (BS) cells, meanwhile, additional procambium is initiated from ground tissue (C). At the latter stages of leaf development, pre-cursor mesophyll (M) cells (middle layer highlighted dark pink) divide to give the BS-M-M-BS radial tissue patterning seen in Kranz anatomy (D). **E)** Diagram of transverse cross section of mature maize leaf showing lateral vein (blue), rank 1 (orange) and rank 2 (black) intermediate veins, and the middle layer of mesophyll cells (pink). Images adapted from Wang *et al.* (2016) and T. Hughes (personal communication).

Maize leaves have several orders of vascular bundles that follow a conserved developmental pattern. The mid-rib vascular bundle (VB) develops first with procambium differentiation occurring from the leaf base to the tip (acropetal growth). This is followed by the acropetal differentiation of large lateral VBs. The mid-rib and laterals extend through both the leaf sheath and leaf blade at maturity. In the leaf blade, interspersed between lateral bundles are two different ranks of smaller, later developing intermediate veins which initiate at the leaf tip and grow down to the base of the developing leaf blade (basipetal growth) (fig.1.5, A; p.14). These intermediate veins merge (anastomise) in the ligule tissue between sheath and blade (Sharman, 1942).

1.2.2 Regulators of leaf development

1.2.2.1 Procambium initiation

Procambium initiation correlates with auxin maxima within the developing monocot or eudicot leaf. These maxima result from regulatory networks that control the localisation of auxin efflux transporters (PIN-FORMED; PIN proteins) within the ground tissue thereby localising auxin to specific cell files (canalisation) (Sachs, 1981; Scarpella *et al.*, 2002, 2006). Several genes have been implicated in the auxin regulatory network during arabidopsis leaf vein patterning. For example, *COTYLEDON VASCULAR PATTERN 2* (*CVP2*) and *CVP2-LIKE 1* (*CVL1*) act to modulate inositol 1,4,5-triphosphate (IP_3) signalling (Carland & Nelson, 2009). This signalling network regulates a complex containing *SCARFACE* (*SFC*/*VAN3*), which has a role in auxin efflux regulation (Sieburth *et al.*, 2006; Koizumi *et al.*, 2000). These action of the encoded proteins, therefore, contributes to the control of auxin maxima localisation.

Several other proteins have been associated with the earliest stages of procambium initiation in arabidopsis leaves. For example, MONOPTEROS (MP, ARF5)

acts as a transducer of the auxin signal to activate procambium related gene expression (Mattsson *et al.*, 2003) in part by activating *ATHB8* (*ARABIDOPSIS THALIANA HOMEOBOX 8*) expression. This HD-ZIP III transcription factor (TF) marks pre-procambial cell identity (Donner *et al.*, 2009) and acts to maintain cell multipotency (meristematic competence) in developing procambium (Kang & Dengler, 2002). Interestingly, *SHORTROOT* (*SHR*), a gene encoding a GRAS TF which is a key regulator of root tissue patterning (reviewed in Fisher & Sozzani, 2016), is co-expressed with *ATHB8* during procambium initiation and development in the arabidopsis leaf (Gardiner *et al.*, 2011). Other gene products also localise to pre-procambial tissues. For example, several *DOF* (*DNA-BINDING WITH ONE ZINC FINGER*) genes (*DOF2.1*, *DOF4.6* and *DOF5.3*) are expressed in developing arabidopsis leaves and mRNA localisation patterns overlap with *ATHB8* in pre-procambial cells, although the genes are also expressed in ground tissue. As procambium develops, *DOF* gene expression becomes restricted to the vasculature, however the roles of the proteins during vein development are not known (Gardiner *et al.*, 2010). Overall, procambium initiation is dependent on the localisation of auxin and the activation of downstream gene regulatory networks mediated by TFs, however, the regulatory networks involved in leaf procambium initiation have not been well characterised.

Auxin is antagonistic to cytokinin signalling, and given that feedback loops between the two hormones are important for regulating many aspects of plant organ development (reviewed in Chandler & Werr, 2015), cytokinin may also play a role in procambium initiation. However, to date that role appears to be indirect through the regulation of cell proliferation. For example, high cytokinin levels maintain cell proliferation in the arabidopsis shoot meristematic tissues and ground leaf tissue (Werner *et al.*, 2003), and the hormone is important for maintaining procambium cell proliferation during root vascular development (Mähönen *et al.*, 2000; Bishopp *et al.*, 2011). Intriguingly, there is evidence that

ground tissue proliferation is necessary for procambium initiation in arabidopsis leaves; over-expression of the negative cell-cycle regulator *ICK1* led to decreased cell proliferation and less vein initiation, and in contrast, constitutive expression of the *AP2-EREBP* (*APETALA2-ETHYLENE RESPONSIVE ELEMENT BINDING PROTEIN*) *AINTEGUMENTA* (*ANT*) increased cell proliferation and vein initiation (Kang *et al.*, 2007). Thus, although there are links between cytokinin signalling and ground tissue proliferation, little is known about the interaction between cytokinin and auxin during procambium initiation in arabidopsis leaves.

Mutagenesis studies have been conducted to identify other genes involved in arabidopsis vein patterning. Koizumi *et al.* (2000) identified mutations in six genes that led to perturbed patterning, but only the function of *VAN3* has been characterised (see above). More recently, Petricka *et al.* (2008) identified 13 mutated *DEFECTIVELY ORGANISED TRIBUTARIES* (*DOT*) genes that perturbed vein patterns. Eight of these genes were analysed and found to encode a range of proteins, including three TFs; a CCHC zinc finger, a SART-1 leucine zipper and a C2H2 zinc finger protein. Interestingly, the C2H2 protein (DOT5/WIP6) is a member of the WIP clade, a family that includes developmental regulator proteins which contain transcriptional repressor domains (Mitsuda & Ohme-Takagi, 2009) and form complexes with other TFs (Appelhagen *et al.*, 2011). Transcriptomics approaches have also been used to discover genes involved in procambium initiation. For example, Gandotra *et al.* (2013) sequenced RNA from procambial and provascular cells and identified 94 TFs present in the developing procambium. Most of the genes identified in these studies have not been functionally characterised, but given the effects on veins, many of these genes will have roles in regulating procambium initiation or development.

Procambium initiation in monocots is less well understood than in arabidopsis. Interestingly, no genes have been implicated in maize procambium development, however, two rice mutants with perturbed initiation have been identified. In *ral1*

(*radicleless1*) rice plants, procambium initiation is delayed and there is reduced expression of *OsHOX1* (a gene encoding a positive regulator of auxin transport during vascular development). The function of RAL1 is not known, but *ral1* plants displayed increased sensitivity to cytokinin, indicating that the protein may be involved in balancing cytokinin vs. auxin-mediated responses during procambium initiation (Scarpella, 2003). The *avb* (*abnormal vascular bundles*) rice mutant line also displays reduced procambium initiation, apparently due to decreased ground tissue cell proliferation. Given that *AVB* expression is upregulated by auxin, it was postulated that AVB (a protein of unknown function) acts in an auxin-dependent manner to promote procambium initiation (Ma *et al.*, 2016). Given that ground tissue proliferation is promoted by high cytokinin levels (Werner *et al.*, 2003), it is possible that AVB could play a role in modulating auxin hormone levels through a negative feedback loop. Overall, these studies suggest that the interplay between auxin and cytokinin may be important for procambium initiation in rice.

Rice *PINHEAD 1* (*OsPNH1*) has also been implicated in procambium initiation. Gene knock-down experiments using RNAi give rice lines that have perturbations in procambium initiation and vein positioning (Nishimura *et al.*, 2002). *PNH1* is expressed in pre-procambium tissue as well as in the SAM where it acts to maintain meristem cell proliferation. The gene is a member of the *ARGONAUTE* (*AGO*) clade, a family that contains genes involved in RNA silencing through RISC (RNA-induced silencing complex) (Yang *et al.*, 2013) and therefore this mechanism may play a role during procambium initiation in rice. Interestingly, *OsPNH1* expression in veins is downregulated by FON2-LIKE CLE PROTEIN 1 (FCP1) (Ohmori *et al.*, 2014). FCP1 also downregulates another gene that displays vein expression patterns as well as having a role in meristem maintenance (*WUSCHEL-RELATED HOMEOBOX4*: *WOX4*). Notably, WOX4-mediated meristem maintenance appears to be due to the upregulation of cytokinin levels (Ohmori *et al.*, 2013) and the arabidopsis ortholog is a positive regulator of cell

division in the procambium (Ji *et al.*, 2010). Two other WUSCHEL-related homeobox genes *NARROW LEAF 2* and *3* (*NAL2* and *NAL3*) have also been implicated in procambium initiation, in that a double mutant rice line showed decreased procambium initiation compared to wild-type, whilst over-expression increased leaf width and intermediate vein number (Ishiwata *et al.*, 2013). Interestingly, weak *NAL2* gene expression has been observed in veins, and strong signal was seen at the leaf tips (Cho *et al.*, 2013; Zeng *et al.*, 2016), an area that has been implicated in meristematic activity (Hagemann & Gleissberg, 1996). These findings suggest that maintenance of meristematic competence in ground leaf tissue may be required for procambium initiation in rice.

1.2.2.2 Vascular bundle development

After procambium is initiated, the procambial cells divide and differentiate to form vascular bundles. The genetic networks regulating these processes have been studied in arabidopsis embryos, roots and shoots, but less in known about leaf vein development and very little research has been conducted in monocot leaves. The tissue structures within the root, shoot and leaf vascular bundles are similar (for example, xylem, phloem and companion cells), however, the locations of the structures within the bundle are different (Sharman, 1942; Esau, 1943). It is therefore likely that gene networks controlling cell differentiation are conserved between vascular development in roots and shoots, and leaf vein development, but that the specific spatio-temporal regulation of the genes differs.

The development of a vascular bundle requires procambium proliferation. During arabidopsis embryo development, procambium cell proliferation is positively regulated in a cytokinin-mediated process by *LONELY GUY* (*LOG*) genes, these encode cytokinin biosynthetic enzymes (Tokunaga *et al.*, 2012). The *LOG* genes are regulated by basic helix-loop-helix (bHLH) TFs LONESOME HIGHWAY (LHW)

and TARGET OF MONOPTEROS (TMO5) (De Rybel *et al.*, 2013). Interestingly, *LHW* and *TMO5* are upregulated by auxin signalling through MP activity (Schlereth *et al.*, 2010). In vascular procambium, long meristematic cells show periclinal division patterns to produce files of cells, these division patterns are regulated by CLE (CLAVATA3/ESR-RELATED) peptide signalling (Strabala *et al.*, 2006). CLV3/ESR1-LIKE 41, 42 and 44 (CLE41, CLE42 and CLE44) interact with the leucine-rich repeat receptor-like kinase (LRR-RLK) PHLOEM INTERCALATED WITH XYLEM (PXY) (Etchells & Turner, 2010; Zhang *et al.*, 2016). CLE-PXY signalling acts to increase the expression of genes such as *WOX4* and *WOX14* and these gene products act to promote procambium cell division (Hirakawa *et al.*, 2010; Etchells *et al.*, 2013). Interestingly, there is evidence that *WOX4* is also regulated by a PXY-independent mechanism (summarised in Etchells *et al.*, 2012). Ethylene signalling through AP2/ERF (APETALA2/Ethylene Responsive Factor) proteins can also act to regulate cell division in procambium tissues. For example, *ERF109* and *ERF018* expression is required to promote procambium cell division in arabidopsis *pxy* mutant plants. It is proposed that ethylene/ERF signalling acts in parallel to CLE-PXY signalling, and is repressed by WOX4 activity (Etchells *et al.*, 2012). The cell proliferation in procambium tissue forces older cells towards the periphery of the vascular bundle where the cells differentiate into xylem and phloem tissues.

Protophloem is the first tissue type to differentiate at the periphery of the bundle during vascular development (Rodriguez-Villalon *et al.*, 2014). Protophloem differentiation is regulated through the inhibitory action of CLE45-BARELY ANY MERISTEM 3 (BAM3) signalling (Rodriguez-Villalon *et al.*, 2014), and promoted by the MYB protein ALTERED PHLOEM DEVELOPMENT (APL) (Bonke *et al.*, 2003). The APL TF acts to upregulate *NO APICAL MERISTEM/ATAF/CUP-SHAPED COTYLEDON 45* and *86* (*NAC45* and *NAC86*), and these NAC TFs promote the morphogenesis of sieve elements in the phloem tissue (Furuta *et al.*, 2014).

There is evidence that brassinosteroid (BZ) signalling is also involved in phloem development. OCTOPUS (OPS), a plasma membrane-associated protein that displays polar localisation in cells (at the shoot-ward end of the cell) (Truernit *et al.*, 2012), positively regulates BZ signalling by sequestrating BRASSINOSTEROID INSENSITIVE 2 (BIN2) protein at the plasma membrane, preventing BIN2 nuclear localisation. The repression of BIN2 allows nuclear accumulation of TFs BRI1-EMS-SUPPRESSOR 1 (BES1) and BRASSINAZOLE-RESISTANT 1 (BZR1), these proteins regulate gene expression to promote protophloem differentiation (Anne *et al.*, 2016). Interestingly, *CVP2* (see 1.2.2.1; p.16) is also implicated in phloem differentiation, downstream of *OPS*. This indicates that IP_3 signalling is another important regulatory pathway. Significantly, *CVP2* overexpression partially rescues the knockout mutant phenotype of another gene implicated in phloem differentiation: *BREVIS RADIX* (*BRX*) (Rodriguez-Villalon *et al.*, 2015). Similarly to OPS, the BRX protein also shows polar localisation on cell plasma membrane, but it is localised on the root-ward end and translocates to the nucleus upon auxin treatment (Scacchi *et al.*, 2009). *BRX* expression is activated in an auxin-dependent manner by MP and there is evidence that the protein mediates a brassinosteroid and auxin signalling feedback loop in the arabidopsis root (Mouchel *et al.*, 2006; Scacchi *et al.*, 2010). Given the known role of CVP2 and IP_3 in regulating auxin efflux, CVP2 may act as part of this BZ-IAA regulatory feedback system. BRX also acts in cytokinin-auxin signalling in the boundary region between protophloem and procambium tissues. Cytokinin signalling induces *SHORT HYPOCOTYL 2/AUXIN-INDUCED AUX/IAA 3* (*SHY2/IAA3*) expression (a gene that does not appear to be auxin responsive). The SHY2 protein acts to downregulate *PIN* expression, and inhibit auxin signalling, this therefore inhibits *MP* and *BRX* expression and promotes procambium cell identity. SHY2 and BRX1 act in negative feedback loop between auxin and cytokinin because *SHY2* expression requires *BRX* expression (Scacchi *et al.*, 2010). Taken together, these

data show that cross-talk between several signalling networks is required for correct development of phloem tissue, with brassinosteroid, cytokinin and auxin playing key roles during vascular development.

Auxin signalling has also been implicated in xylem differentiation for several decades (Jacobs, 1952). More recently, it was shown that an inhibitory feedback loop between auxin and cytokinin instigated by the LHW/TMO5 regulatory pathway is necessary for the localisation of xylem tissue within the vascular bundle (Bishopp *et al.*, 2011). This feedback loop is produced because the LHW/TMO5 complex activates *HISTIDINE PHOSPHOTRANSFER PROTEIN 6* (*AHP6*) . This gene encodes an inhibitory pseudophosphotransfer protein which acts to inhibit cytokinin signalling. *AHP6* expression localises to cell files on the periphery of the vascular bundle that subsequently form protoxylem (Mähönen *et al.*, 2006). The xylem-procambium boundary also appears to be regulated by the cytokinin inducible genes *AT-HOOK MOTIF NUCLEAR LOCALIZED 3* and *4* (*AHL3* and *AHL4*), these genes encode mobile proteins that localise to the boundary region between xylem and procambium (Zhou *et al.*, 2013). Alongside auxin-cytokinin signalling, the differentiation of the root vascular tissues into protoxylem and metaxylem is also regulated by the *SHORTROOT-SCARECROW* (*SHR/SCR*) pathway (fig. 1.6; p.24. The expression of several *HD-ZIP III* genes (*PHABULOSA*, *PHB*; *PHAVOLUTA*, *PHV*; *REVOLUTA*, *REV*; *ATHB8* and *ATHB15*) promotes metaxylem cell differentiation (Prigge *et al.*, 2005; Kim *et al.*, 2005). These genes are negatively regulated by miRNA165/6 in a SHR/SCR mediated pathway (Carlsbecker *et al.*, 2010), and by KANADI (KAN) proteins through the inhibition of *PIN1* expression and therefore the regulation of auxin transport (Ilegems *et al.*, 2010). These studies again highlight the importance of auxin-cytokinin cross talk during vascular development, and show that genes important for procambium development are also active later in vascular development, enabling cell specification.

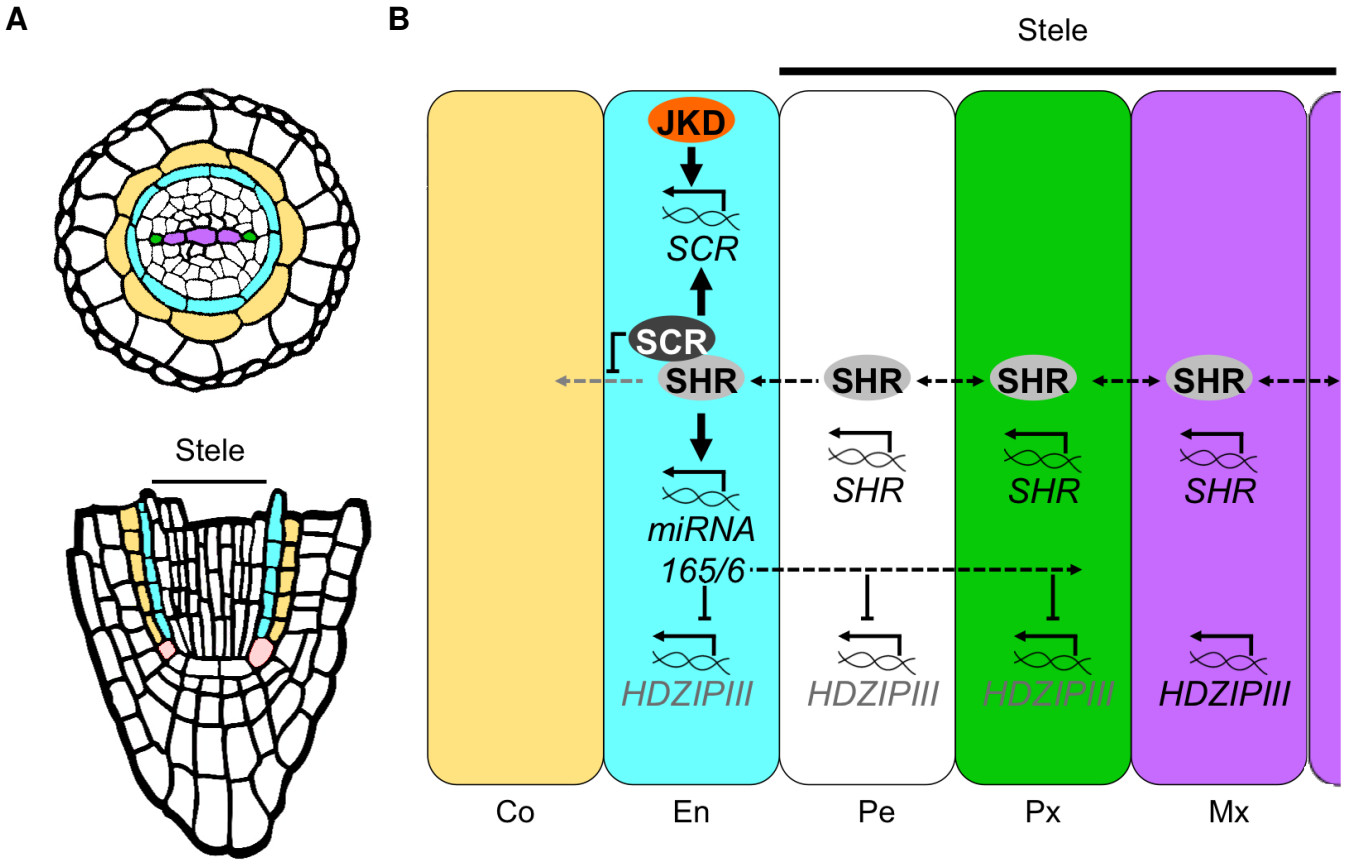


Figure 1.6: *SHORTROOT-SCARECROW* pathway regulates xylem and endoderm cell differentiation in arabidopsis roots. Figure legend is on p.25.

Figure 1.6: *SHORTROOT-SCARECROW* pathway regulates xylem and endoderm cell differentiation in arabidopsis roots. A) Arabidopsis root diagrams showing transverse (top) and longitudinal (bottom) cross-sections. The cortex (yellow), endodermis (blue), protoxylem (green) and metaxylem (purple) tissues and the cortex/endodermis initial cell (pink) are shown. **B)** The GRAS transcription factors SHORTROOT and SCARECROW (SHR and SCR) regulate cell differentiation of the xylem and endodermal tissue. *SHR* is expressed in the stele, SHR (light grey) is mobile and moves out of the vasculature to surrounding cells. In this diagram, *SHR* expression is shown in the pericycle (Pe), protoxylem (Px) and metaxylem (Mx) cells layers, but it is also expressed in other cell files. *SCR* expression is activated by INDETERMINATE DOMAIN (IDD) proteins, including JACKDAW (JKD, orange), which is expressed in the endodermis (En) cell file. The SCR protein (dark grey) acts to restrict SHR movement by sequestering the protein in the nucleus, where SHR/SCR act as a transcriptional complex. Px and Mx cell identities are specified by activation of *miRNA165/166* expression. A concentration gradient of this mobile RNA molecule is formed across the stele, and the miRNA targets *HD-ZIP III* RNA molecules (*PHABULOSA*, *PHAVOLUTA*, *REVOLUTA*, *ATHB8* and *ATHB15*) for degradation in the Pe and Px cell files. In the Mx cell file, the level of *miRNA165/166* is low, so *HD-ZIP III* RNA molecules are not degraded and can specify Mx cell identity (reviewed in De Rybel *et al.*, 2016). The radial tissue patterning of cortex (Co) and En cells is also regulated by SHR/SCR. In the En, the complex activates genes that confer En cell identity, and the absence of the SHR/SCR complex in the Co cell file allows Co-specifying gene expression (Di Laurenzio *et al.*, 1996; Helariutta *et al.*, 2000; Ogasawara *et al.*, 2011). **Ovals**- protein. **Helices and hooked arrows**- gene expression, **dashed lines-black**: protein movement, **grey**: inhibition of protein movement. Image adapted from Van Norman *et al.* (2011) and Ogasawara *et al.* (2011).

1.2.2.3 Bundle sheath and mesophyll cell differentiation

Gene products expressed in developing vascular bundles also affect differentiation of cells in surrounding tissues. Similarly to the role in root cortex cell specification, the *SHORTROOT-SCARECROW* pathway (fig. 1.6; p.24) has been implicated in arabidopsis BS cell differentiation. *SCR* and *SCARECROW-LIKE 23* are expressed in BS precursor cells, and act in concert with SHR moving out of the vascular bundle to specify BS identity (Cui *et al.*, 2014). This is in agreement with Kinsman & Pyke (1998) who noted that BS cell differentiation in arabidopsis was position-specific, as opposed to lineage specific. The role of SHR/SCR in BS differentiation may also be conserved in maize. Slewinski *et al.* (2012) reported that a *scr* maize mutant showed some perturbations in Kranz patterning. This is significant as a hypothesis for Kranz BS/M cell differentiation was proposed by Langdale & Nelson (1991) which postulated a mobile signal from the veins acting with BS- and M- cell specific factors to control BS and M cell identity; this mobile signal may be SHR.

Interestingly, a conflicting cell-lineage dependent hypothesis for maize BS cell identity has been proposed from the study of the *tangled 1* (*tan1*) mutant; this line displays disorganised vascular bundles and BS cells that are not adjacent to a vein. The BS cells are formed in clusters or chains originating from one precursor cell and therefore the authors suggest a lineage based mechanism for BS cell identity (Jankovsky *et al.*, 2001). Given that morphological analysis of maize vein development suggests BS cells differentiate from procambium tissue (Sakaguchi & Fukuda, 2008), the *tan1* BS cell phenotype may be due to changes in *SHR* expression overlaid with perturbations in procambium and ground tissue cell division patterns. There is evidence that the ontogeny of rice BS cells differs from maize, in that rice BS cells differentiate from ground tissue (Sakaguchi & Fukuda, 2008). Given that maize bundle sheath cells show morphological and

expression differences compared to C_3 BS cells (Langdale *et al.*, 1988b), it is possible that TF network differences in the originating tissue (procambium versus ground tissue) may enable C_4 BS cell differentiation during Kranz development.

Developmental differences in BS and M cell differentiation give rise to cells with differences in metabolism and chloroplast morphology in both C_3 and C_4 plants (Kinsman & Pyke, 1998; Leegood, 2008). Notably, arabidopsis mutagenesis screens have identified a 'reticulated' class of mutants, plants with normal BS differentiation, but abnormal M cell chloroplast development. The existence of reticulated mutants shows that there are regulatory differences between M and BS chloroplast development in C_3 plants and therefore functional characterisation of these mutants has been conducted to study the control of cell specification in the arabidopsis leaf (Lundquist *et al.*, 2014).

Interestingly, signals from vascular bundles and BS cells appear to be important for the development of mesophyll chloroplasts. Most of the cloned reticulated mutants encode plastid-localised proteins that are implicated in primary metabolism, however, all the promoter-reporter experiments that have been conducted show vein or BS-specific expression patterns. Two models for mesophyll development have been proposed; the signalling hypothesis proposes that a metabolic signal from the vein acts to promote mesophyll differentiation (Knappe *et al.*, 2003; Voll *et al.*, 2003). Alternatively, the supply model hypothesises that the supply of metabolites from vascular tissue to M cells enables normal development. Many of the mutant loci encode genes involved in the shikimate pathway or in the synthesis of purine and pyrimidine nucleotides (reviewed in Lundquist *et al.*, 2014). Strikingly, the shikimate pathway is implicated in the biosynthesis of several developmentally significant metabolites, for example, auxin (Ljung, 2013). Further, purine metabolism is involved in cytokinin biosynthesis (Mok & Mok, 2001). Intriguingly, the reticulated mutant *dov1* (*differential development of vascular associated cells 1*) was found to have significantly lower levels of cytokinin, possibly

due to perturbations in purine metabolism, and it is possible that this hormonal alteration contributed to the mutant phenotype (Rosar *et al.*, 2012). Overall, these results suggest that vein and/or BS cell activity regulates M cell development.

Reticulated mutants have not been reported in monocots, however, maize mutants that have BS-specific perturbations in chloroplast development have been identified (*bundle sheath defective*: *bsd* mutants). For example *bsd2* maps to a locus that encodes a protein required for RUBISCO assembly in BS cells (Brutnell *et al.*, 1999), and a *bsd-pale green* (*bsd-pg*) mutant has been identified, however the mutated locus has only been resolved down to a genomic area that contains 30 possible gene candidates (Pan *et al.*, 2015). Pertinently to the regulation of BS cell differentiation, a TF has also been identified. *bsd1-m1* has BS-specific chloroplast development perturbations (Langdale & Kidner, 1994) and this mutated locus encodes the MYB-related GARP TF GOLDEN 2 (G2). *G2* expression is localised to BS cells in maize foliar blade tissue (Jenkins, 1926; Hall *et al.*, 1998; Rossini *et al.*, 2001) and further study showed that G2 regulates BS-specific chloroplast development in C_4 leaf tissue. Interestingly, in C_3-like maize tissues, G2 affects chloroplast development in both M and BS cells (Cribb *et al.*, 2001). The related *GOLDEN2-LIKE 1* (*GLK1*) gene shows M-cell specific expression in maize blade tissue so is hypothesised to regulate M-cell specific chloroplast development in this tissue (Rossini *et al.*, 2001). Notably, based on phylogenetic and expression analysis in other C_3 and C_4 species, it is hypothesised that differential expression of *G2* and *GLK1* in BS and M cells may have enabled C_4 evolution through sub-functionalisation of GLK protein activity between the two cell types (Wang *et al.*, 2013a).

Overall, these data show that current understanding of the genetic network regulating BS and M cell differentiation during Kranz development is sparse. To engineer a two-cell type C_4 system into rice, the identification of Kranz regulators is essential. In recent years, a range of approaches have been used to discover

the key genes involved in anatomy development, from mutagenesis screens to gene expression analysis using next generation sequencing techniques.

1.3 Identification of Kranz regulators

1.3.1 Mutagenesis screens

Large-scale mutagenesis screens for C_4 mutants displaying C_3 characters have been reported, however there has been very limited success in identifying Kranz regulators. Ethyl methanesulfonate (EMS) mutagenesis of *Panicum maximum* (a C_4 grass with Kranz anatomy) produced several lines with abnormal inter-vein spacing and one mutant (*large interveinal space 1*, *lis1*) displayed a significant increase in the CO_2 compensation point (a high CO_2 compensation value is characteristic of C_3 photosynthesis) and increased vein spacing (Fladung, 1994), however, there has been no further work with this line.

Large-scale screens using *Sorghum bicolor* populations treated with EMS or γ-irradiation have also been analysed. In this case, two mutations were identified that showed decreased vein density due to an increase in M cell size and number. Both these mutations mapped to the *CYTOCHROME P450-90D2* (*CYP90D2*) gene which encodes a BZ biosynthesis enzyme, suggesting that BZ signalling may be involved in Kranz vein patterning. However, the mutant blade tissue was also thicker, which may indicate a sheath-like identity, therefore it is proposed that the perturbations in BZ signalling resulted in a shift of domain identity from blade to sheath in the foliar leaf (Rizal *et al.*, 2015).These results show that, although these kinds of screens can give useful data, the identification and characterisation of the mutated loci can be difficult and potentially expensive (Gilchrist & Haughn, 2010). This disadvantage, when combined with the expense and time involved in setting up and screening the large populations, has led to some

researchers utilising approaches based on C_3-C_4 gene expression differences to identify candidate regulators.

1.3.2 Gene expression studies

During the elucidation of the C_4 biochemical pathway, it was noted that the photosynthetic enzymes had already been identified in other (C_3) plant metabolic pathways (Hatch & Slack, 1966). This feature of the C_3-C_4 transition means that regulatory and functional differences between ancestral C_3 and derived C_4 forms can be investigated by examining known C_4 genes and their C_3 orthologs (reviewed in Aubry *et al.*, 2011). For example, *RUBISCO small subunit* gene (*RbcS*) expression is BS cell-specific in maize, but M cell-specific in rice. Using a promoter-GUS (β-glucuronidase) construct transformed into maize lines, Nomura *et al.* (2000) showed that the rice promoter drove M-cell specific expression, whilst the maize promoter gave BS-cell specificity. More recently, a nine-nucleotide motif has been reported from the C_4 eudicot *Gynandropsis gynandra* which causes M-cell specific mRNA accumulation when it is present in the 5′ untranslated region (UTR) of a transcript (Williams *et al.*, 2016). These studies show that differences in the promoter regions of genes are important for regulating BS and M cell specific protein localisation. Recent systems biology studies have widened this approach. For example, *in silico* comparison of *cis*-regulatory regions upstream of known C_4 proteins and their C_3 orthologs identified several motifs that were correlated with C_4 expression patterns (Hibberd & Covshoff, 2010; Xu *et al.*, 2016). These types of data can be used to infer key changes during the evolution of C_4 and potentially identify regulatory sequences that can be used in engineering C_4 rice. However, as noted previously, very little is known regarding the regulators of cellular patterning in C_4 plants and therefore it is premature to use this approach for studying Kranz development.

Transcriptome analysis has recently come to the forefront for investigating C_3-C_4 differences and identifying candidate C_4 developmental regulator genes. Several studies have been conducted over recent years to investigate expression differences between species, developmental stages and tissue types (reviewed in Huang *et al.*, 2016; Fouracre *et al.*, 2014). For example, changes in gene expression along the maize leaf proximo-distal axis have been investigated by Li *et al.* (2010). mRNA samples were extracted from four points along a fully expanded maize leaf and illumina sequencing was used to analyse transcript abundance. Perhaps unsurprisingly, the group found that transcript levels of many genes changed along the leaf gradient and that genes could be categorised into clusters by shared expression profiles. Pick *et al.* (2011) conducted a similar study but increased the resolution by sampling at ten positions. Both studies found that TFs were most highly expressed at the base of the leaf (which is the youngest tissue), whilst photosynthetic genes were expressed in older tissues towards the tip. These data are valuable as they provide expression profile for specific genes and it may be possible to identify genes involved in particular biological processes occurring in the C_4 leaf on the basis of these profiles. Unfortunately, these studies are not particularly useful for the identification of Kranz regulator candidates because the earliest developmental stages used in both studies already displayed differentiated Kranz anatomy.

Li *et al.* (2010) also analysed differences in expression between mature BS and M maize cell. This cell-type comparison has been the focus of other studies using illumina sequencing (Chang *et al.*, 2012; Tausta *et al.*, 2014) and microarray analysis (Sharpe *et al.*, 2011). Expression patterns in BS and M cells have also been compared in mature *Panicum virgatum* and *Sorghum bicolor* leaves (Rao *et al.*, 2016; Döring *et al.*, 2016). These experiments are useful as they show at a genome-wide level which of the transcripts are differentially expressed between the two cell types, unfortunately, analysis of mature cells has limited value in identifying

candidate developmental regulators. However, as well as sampling mature cells, Sharpe *et al.* (2011) and Tausta *et al.* (2014) also investigated expression changes during the latter stages of BS and M cell differentiation. These data are a useful resource for engineering C_4 rice, as although they cannot provide information regarding the early events leading to Kranz anatomy differentiation, the regulation of later stages of cell development can be inferred. For example, cell wall synthesis occurs after cell expansion, so the transcriptome analyses could identify regulatory networks involved in the suberisation of bundle sheath cells. Further, some of the early-acting regulators that influence cell-type differentiation may still be expressed. This means that these data could prove valuable in the identification of candidate regulators. Significantly, 14 BS-specific and 2 M-specific transcription factors were identified as candidate BS or M cell-type regulators in both illumina studies. The BS-specific TFs include two *DAG*-like (*DOF AFFECTING GERMINATION-LIKE*) *DOF* genes, as well as *ATHB HD-ZIP III* TFs and a *GLK* gene. Interestingly, the two M cell TFs identified were *GLK1* and a *JACKDAW-LIKE IDD* gene. The latter gene implicates the SHR/SCR pathway in M cell differentiation (Tausta *et al.*, 2014). Overall, these studies have identified candidate TFs that may be involved in regulating BS and M cell differentiation in maize

Studies comparing expression differences between closely related C_3 and C_4 species may also yield valuable information regarding transcriptional changes during the C_3-C_4 transition. Both monocot and eudicots have been used for these studies. For example, Bräutigam *et al.* (2011) compared expression between a C_3 and a C_4 species in *Cleome*, and found transcript abundance differences in 603 transcripts, including 43 transcripts assigned as having regulatory functions. *Flavaria* species have also been studied, Gowik *et al.* (2011) analysed two C_3 and C_4 species and one C_3-C_4 intermediate species and found 183 TFs that clustered in C_4-specific groups. However, mature leaf tissue was used in this study, so the data are unlikely to contain information that would help select

candidates for early Kranz regulators. In contrast, Külahoglu *et al.* (2014) used a leaf developmental series comparing a C_3 and C_4 species of *Cleome* to investigate early differences in transcript abundance. Interestingly, this study found that TFs were expressed for longer in developing C_4 tissues compared with C_3 and that mesophyll differentiation was delayed. This may suggest that the C_4 leaves have a prolonged developmental window for vein initiation competency. This dataset has also been re-analysed by Huang *et al.* (2016), who highlighted a set of vein development related TFs that are differentially expressed between the C_3 and C_4 species. Similarly to other studies, genes related to the SHR/SCR pathway were identified as candidate regulators of C_4 leaf development.

More distantly related C_3 and C_4 species have also been compared. For example, Wang *et al.* (2014a) analysed transcript abundance and metabolite differences between developing maize and rice leaves with 118 TFs identified as C_4 regulatory candidates. In this experiment, the BS and M cells were not separated and samples analysed did not include early stages of Kranz development, so the candidate regulators identified are likely to act in later stages of cell differentiation. However, many of the genes reported as candidate TFs in this study were also identified in the BS and M cell transcriptome analyses described above.

Some plants display both C_3- and C_4-like tissues (usually in different tissue types or at different developmental stages). Comparing gene expression in developing C_3 and C_4 tissues in one species can prevent phylogenetic noise, and so give robust data on changes in transcription occurring during the transition from C_3 to C_4 photosynthesis. For example, the eudicot *Salsola soda* has cotyledon leaves that show C_3 photosynthesis, whilst adult leaves use C_4. Lauterbach *et al.* (2016) analysed genome-wide gene expression at five developmental stages during *Salsola* development, from cotyledons to mature leaves. Given that *Salsola* cotyledons have C_3-like vein patterning and adult leaves show Kranz anatomy, this analysis could identify candidate Kranz regulators. 87 TFs were identified

as candidates, these included genes related to the SHR/SCR pathway, many genes implicated in auxin responses and several AP2-EREBP TFs. Similarly, in C_4 monocots, maize has C_4 anatomy in foliar blade tissue, but also displays C_3-like anatomy in husk leaves (Langdale *et al.*, 1988a). Wang *et al.* (2013b) sampled foliar and husk tissue at five developmental stages, two of which were before the morphological development of Kranz anatomy. This approach yielded data that provides a key resource for Kranz regulator identification in monocots.

1.3.3 Identifying candidate Kranz regulators

To identify candidate Kranz regulators, Wang *et al.* (2013b) analysed genome-wide transcript differences between developing maize husk and foliar leaf tissues. A series of filters were used to identify candidate positive and negative Kranz regulators. For example, for a candidate positive regulator, transcript levels had to be significantly higher in foliar than husk leaves at the stage at which the tissue was undergoing procambium differentiation. In total 283 putative positive and 142 putative negative regulators were identified. Ribosomal, cytoskeleton-related and housekeeping genes were removed from this filtered set, producing a candidate list of 48 putative positive regulators and 23 putative negative regulators.

Based on orthology to known developmental regulators, and on expression profiles during Kranz development and in BS and M cells, Fouracre *et al.* (2014) proposed a subset of twelve genes as key candidate Kranz regulators. Seven were proposed to function in procambium development, and five in BS/M cell differentiation (table 1.1; p.37). The authors proposed that the procambium associated genes were likely to have several different roles in plant development, but that they had been co-opted into Kranz development. For BS and M cell differentiation, a mechanistic model was proposed based on the SHR/SCR pathway and on recognition that the IDD genes *ZmRAVEN1* (*ZmRVN1*) and *ZmNAKED1* (*ZmNKD*) are differentially expressed in BS and M cells, respectively. The model

postulates that cell identity is conferred through the combined action of SHR-SCR complexes and cell-specific factors in each cell type (fig. 1.7; p.38). However, this model has not yet been tested experimentally.

Given that the proteins highlighted in Fouracre *et al.* (2014) may act in more than one developmental pathway during leaf development, analysis of spatial and temporal expression patterns in young leaf primordia could prioritise or downgrade the key Kranz candidates further, as well as giving evidence that could support or refute the proposed model. Once the candidate genes are prioritised, their potential role in Kranz anatomy must be analysed. Transgenic approaches to produce gain and loss of function lines are powerful tools for characterising candidate gene function. For example, over-expressing a positive Kranz regulator in a C_3 plant should give lines with C_4-like characters if the encoded protein is sufficient to induce Kranz development. Conversely, suppressing gene activity in a C_4 plant should give lines with C_3-like characters if the gene is necessary for the regulation of Kranz development.

1.4 Thesis aims and objectives

Given the potential of C_4 photosynthesis for improving plant yields, understanding the development of Kranz anatomy may enable this photosynthetic system to be engineered in globally important C_3 crops. The work in this thesis begins in chapter 3 by investigating expression patterns of a subset of candidate Kranz regulators using *in situ* hybridisation on maize shoot apices.These experiments aim to visualise the spatio-temporal dynamics of RNA accumulation, giving information that can be used to build hypotheses about the role of the candidate genes during Kranz development. Although the testing of these hypotheses is outside the scope of this thesis, the data are used to refine the current hypothetical model of Kranz development. Chapter 4 goes on to explore the function of one of these

candidate maize genes, *AINTEGUMENTA* (*ZmANT1*). The effect of the gene on rice and arabidopsis leaf development is tested using a transgenic approach. *ZmANT1* is constitutively expressed in the two species and vein patterning is phenotyped to investigate if the gene confers Kranz-like traits in the C_3 plant leaves. Putative ANT target genes are also identified, using *in silico* promoter region analysis for maize and arabidopsis. These data are used to investigate potential differences in ANT-mediated developmental pathways between the species. Lastly, the interaction of ZmANT1 with cytokinin signalling is investigated in rice, using phenotypic analysis of kinetin-treated plants. Chapter 5 explores the function and transcriptional regulation of two more candidate Kranz regulators, a closely related pair of *DAG-LIKE DOF* genes, one of which is predicted to be an ANT target gene. Again, phenotypic analysis of rice leaves constitutively expressing the candidate regulators is used to investigate the capacity of these genes to confer Kranz-like traits during C_3 leaf development. Further, *in silico* analysis of known promoter motifs in the two genes' putative promoter regions is used to investigate the potential transcriptional regulatory mechanisms for these genes, and a probable RNAi silencing mechanism is experimentally tested using a transgenic approach. The research implications of all these experiments for engineering C_4 photosynthesis into rice are discussed in chapter 6.

Accession	Gene family	Gene subfamily	Gene name	Ortholog functions
Procambium				
GRMZM2G045883	bHLH	SPCH clade	*ZmSPCH-like 1 (ZmSPCHL1)*	*At4g01460*: Protein-protein interaction with BRAHMA (BRM). Cytokinin responsive chromatin remodelling in leaf development (Efroni *et al.*, 2013). *bHLH71* (*At5g46690*): Protein-protein interaction with FAMA; cell division/differentiation switch during stomata development (Ohashi-Ito & Bergmann, 2006).
GRMZM2G132794	GRAS	SHR clade	*ZmSHR1*	*SHR* (*At4g37650*): Procambium initiation, vascular cell patterning, bundle sheath and endodermis cell identity. Protein-protein interaction with SCR (reviewed in Moreno-Risueno *et al.*, 2015).
GRMZM2G172657	GRAS	SHR clade	*ZmSHR2*	
GRMZM2G028046	C2H2	MRPI clade	*ZmMRPI-like 1 (ZmMIL1)*	*ZmMRPI1* (*GRMZM2G139160*), *ZmMRPI2* (*GRMZM2G105224*): Protein-protein interaction with MYB TF ZmMRP1. Endosperm transfer cell differentiation (Royo *et al.*, 2009).
GRMZM2G136494	C2H2	MRPI clade	*ZmMIL2*	
GRMZM2G150011	C2H2	DOT5 clade	*ZmDOT1*	*AtDOT5* (*At1g13290*): Misaligned veins in mutant (Petricka *et al.*, 2008).
GRMZM2G039074	MYB	KAN clade	*ZmATS-like 1 (ZmATL1)*	*ATS* (*At5g42630*): Polarity specification in ovule development (McAbee *et al.*, 2006). Protein-protein interaction with ARF3: auxin mediated responses (Kelley *et al.*, 2012).
BS/M cell differentiation				
GRMZM2G114998	C2C2-Dof	DAG1 clade	*ZmDAG-like 1 (ZmDLK1)*	*HvSAD* (*MLOC60907*): Protein-protein interaction with HvGAMYB in endosperm (Diaz *et al.*, 2005).
GRMZM2G171852	C2C2-Dof	DAG1 clade	*ZmDLK1H*	*AtDOF4.6* (*At4g24060*): Role in leaf development (Ahmad *et al.*, 2012).
GRMZM2G129261	C2H2	IDD clade	*ZmNKD1 (aka ZmJAY1)*	*ZmNKD1*: Aleurone cell differentiation in endosperm (Yi *et al.*, 2015).
GRMZM2G143723	C2H2	IDD clade	*ZmRAVEN1 (ZmRVN1)*	*AtJKD* (*At5g03150*, *AtBIB* (*At3g45260*): Upregulate *AtSCR*, (reviewed in Moreno-Risueno *et al.*, 2015).
GRMZM2G131516	GRAS	SCR	*ZmSCR1*	*SCR* (*At3g54220*): Asymmetric cell division, vascular cell patterning, bundle sheath and endodermis cell identity, sugar signalling. Protein-protein interaction with SHR (reviewed in Moreno-Risueno *et al.*, 2015).

Table 1.1: Kranz regulator genes proposed to function in procambium development and cell differentiation. Set of genes identified by Fouracre *et al.* (2014) as high priority candidate regulators for procambium development, bundle sheath (BS) or mesophyll (M) cell differentiation during Kranz development.

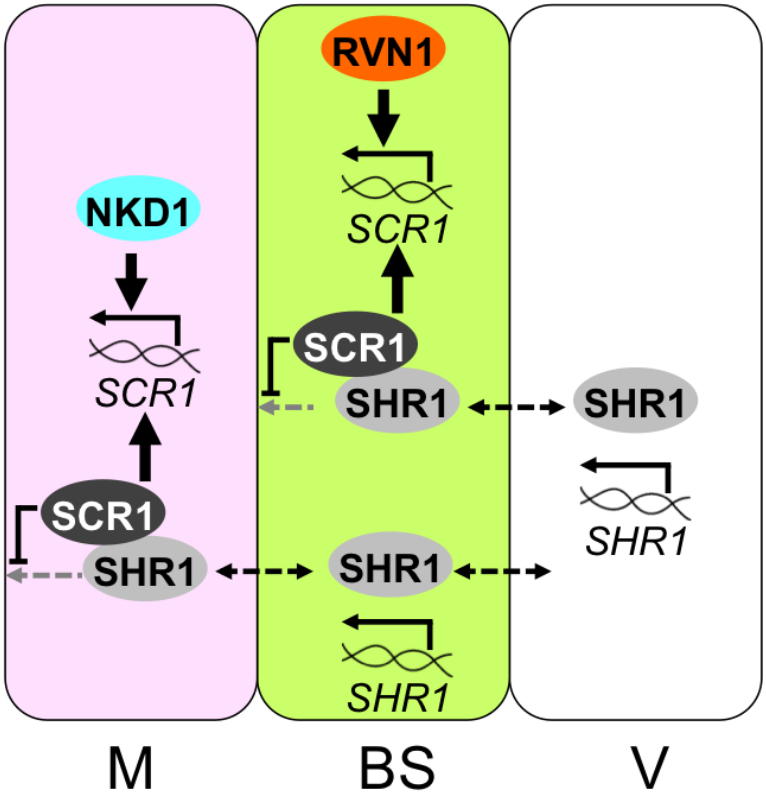


Figure 1.7: *SHR-SCR* model for Kranz cell differentiation. Fouracre *et al.* (2014) proposed a *SHORTROOT-SCARECROW* mediated model for bundle sheath (BS, green) and mesophyll (M, pink) cell differentiation during Kranz development. *ZmSHR1* is expressed in vein tissue (V, white) and BS cells. The ZmSHR1 protein (light grey) moves to the adjacent cell layer where it is sequestered into the nucleus by ZmSCR1 (dark grey), activating transcription of gene targets. The cell-specific factors ZmRVN1 (orange) and ZmNKD1 (blue) promote *ZmSCR1* expression in BS and M cells, respectively. The SHR/SCR complex may act alongside other cell-specific factors to promote the different cell identities. **Ovals**- protein, **helices and hooked arrows**- gene expression, **dashed lines**- **black**: protein movement, **grey**: inhibition of protein movement. Image adapted from Fouracre *et al.* (2014).

2

Materials and Methods

2.1 General materials and methods

All solutions, unless specified, were made using ultrapure water from a Milli-Q system (Merck Millipore). Materials were sterilised, where necessary, by autoclaving (15 min at 120°C, 100 kPa). For RNAse-free work, glassware was heated at 200°C for 12 h to denature RNAse enzymes and plastic was treated with RNAseZap RNase Decontamination Solution (Thermo Scientific). Unless specified, chemicals were purchased from Sigma-Aldrich or Melford.

2.2 Plant material and growth conditions

2.2.1 Rice

Oryza sativa (L.) ssp. Japonica cv. Kitaake (IRRI) was used for all experiments. Plants were grown in a glasshouse at 28°C (day) and 23°C (night), with up to 16 h supplemental light (100 µmol m^{-2} s^{-1}) per day in winter.

Plants used for morphological analysis were grown in a controlled environment growth room (Foster). The conditions were 16 h (100 µmol m^{-2} s^{-1}) light, 28°C;

8 h dark, 23°C. Seed was planted directly into water saturated John Innes Nº2 compost (J. Arthur Bower's), at a density of 180 plants per m^2.

For *in situ* hybridisation seed was planted into water saturated medium grade vermiculite (Sinclair) and the seedlings were grown in the glasshouse as described above.

2.2.2 Maize

Zea mays (L.) ssp. Mays cv. B73 (USDA) was used and plants were grown in the same glasshouse conditions as described in 2.2.1. Seed was sterilised (120 min in 50% ethanol, washed with sterile H_2O) and planted in damp medium grade vermiculite.

2.2.3 Arabidopsis

Arabidopsis thaliana cv. Columbia (Col-0) identified to be heterozygous for a T-DNA insertion in *AtSHR1* (mutant line *shr-6*, Yu *et al.*, 2010) (NASC) was used. Plant were grown in a glasshouse at 21°C with 16 h of supplemental LED light. Seed were sown onto Levington Growth Medium M2 (Everris) and vernalised for 3 days at 4°C prior to transfer to the glasshouse.

2.3 Molecular cloning I: Gateway cloning

2.3.1 DNA template

2.3.1.1 Genomic DNA

Genomic DNA was extracted from young maize or rice leaves. The extraction protocol was based on Murray & Thompson (1980), with significant modifications. 1.5 cm leaf samples were flash frozen in liquid nitrogen (N_2) and powdered

using metal beads and a TissueLyser (Qiagen) (2 min, maximum speed); the samples were kept frozen throughout. 650 µl of extraction buffer (1.5% w/v cetyltrimethylammonium bromide (CTAB), 75 mM Tris-hydrochloride (Tris-HCl) pH 8.0, 15 mM ethylenediaminetetraacetic acid (EDTA) pH 8.0, 1.05 M sodium chloride) was added and samples were incubated for 20 min at 65°C. An equal volume of 24:1 chloroform (VWR):isoamyl alcohol was added and samples were vortexed and centrifuged at 12 000 rpm for 5 min. 2.5 vol of 100% ethanol was used to precipitate the DNA, followed by a further centrifugation. The pellet was washed with 70% ethanol, centrifuged and air dried. The DNA was re-suspended in 100 µl Tris-EDTA (TE) buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA).

2.3.1.2 Complementary DNA

RNA was extracted from shoot primordia enriched maize or rice samples. 7 day old shoot apices were dissected to remove the outer 3 leaves, and flash frozen in liquid N_2. The tissue was homogenised (see 2.3.1.1; p.40) and RNA was extracted using TRIzol reagent (38% water-saturated phenol pH 4.5 (AMRESCO), 0.8 M guanidine thiocyanate, 0.4 M ammonium thiocyanate, 0.1 M sodium acetate pH 5.0, 5% glycerol [Thermo Scientific]) according to the Invitrogen protocol (Life Technologies, 2012). The optional step to remove insoluble material was included and the purified RNA was re-suspended in diethylpyrocarbonate (DEPC) treated H_2O.

The RNA was treated with TURBO DNase (Invitrogen) and quantified with a Nanodrop 1000 spectrophotometer (Thermo Scientific). 500 ng of total RNA was reverse-transcribed into complementary DNA (cDNA) through first-strand cDNA synthesis using Superscript III Reverse Transcriptase (Invitrogen). Oligo$(dT)_{20}$ (Eurogentec) primers were used, and the cDNA synthesis was carried out at 55°C. All steps were conducted according to the manufacturer protocols.

2.3.2 Polymerase chain reaction to amplify template DNA

Polymerase chain reaction (PCR) was used to amplify the genes of interest (see C.1.1; p.264) from gDNA or cDNA. Primers were designed using Primer3Plus software (Untergasser *et al.*, 2012). Primer specificity to the target genomic region was inferred using the BLASTN (Nucleotide Basic Local Alignment Search Tool) algorithm on the genome of interest in Phytozome v.10 (Goodstein *et al.*, 2012). Gateway (Invitrogen) flanking regions attB1 (5′-GGGGACAAGTTTGTAC-AAAAAAGCAGGCTTC-3′) and attB2 (5′-GGGGACCACTTTGTACAAGAAAGCT-GGGTT-3′) were added to the 5′ end of the forward and reverse primer, respectively. Primers were synthesised by Eurogentec (for sequence details see digital appendix C.1.1; p.264).

Phusion High-Fidelity DNA polymerase (New England Biolabs) was used to amplify DNA in a 20 µl reaction (1X Phusion GC buffer, 200 µM dNTPs, 0.5 µM forward primer, 0.5 µM reverse primer, 8% dimethyl sulfoxide (DMSO), 200-500 ng DNA, 1.0 units/50 µl Phusion enzyme) using a thermal cycler (Applied Biosystems) with settings shown in table 2.1 (p.42).

2.3.3 DNA gel electrophoresis and gel extraction

Amplified DNA template (see 2.3.2; p.42) was mixed with DNA loading buffer (Thermo scientific) and loaded onto an agarose gel (40 mM Tris, 20 mM acetic

Step	Temperature	Time	Nº of Cycles
Initial Denaturation	98°C	30 sec	1
Denaturation	98°C	10 sec	35
Annealing	see appx.C	30 sec	
Extension	72°C	see appx.C	
Final Extension	72°C	2 min	1
Final Cooldown	15°C	∞	1

Table 2.1: Phusion enzyme PCR conditions

acid, 1 mM EDTA, 0.7-1.5% w/v agarose [depending on size of PCR product], 0.5 µg µl^{-1} ethidium bromide). The gel was run at 6 V/cm in TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) until there was good band separation. Depending on the size of the expected DNA bands, 1 kbp or 100 bp DNA ladder (Thermo Scientific) was also loaded on to the gel.

The band of correct size was quickly excised whilst gel was exposed to 360 nm UV light. The DNA was extracted from the gel slice using the GenElute Gel Extraction Kit (Sigma-Aldrich) into 50 µl of Tris-HCl pH 9.0. To confirm correct band excision, the remaining gel was subsequently imaged using a Biospectrum Image System with Vision Works software (UVP).

2.3.4 Plasmid construction

The purified DNA (see 2.3.3; p.42) was cloned into the pDONR207 vector (Invitrogen) using the Gateway BP reaction (Invitrogen) according to manufacturer protocol. The resulting plasmid was transformed into *Escherichia coli* and purified (see 2.5.1; p.46). This Gateway entry clone was then used in a LR reaction (Invitrogen) with the destination vector pVec8 (over-expression experiments) (Wang *et al.*, 1997) or pANDA (silencing experiments) (Miki & Shimamoto, 2004), according to manufacturer protocols. These binary vector plasmids were purified and transformed into *E. coli* (see 2.5.1; p.46), then *Agrobacterium tumefaciens* (see 2.5.2; p.47). The resulting binary vector was used for *Agrobacterium* mediated transformation in rice or arabidopsis (see 2.6; p.48).

In silico BP and LR reactions were carried out using ApE Plasmid Editor (Davis, 2013) to produce vector maps. Entry clone and binary vector details can be found in the digital appendix (see C.1.2; p.264).

2.3.5 Sanger sequencing

To confirm the correct genomic sequence has been cloned, sequencing reactions were conducting using the final Gateway binary vectors (see 2.3.4; p.43). The BigDye Terminator v3.1 Cycle sequencing kit (Applied Biosystems) was used, as per manufacturer instructions, but with an addition of 10% v/v DMSO, and with the PCR programme outlined in table 2.2 (below). Sequences for the primers used in these reactions can be found in the digital appendix (see C.1.3; p.264).

The DNA was precipitated on ice with 2.5 vol ethanol and 0.08 M sodium acetate pH 5.2 and centrifuged at 4600 rpm for 40 min. The pellet was washed with 70% ethanol, centrifuged, and then air-dried. The reactions were sequenced using an ABI PRISM 3100 Genetic Analyzer (Applied Biosciences) by a third party, and aligned to the expected nucleotide sequence for the binary vector using ApE Plasmid Editor (Davis, 2013).

2.4 Molecular cloning II: Golden Gate assembly

The Golden Gate assembly methodology is described in Engler *et al.* (2008). Simply, the system is based on modules, with level 0 plasmids containing sequence for one part of a transcriptional unit, such as a promoter region, gene coding sequence etc. The units are ligated with a backbone vector to produce a level 1 plasmid that contains a transcriptional unit (e.g. promoter-gene-terminator).

Step	Temperature	Time	Nº of Cycles
Initial Denaturation	98°C	1 min	1
Denaturation	98°C	10 sec	40
Annealing	55°C	5 sec	
Extension	60°C	4 min	
Final Cooldown	15°C	∞	1

Table 2.2: BigDye Terminator v3.1 sequencing reaction PCR conditions

The transcriptional units can be stacked to produce a level 2 binary vector with several transcriptional units for transformation into the organism(s) of interest. All plasmids used in this study are listed in the digital appendix (see C.1.4; p.264).

2.4.1 DNA template and level 0 plasmid synthesis

Maize coding sequence for the gene of interest was extracted from the Phytozome v.11 database (Goodstein *et al.*, 2012). If BpiI, BsaI, DraIII or Esp3I restriction endonuclease target sites were present, the sites were removed through 'domestication': the nucleotide sequence was altered, whilst retaining the endogenous amino acid sequence. These sequences were also codon optimised for rice using the Codon Optimisation Tool (Integrated DNA Technologies) (www.idtdna.com/CodonOpt). The nucleotide sequence was synthesised and cloned into a level 0 plasmid by GenScript.

2.4.2 Level 1 plasmid ligation

100 ng of the required level 1 backbone and the level 0 plasmids (200 ng each) were digested and ligated in a 20 µl reaction using 1X Cutsmart Buffer (New England BioLabs), 10 mM ATP (Thermo Scientific), 10 units BsaI (New England BioLabs) and 5-10 units T4 DNA ligase (Promega). The thermal cycle outlined in table 2.3 (p.46) was used. The ligated level 1 plasmids were transformed into *E. coli* and purified (see 2.5.1; p.46). Restriction endonuclease enzymes (New England Bioscience) were used to confirm correct ligation, according to manufacturer protocols.

2.4.3 Level 2 plasmid ligation

100 ng of the required level 2 backbone and the level 1 plasmids (200 ng each) were digested and ligated in a 20 µl reaction using 1X Buffer G (Thermo Scientific),

10 mM ATP, 10 units BpiI (Thermo Scientific) and 5-10 units T4 DNA ligase. The thermal cycle in table 2.3 (below) was used and the ligated level 2 plasmid (the binary vector) was transformed into *E. coli* and purified (see 2.5.1; below), prior to *Agrobacterium* transformation (see 2.5.2; p.47). Restriction endonuclease enzymes (New England Bioscience) were used to confirm correct ligation, according to manufacturer protocols. The level 2 plasmid sequence was confirmed using Illumina DNA sequencing by IMGM Laboratories.

2.5 Bacterial transformation

2.5.1 *Escherichia coli*

Chemically competent *E. coli* cells were transformed with the plasmid of interest through heat shock treatment (see Sambrook, 2001). 2.5 µl DNA per 50 µl DH5α competent cells (made using the Inoue Method for Ultra-Competent cells described in Inoue *et al.*, 1990) were used for Gateway plasmids, and 1 µl DNA per 4 µl MAX Efficiency DH10B cells (Thermo Scientific) was used for Golden Gate plasmids.

Transformed cells were selected on Luria-Bertani broth (LB) agar plates grown overnight at 37°C with appropriate antibiotic (see C.1.2; p.264) and, for DH10B cells, with 40 µg ml^{-1} 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for blue/white selection.

Step	Temperature	Time	Nº of Cycles
Digestion	37°C	10 min	3
Ligation	16°C	10 min	
Final Digestion	37°C	10 min	1
Enzyme Inactivation	65°C	20 min	1
Final Cooldown	15°C	∞	1

Table 2.3: Golden Gate digestion/ligation thermal cycle

Single colonies were individually transferred into 5 ml LB broth with antibiotic and grown overnight at 37°C with shaking at 200 rpm. A glycerol stock was made by adding 300 µl culture to 700 µl 50% glycerol, this mixture was flash frozen using liquid N_2, and stored at -80°C. The remaining culture was taken for plasmid purification using the QIAprep Spin Miniprep Kit (Qiagen) according to manufacturer protocols. The presence of transformed cells was confirmed using colony PCR (see 2.5.3; below).

2.5.2 *Agrobacterium tumefaciens*

Agrobacterium strain EHA105 electrocompetent cells (made by Z. Lewis using protocol in Sambrook, 2001) were transformed with the binary vector through electroporation. 1 µl of binary vector per 40 µl ice-cold competent cells was used, and the cells were shocked using a Gene Pulser II (Bio-rad) (25 µF capacitance, 2.5 kV cm^{-1} field strength, 600 Ω resistance). After electroporation, 1 ml ice-cold LB was added, and cells were left at 28°C for 3 h with 150 rpm shaking. The transformed cells were selected on LB agar plates grown at 28°C for 2 days with antibiotic selection appropriate to the plasmid (see C.1.2; p.264) and 50 µg ml^{-1} rifampicin. Single colonies were individually transferred into 5 ml LB broth with appropriate antibiotic and grown for 2 days at 28°C with 150 rpm shaking. A glycerol stock was made by adding 300 µl culture to 700 µl 50% glycerol, this mixture was flash frozen using liquid N_2, and stored at -80°C.

2.5.3 Colony PCR

Individual colonies of *E. coli* and *A. tumefaciens* surviving on selection medium were tested for transformation using colony PCR to amplify plasmid-specific DNA regions (see C.1.5; 264). GoTaq G2 DNA polymerase (Promega) was used in a 10 µl reaction (1X GoTaq buffer, 200 µM dNTPs, 0.5 µM forward primer, 0.5 µM

reverse primer, bacterial DNA, 1.25 units/50 µl GoTaq enzyme), using the thermal cycle settings in table 2.4 (below). Gel electrophoresis was used to visualise DNA bands (see 2.3.3; p.42).

2.6 Plant tranformation

2.6.1 Rice callus transformation

2.6.1.1 Tissue culture media

R1 medium:

4.3 g L^{-1} Murashige and Skoog (MS) salts and vitamins (Duchefa Biochemie), 30 g L^{-1} sucrose, 0.5 g L^{-1} 2-morpholinoethanesulfonic acid (MES) , 300 mg L^{-1} casamino acid (N-Z-Amine), 2.8 g L^{-1} L-proline, 4 g L^{-1} Phytagel, pH 5.8.

AAM medium:

4.3 g L^{-1} MS salts and vitamins, 68.5 g L^{-1} sucrose, 0.5 g L^{-1} MES, 36 g L^{-1} glucose, 500 mg L^{-1} casamino acid, 0.876 g L^{-1} L-Glutamine, 0.266 g L^{-1} L-aspartic acid, 0.174 g L^{-1} L-Arginine, 7.5 mg L^{-1} glycine, pH 5.2. 40 mg L^{-1} sterile 3,5-dimethoxy-4- hydroxyacetophenone (acetosyringone) added before use.

RS2 medium:

4.3 g L^{-1} MS salts and vitamins, 30 g L^{-1} sucrose, 0.5 g L^{-1} MES, 300 mg L^{-1} casamino acid, 10 g L^{-1} glucose, 2 mg L^{-1} 2,4-dichlorophenoxyacetic acid (2,4-D), 4 g L^{-1} Phytagel, pH 5.2. 20 mg L^{-1} sterile acetosyringone added after autoclaving.

Step	Temperature	Time	Nº of Cycles
Initial Denaturation	96°C	5 min	1
Denaturation	96°C	30 sec	35
Annealing	see appx.C	30 sec	
Extension	72°C	see appx.C	
Final Extension	72°C	5 min	1
Final Cooldown	15°C	∞	1

Table 2.4: GoTaq enzyme PCR conditions

R3 medium:

4.3 g L^{-1} MS salts and vitamins, 30 g L^{-1} sucrose, 0.5 g L^{-1} MES, 300 mg L^{-1} casamino acid, 2.8 g L^{-1} L-proline, 2 mg L^{-1} 2,4-D, 4 g L^{-1} Phytagel, pH 5.8. 200 mg L^{-1} ticarcillin disodium and clavulanate potassium (timentin) (Thomas Scientific) and 30 mg L^{-1} hygromcycin B added after autoclaving.

R4 medium:

4.3 g L^{-1} MS salts and vitamins, 30 g L^{-1} sucrose, 0.5 g L^{-1} MES, 2 g L^{-1} casamino acid, 30 g L^{-1} sorbitol, 2 mg L^{-1} kinetin, 1 mg L^{-1} 1-naphthaleneacetic acid (NAA), 4 g L^{-1} Phytagel, pH 5.8. 200 mg L^{-1} timentin and 20 mg L^{-1} hygromcycin B added after autoclaving.

2.6.1.2 Rice callus induction

Rice seed was de-hulled and sterilised (2 min in 70% ethanol, washed with sterile H_2O, 15 min in 2.5% sodium hypochlorite, washed thoroughly with sterile H_2O). The seed was placed embryo down onto R1 media and cultured in constant light (125 µmol m^{-2} s^{-1}) at 32°C for 14 days. Proliferating callus was transferred to a new R1 plate and cultured for a further 7 days.

2.6.1.3 *Agrobacterium* mediated callus transformation

Agrobacterium cells transformed with the binary vector (see 2.5.2; p.47) were grown from glycerol stock in 10 ml LB medium with appropriate antibiotics (see C.1.2; p.264) for 24 h at 28°C with 150 rpm shaking. 5 ml of culture was centrifuged at 3500 *g* for 5 min at room temperature and the pellet was resuspended in 4 ml AAM medium. After *Agrobacterium* cells had equilibrated for 1 h, rice calli were immersed in the suspension for 5 min then left to air-dry on filter paper for 20 min. The calli were transferred to filter paper placed on R2 medium and co-cultured for 3 days at 25°C in the dark.

2.6.1.4 Transformed rice callus selection

After co-culture, rice calli were transferred to R3 medium to select transformed cells. The calli were cultured at 32°C with constant light (125 µmol m^{-2} s^{-1}) for 14 days. Proliferating tissue was transferred to new R3 medium and cultured for a further 14 days.

2.6.1.5 Transformed rice callus regeneration

After a total of 28 days on R3 medium, surviving calli were moved to R4 medium, and cultured at 32°C with constant light (125 µmol m^{-2} s^{-1}). Proliferating callus tissue was transferred to new R4 medium every 14 days until shoots and roots had regenerated. The plants were transferred to a growth chamber with a 16 h 100 µmol m^{-2} s^{-1} light / 8 h dark photoperiod at 30°C and grown hydroponically, supplemented with 1 g L^{-1} MS salts and vitamins every 7 days, until plant height was 20 cm. The plants were transplanted into water saturated John Innes Nº2 compost, at a density of 180 plants per m^2 and grown under glasshouse conditions (see 2.2.1; p.39).

2.6.2 Arabidopsis transformation using floral dipping

2.6.2.1 Arabidopsis plants for floral dipping

Arabidopsis plants were grown in glasshouse conditions (see 2.2.3; p.40) at a density of 360 plants per m^2. After bolting, inflorescence shoots were pruned to encourage new shoots and plants were used for transformation once new flowers were visible.

2.6.2.2 *Agrobacterium* mediated arabidopsis transformation

Agrobacterium cells transformed with the binary vector (see 2.5.2; p.47) were grown from glycerol stock in 10 ml LB medium with appropriate antibiotics (see C.1.2; p.264) for 48 h at 28°C with 180 rpm shaking. 5 ml of this pre-culture were used to inoculate 500 ml LB broth. The culture was incubated for 18 h at 28°C with 150 rpm shaking, then centrifuged at 2700 *g* for 20 min at 4°C. The cells were resuspended in 500 ml of dipping solution (2.15 g L^{-1} MS, 50 g L^{-1} sucrose, 500 µl L^{-1} Tween 20). Arabidopsis plants were dipped into the solution for $\sim$1 min, placed into a plastic bag and left at 4°C overnight. The plants were then transferred to glasshouse conditions to set seed (see 2.2.3; p.40).

2.6.3 Identification of transformed plants

2.6.3.1 Hygromycin resistance assay to select arabidopsis transformants

To identify transformed arabidopsis plants, a hygromycin growth assay was used. T_1 arabidopsis seed were first sterilised (10 min in 70% ethanol plus 0.1% tween 20, brief wash in 100% ethanol and air-dried). To select transformed plants, the seed was then placed on MS medium with hygromycin (2.15 g L^{-1} MS salts with vitamins, 0.5 g L^{-1} MES, 4 g L^{-1} Phytagel, pH 5.8. 20 mg L^{-1} hygromcycin B added after autoclaving), vernalised at 4°C and grown in a controlled environment room (Foster) with constant 150 µmol m^{-2} s^{-1} light at 21°C. Surviving individuals were transplanted into soil and grown in the glasshouse (see 2.2.3; p.40).

2.6.3.2 SDS protocol for DNA extraction

Leaf tissue from arabidopsis and rice plants that had undergone hygromycin selection was harvested and DNA was extracted in a high throughput approach using sodium dodecyl sulfate (SDS). Tissue samples were placed in 96-well

collection microtubes with metal beads. 375 µl extraction buffer (200 mM Tris pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) was added at room temperature and samples were homogenised using a TissueLyser (2 min, maximum speed). The samples were centrifuged for 10 min at 6000 rpm and 300 µl isopropanol was added to 250 µl supernatant. After mixing, samples were incubated at room temperature for 10 min to precipitate DNA and then centrifuged for 10 min at 6000 rpm. The resulting DNA pellet was washed with 70% ethanol, centrifuged, then left to air dry. Once ethanol had evaporated, the pellet was resuspended in 100 µl TE buffer.

2.6.3.3 PCR to identify trangenic plants

To confirm plant transformation, 1 µl of the extracted genomic DNA was used in a PCR reaction (see 2.5.3; p.47). The primers used amplified a region which included the cloned gene and backbone sequence (see C.1.5; p.264 for primer details). Gel electrophoresis was used to visualise DNA bands, as described in 2.3.3 (p.42).

2.6.4 Transgene expression analysis

2.6.4.1 RNA extraction and cDNA synthesis

Young leaf tissue from transgenic arabidopsis and rice plants was harvested and flash frozen in liquid N_2. Tissue was homogenised, RNA extracted and cDNA synthesised as in 2.3.1.2 (p.41).

2.6.4.2 Reverse-transcription PCR

Diluted cDNA (0.1 x dilution, 1 µl) was used in PCR reactions (see 2.5.3; p.47) with primer pairs designed to amplify the cloned gene and endogenous *UBIQUITIN* mRNA (see C.1.5; p.264 for primer details).

2.7 Morphological analysis

2.7.1 Rice leaf sectioning

The 'flag leaf minus 1' was harvested from 3 month old rice plants, and two 2 cm segments spanning the whole leaf width were cut from the mid-point of the leaf blade. The tissue was mounted in 7% agar and fresh leaf sections (40-80 µm thick) were cut using a Vibratome 1000 (The Vibratome Company). Sections were cleared overnight in 3:1 ethanol:acetic acid and mounted on glass slides (Thermo Scientific) in 75% ethanol. The sections were imaged under UV light using a Leica DMR8 microcope, QImaging MicroPublisher (QImaging) camera and Image-Pro Insight (MediaCybernetics) software.

2.7.2 Arabidopsis leaf clearing

Leaf 5 from 21 day old arabidopsis plants was cleared with chloral hydrate solution (100% w/v chloral hydrate, 1% v/v glycerol) for 48 h, mounted on glass slides using 50% glycerol and sealed with nail varnish. The leaves were imaged using darkfield photography with a Nikon D300 camera, and vein patterning analysed using LIMANI (Leaf Image Analysis Interface) software (Dhondt *et al.*, 2012). Leaves were also imaged using differential interference contrast (DIC) microscopy with an Olympus BX50 microscope.

2.7.3 Measurements and statistical analysis

Anatomical features were measured using ImageJ software (Schneider *et al.*, 2012). Statistical analysis was conducted using R software (R Core Team, 2013). All non-counts datasets were tested for significant deviance from normality and equal variances between groups (Shapiro & Wilk, 1965; Levene, 1960). If data

violated the parametric assumptions, appropriate functions were used for data transformation. Linear mixed effects models were fitted to test for significant differences between groups. Where data were nested, appropriate ANOVA tests were used to analyse differences. For counts data, a poisson distribution was assumed and general linear mixed effects models were fitted (see Whitlock & Schluter, 2009). Where appropriate, *post hoc* Tukey's Honesty Significant Difference tests were used for pair-wise analysis between groups (Tukey, 1949).

2.8 *In situ* hybridisation

2.8.1 RNA probes

2.8.1.1 RNA probe design

For each gene of interest, 80 bp - 1 kbp long probes were designed based on the transcript sequence. BLASTN software was used to identify genomic regions that had $<$80% homology to other predicted expressed sequences. Primers were designed using Primer3Plus (Untergasser *et al.*, 2012) and flanking sequences containing RNA polymerase binding sequences were added (the T3 RNA polymerase sequence 5′-AATTAACCCTCACTAAAGGG-3′ was added to the 5′- end of the forward primer, and the T7 RNA polymerase sequence 5′-TAATACGACTCACTATAGGG-3′ to the 5′ end of the reverse primer). Probe details are in the digital appendix (see C.2.1; p.264).

2.8.1.2 Digoxigenin-labelled RNA probe synthesis

Primordia enriched cDNA (see 2.3.1.2; p.41) was used to amplify template DNA using a 10 µl PCR reaction (see 2.3.2; p.42). To increase yields, 2 µl of the PCR product was then used in a 100 µl PCR reaction. The DNA template

was purified using gel extraction (see 2.3.3; p.42) and RNAse contamination was removed through phenol/chloroform/isoamylalcohol extraction (see Sambrook, 2001). The DNA was air dried and re-suspended in 10 µl DEPC H_2O. DNA concentration was measured using a Nanodrop 1000 and adjusted to give $\sim 2.7 \times 10^{11}$ DNA molecules/µl.

Digoxigenin (DIG)-labelled RNA probes was synthesised as described in Langdale (1994) but with the following reaction mix: 1X RNA ploymerase buffer, 0.5 mM ATP, 0.5 mM GTP, 0.5 mM CTP, 0.1 mM DIG-UTP, 0.8 U μl^{-1} RNA polymerase (all Thermo Scientific), 1.6 U μl^{-1} RNAseOUT (Invitrogen). The reaction was incubated for 1.5 h at 37°C and terminated by 10 min incubation at 37°C with 3 vol DNAse solution (10 mM Tris-HCl pH 7.5, 10 mM magnesium chloride, 50 mM sodium chloride, 2.5 µg μl^{-1} tRNA, 0.02 U μl^{-1} DNAse I). RNA was resuspended in 50 µl DEPC H_2O. A 1 µl aliquot was taken to analyse DIG-incorporation efficiency (see 2.8.1.3; p.55) and the rest was hydrolysed to give $\sim$200 bp fragments. Hydrolysed RNA was precipitated as per Langdale (1994), re-suspended in 49 µl DEPC H_2O and stored at -80°C in 50% formamide.

2.8.1.3 RNA probe DIG-incorporation and hydrolysis efficiency

Using RNAse-free protocols, hydrolysed and unhydrolysed RNA samples were mixed with 1 vol RNA loading buffer (Thermo), denatured at 70°C for 10 min and loaded onto a denaturing RNA formaldehyde gel (1.5% w/v agarose, 2.2 M formaldehyde, 20 mM 3-morpholinopropane-1-sulfonic acid [MOPS] pH 8, 5 mM sodium acetate, 1 mM EDTA). The gel was run for 1.5 h at 6 V/cm in MOPS running buffer (20 mM MOPS pH 7.5, 5 mM sodium acetate, 1 mM EDTA).

After electrophoresis, the gel was soaked in 20X SSC (3 M sodium chloride, 0.3 M sodium citrate, pH 7.6) for 15 min, and RNA was transferred to a nylon membrane (Scientific Laboratory Supplies) using the northern blotting protocol

in Sambrook (2001). The RNA was cross-linked to the membrane using a UV light Stratlinker 1800 (Stratgene) set at 254nm.

DIG-label was visualised by antibody staining; the membrane was washed for 5 min in Tris-NaCl (100 mM Tris, 150 mM sodium chloride, pH 7.5), then treated with blocking solution (Tris-NaCl, 0.5% w/v blocking reagent [Roche]) for 30 min, subsequently the Tris-NaCl step was repeated three times. Alkaline-phosphatase conjugated anti-DIG antibody (Anti-DIG-AP, Fab fragments [Roche]) was bound to DIG-labelled RNA by incubating the membrane in antibody solution (Tris-NaCl, 0.15 U ml^{-1} Anti-DIG-AP). Subsequently, the membrane was washed twice with Tris-NaCl for 15 min per wash, then briefly washed in Tris-MgCl solution (100 mM Tris, 100 mM sodium chloride, 50 mM magnesium chloride, pH 9.5). To stain the antibody, the membrane was incubated in NBT/BCIP solution (Tris-MgCl, 100 µg ml^{-1} 4-nitro-blue-tetrazolium chloride (NBT), 100 µg ml^{-1} 5-bromo-4-chloro-indolylphosphate (BCIP)) until signal was visible, then rinsed with H_2O to terminate the reaction.

2.8.2 Slide preparation and pre-treatment

2.8.2.1 Tissue fixation

Maize and rice plants (3 day old and 17 day old, respectively, grown under conditions described in 2.2.1; p.39) were harvested on ice by cutting 1 cm segments of shoot in the region between the top of the mesocotyl and the base of the coleoptyle, to give samples that included the youngest leaf primordia and the shoot apical meristem (SAM). The primordia samples were vacuum infiltration with ice-cold 4% paraformaldehyde solution, and fixed at 4°C overnight (Jackson, 1992). The tissue was taken through a graded ethanol series (2 h in 10%, 30%, 50%, 70% ethanol) and stored at 4°C in 70% ethanol.

2.8.2.2 Sectioning

Fixed tissue was infiltrated with Paraplast Plus (paraffin wax, 0.8% DMSO) using a Tissue-Tek VIP 5 vacuum infiltration processor (VWR) (see digital appendix for cycle details: C.2.2; p.264). The samples were then embedded in Paraplast Plus and mounted for sectioning. A N°820 rotary microtome (Spencer Lens Co.) was used to cut 8 µm sections that were then mounted on Superfrost Plus glass slides (Thermo Scientific). The slides were dried at 37°C for 16 h and stored at 4°C with desiccant for a maximum of 5 days prior to use.

2.8.2.3 Tissue pre-treatment

Prior to hybridisation, sections were pre-treated using a protocol modified from Javelle *et al.* (2011). The 30 s 10% ethanol and 1 min H_2O washes were replaced with a 2 min 0.85% w/v sodium chloride wash. Further, the proteinase K incubation time was increased to 30 min and the required concentration was determined empirically for each enzyme batch, ranging between 0.8-1 µg ml^{-1}. After the final 100% ethanol dehydration step, slides were air-dried.

2.8.3 Hybridisation and post-hybridisation treatments

2.8.3.1 Hybridisation

Hybridisation of RNA probes to mRNA was carried out according to Langdale (1994), except the hybridisation buffer composition was modified by replacing DTT and poly A with 1.25X Denhardt's solution (Thermo Scientific). Slides were incubated at 55°C, and kept in a horizontal position without sealant.

2.8.3.2 Post-hybridisation washes

After hybridisation, slides were washed according to Javelle *et al.* (2011), except the saline sodium citrate (SSC) buffer concentration was adjusted depending on the calculated melting temperature (T_m) of the RNA:RNA complexes; the concentration ranged between 0.005X-0.1X SSC (see C.2.1; p.264 for probe-specific details).

2.8.3.3 Antibody staining and fluorescence counter-staining

DIG-labelled RNA was visualised following the protocol in Javelle *et al.* (2011), except slides were kept horizontal and not sandwiched at the NBT/BCIP incubation step. Once signal was visible, the slides were taken through a modified ethanol series (H_2O, 70%, 95%, 100%, 95%, 70% ethanol, H_2O), counterstained for 5 min with 0.1% w/v calcofluor white (Fluorescent Brightener 28) and briefly washed with H_2O. The slides were mounted in Entellan (Merck Millipore) and imaged using brightfield and UV microscopy as in 2.7.1 (p.53).

2.8.3.4 Image analysis

UV images were used to categorise the stages of vein development visible in sections after *in situ* hybridisation (fig. 2.1; p.60). Cell outlines were traced from UV images using Sketchbook (Autodesk) software (www.sketchbook.com), and overlaid on light microscopy images using GNU Image Manipulation Program (GIMP) software (www.gimp.org).

2.9 *In silico* analyses

2.9.1 Phylogenetic inference

BLASTP (protein) and BLASTN (nucleotide) search algorithms were used on proteomes and genomes from fourteen angiosperm and one bryophyte species stored on the Phytozome v.10 database (Goodstein *et al.*, 2012). Protein and genomic sequences that were significantly similar (threshold: E-value $<1 \times 10^{-3}$) to the amino acid and transcript sequence from genes of interest were identified. For each gene of interest, the amino acid sequences for significantly similar genes were aligned using MergeAlign software (Collingridge & Kelly, 2012). A maximum-likelihood (ML) method was used to infer the phylogenetic tree using RAxML software (Stamatakis, 2014); 100 rapid bootstrap analysis runs were used to produce a best scoring ML tree (the protein substitution model used was selected automatically by the software). Phylogenetic trees were visualised using Interactive Tree of Life software (Letunic & Bork, 2016).

2.9.2 RNAseq analysis

RNA sequencing data from Wang *et al.* (2013b) and van Campen *et al.* (2016) was mapped against reference genomes using TopHat2 software (Kim *et al.*, 2013) and RNA abundance was quantified in transcripts per million (tpm) by Dr S. Kelly. The aligned RNA and DNA sequences were visualised *in silico* using Integrative Genomics Viewer (IGV) software (Robinson *et al.*, 2011).

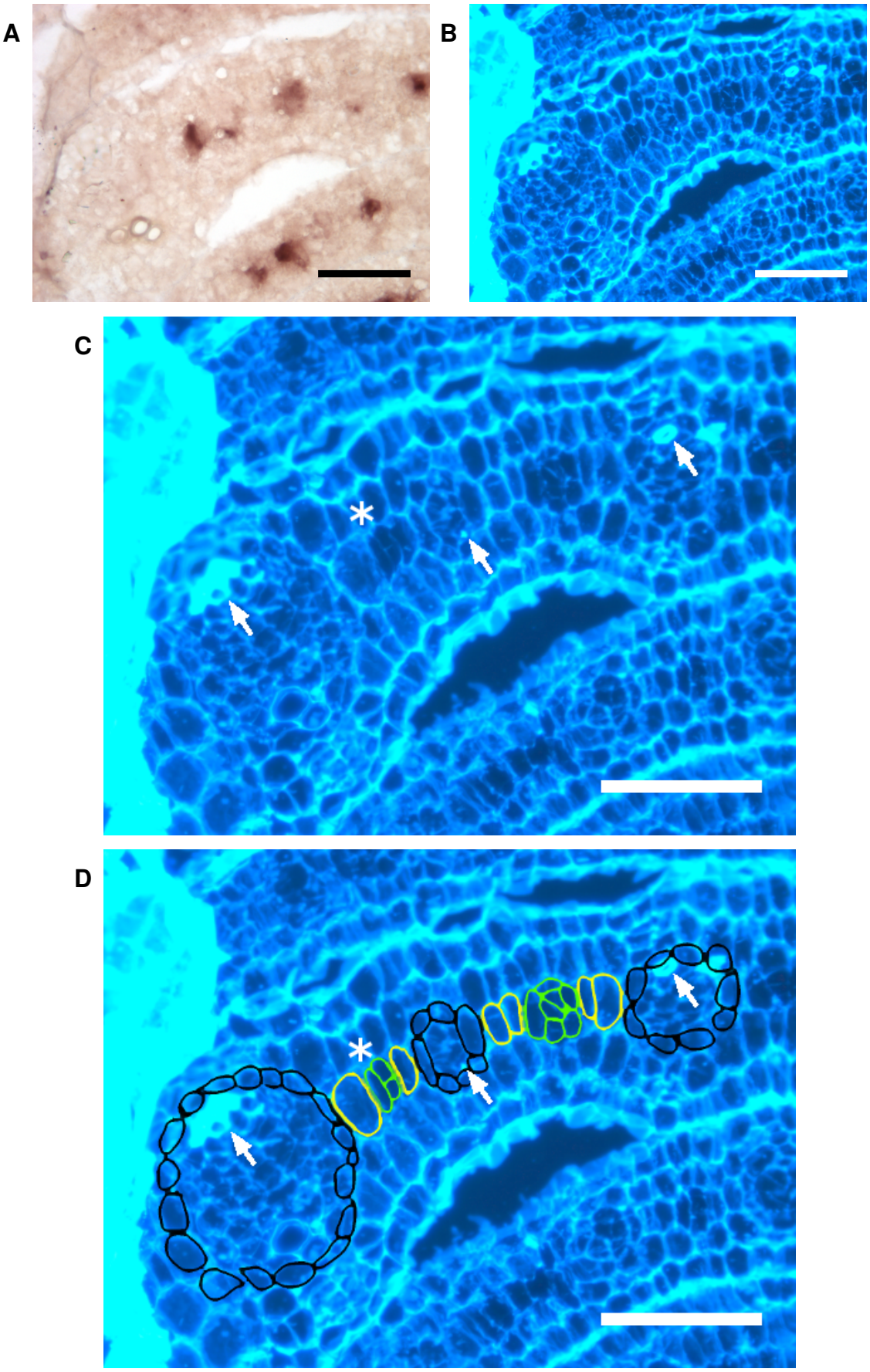


Figure 2.1: Calcofluor counter-staining to catagorise key stages of vein development. Figure legend on p.61.

Figure 2.1: Calcofluor counter-staining to catagorise key stages of vein development. A-B) Transverse section of 4 day old maize seedling, *in situ* hybridised with an anti-sense *ZmDOT1* RNA probe, counter-stained with calcofluor and imaged using light (A) or UV (B) microscopy. **C-D)** Larger scale UV image of the transverse section (C), annotated with key vein developmental stages (D). Ground tissue cells in the inner-most cell layer (yellow) have the competence to develop into procambium tissue (green), the earliest recognisable stage of procambium is highlighted (asterisk). Differentiated bundle sheath cells (black) are annotated once there is evidence of lignification in the vascular bundle (arrows). Scale bar- 50 µm.

2.9.3 *In silico* promoter region analysis

2.9.3.1 Regulatory element identification

NSITE-PL software To identify putative regulatory elements (REs) in promoter regions, 3 kbp of genomic sequence upstream of the genes of interest were searched for statistically significant functional motifs. The REs were from the RegSite Database of Plant Regulatory Elements (Softberry Inc.) and the *in silico* analysis was conducted using NSITE-PL software (Shahmuradov & Solovyev, 2015).

PatMatch software The maize genome sequence and gene annotation files were downloaded from the Phytozome v.10 database and 3 kbp of genomic sequence upstream of each open reading frame (ORF) were extracted by Dr D. Emms. Putative AINTEGUMENTA (ANT) target genes were identified by searching the data for the ANT protein DNA binding consensus sequence (5′-gCAC(A/G)N(A/T)TcCC(a/g)ANG(c/t)-3′), allowing 2 mis-matches, using PatMatch software (Yan *et al.*, 2005). The arabidopsis genome was also analysed; 1 kbp of genomic sequence upstream of ORFs were searched for the ANT binding consensus region using PatMatch hosted on the Arabidopsis Information Resource (TAIR) website (TAIR, 2016).

2.9.4 Gene Ontology term analysis

The phylogenetic relationships between arabidopsis genes and putative ANT target maize genes (see 2.9.3.1; p.62) were inferred by Dr D. Emms using Orthofinder software (Emms & Kelly, 2015). All arabidopsis genes that shared an orthogroup with a putative ANT target maize gene were used for gene ontology (GO) term analysis. GO terms for genes of interest were extracted from the Protein Analysis Through Evolutionary Relationships (PANTHER) and Gene Ontology (GO) databases (Mi *et al.*, 2005; Ashburner *et al.*, 2000) using PANTHER Gene Analysis Tools software (Mi *et al.*, 2013).

3

Using *in situ* hybridisation to refine models of Kranz development

3.1 Introduction

3.1.1 Maize transcriptome analysis identified candidate regulators of Kranz anatomy

Transcriptome analysis compared gene expression in maize foliar and husk leaves at different developmental stages, to identify genes that were upregulated during Kranz development. A subset of these genes were classified as candidate Kranz regulators (Wang *et al.*, 2013b; Fouracre *et al.*, 2014). In this subset, a group of eleven transcription factors were taken forward for further analysis based on RNA expression profiles and relationships to published developmental regulators. This group includes orthologs of genes that are involved in the SHORTROOT-SCARECROW root tissue patterning pathway (*ZmSHR1*, *ZmSHR2*, the homeologs *ZmSCR1* and *ZmSCR1H*, and the *INDETERMINATE DOMAIN* genes *ZmRVN1* and *ZmNKD1*) and also orthologs to arabidopsis genes known to be involved in leaf development (*ZmDOT1* and *ZmANT1*) (table 3.1, p.64).

Gene ID	TF Family	Gene clade	Gene name	FP 1	FP 3&4	FP 5	FI	FE	HP 1	HP 3&4	HP 5	HI	HE	BS	M
GRMZM2G132794	GRAS	SHR	*ZmSHR1*	91	255	298	25	8	40	66	98	1	1	2	0
GRMZM2G172657	GRAS	SHR	*ZmSHR2*	377	874	787	10	1	206	301	353	2	0	0	0
GRMZM2G131516	GRAS	SCR	*ZmSCR1*	1774	2678	2112	111	28	702	1484	1685	33	46	3	1
GRMZM2G015080	GRAS	SCR	*ZmSCR1H*	333	595	419	313	132	295	360	259	29	54	11	13
GRMZM2G028046	C2H2	MRPI	*ZmMIL1*	58	289	308	24	0	10	95	137	1	0	0	0
GRMZM2G136494	C2H2	MRPI	*ZmMIL2*	69	178	173	21	0	16	43	47	3	1	0	0
GRMZM2G150011	C2H2	DOT5	*ZmDOT1*	148	551	348	4	0	28	54	91	0	0	0	0
GRMZM2G129261	C2H2	IDD	*ZmNKD1*	716	1438	1756	919	642	591	620	648	236	299	3	14
GRMZM2G143723	C2H2	IDD	*ZmRVN1*	187	479	387	57	145	37	93	123	95	104	8	0
GRMZM2G151542	AP2	CRF	*ZmCRF1*	27	99	331	6	0	32	48	22	0	0	0	1
GRMZM2G021573	AP2	ANT	*ZmANT1*	1351	1873	922	14	2	425	671	452	6	1	0	0
GRMZM2G409726	*N/A*	UBI	*ZmUBI1*	11383	9948	8703	7403	13509	13402	18522	15457	18290	17934	46180	31727

Table 3.1: Putative Kranz regulators analysed by *in situ* hybridisation. Transcription factor encoding genes identified as putative Kranz regulators in Wang *et al.* (2013b) and Fouracre *et al.* (2014), annotated with gene family and clade. A ubiquitously expressed gene (*ZmUBI1*) is shown as a reference for high expression levels. mRNA levels measured in relative transcript abundance in foliar (F) and husk (H) leaf primordia; plastochron (P)- 1, 3&4 and 5; immature (I) and expanded (E) leaves (Wang *et al.*, 2013b); bundle sheath (BS) and mesophyll (M) cells (Li *et al.*, 2010). Note: P1 samples included the shoot apical meristem.

3.1.2 Spatio-temporal control of transcription factor activity is essential for tissue patterning

Specific spatial and temporal regulation of transcription factor activity is essential for correct tissue patterning during plant development. For example, spatial restriction of *ATHB8* expression during leaf development leads to procambium initiation and xylem differentiation, whereas ectopic expression leads to misplaced xylem development in the arabidopsis leaf (Baima *et al.*, 1995, 2001). Similarly, the spatio-temporal regulation of *SCR* expression is crucial for correct tissue patterning in the root (Di Laurenzio *et al.*, 1996). Although the foliar and husk tissue transcriptome analysis (Wang *et al.*, 2013b) is one of the most detailed published reports of RNA changes during the early stages of maize leaf development, the resolution is not sufficient to identify the spatial-temporal control of candidate Kranz regulators at a cellular level.

3.1.3 *In situ* hybridisation can resolve the spatio-temporal expression patterns of developmental regulators

In situ hybridisation (ISH) is a powerful technique to visualise gene-specific mRNA accumulation in fixed tissues (Cox *et al.*, 1984). The technique has already been successfully used to study monocot leaf development. For example, the light-dependent switch from C_3 to C_4 photosynthesis in maize foliar tissues was identified by studying C_4 photosynthetic gene expression patterns (Langdale *et al.*, 1988b). ISH has also been used to characterise regulators of rice leaf development, for example, showing that the unknown gene *ABNORMAL VASCULAR BUNDLES* (*AVB*) was implicated in procambium initiation (Ma *et al.*, 2016). Therefore, resolution of transcript accumulation on a cellular level using ISH can give valuable information for characterisation of developmental regulators.

This chapter aimed to resolve the spatio-temporal expression patterns of eleven candidate Kranz regulators in maize, and to test whether the expression patterns of orthologous genes are conserved in a C_3 developmental context. ISH was performed on sectioned maize and rice shoot apices, that covered a range of developmental stages from shoot apical meristem (SAM) to plastochron 5 (primordia with mature veins). The data obtained enabled the model of Kranz development (Fouracre *et al.*, 2014) to be refined and informed the prioritisation of genes for future functional analysis.

3.2 Results

3.2.1 Transcript profiles of candidate transcription factors distinguish three distinct stages of Kranz development

Transverse cross sections of 4 day old maize seedlings were used for ISH to visualise the mRNA accumulation patterns of candidate regulators (table 3.1, p.64) at different stages of Kranz development. The mRNA localisation patterns could be partitioned into three groups, dependent on the timing of expression (fig. 3.14; p.100). *ZmMIL2* mRNA was present from the earliest developmental stages, when ground tissue was proliferating. *ZmSHR2*, *ZmANT1*, *ZmMIL1* and *ZmDOT1* transcripts accumulated during procambium initiation, and *ZmSHR1*, *ZmSCR1*, *ZmSCR1H*, *ZmNKD1*, *ZmRVN1* and *ZmCRF1* accumulated even later, during cell type differentiation. These data suggest that the genes examined may have roles in distinct stages of Kranz development.

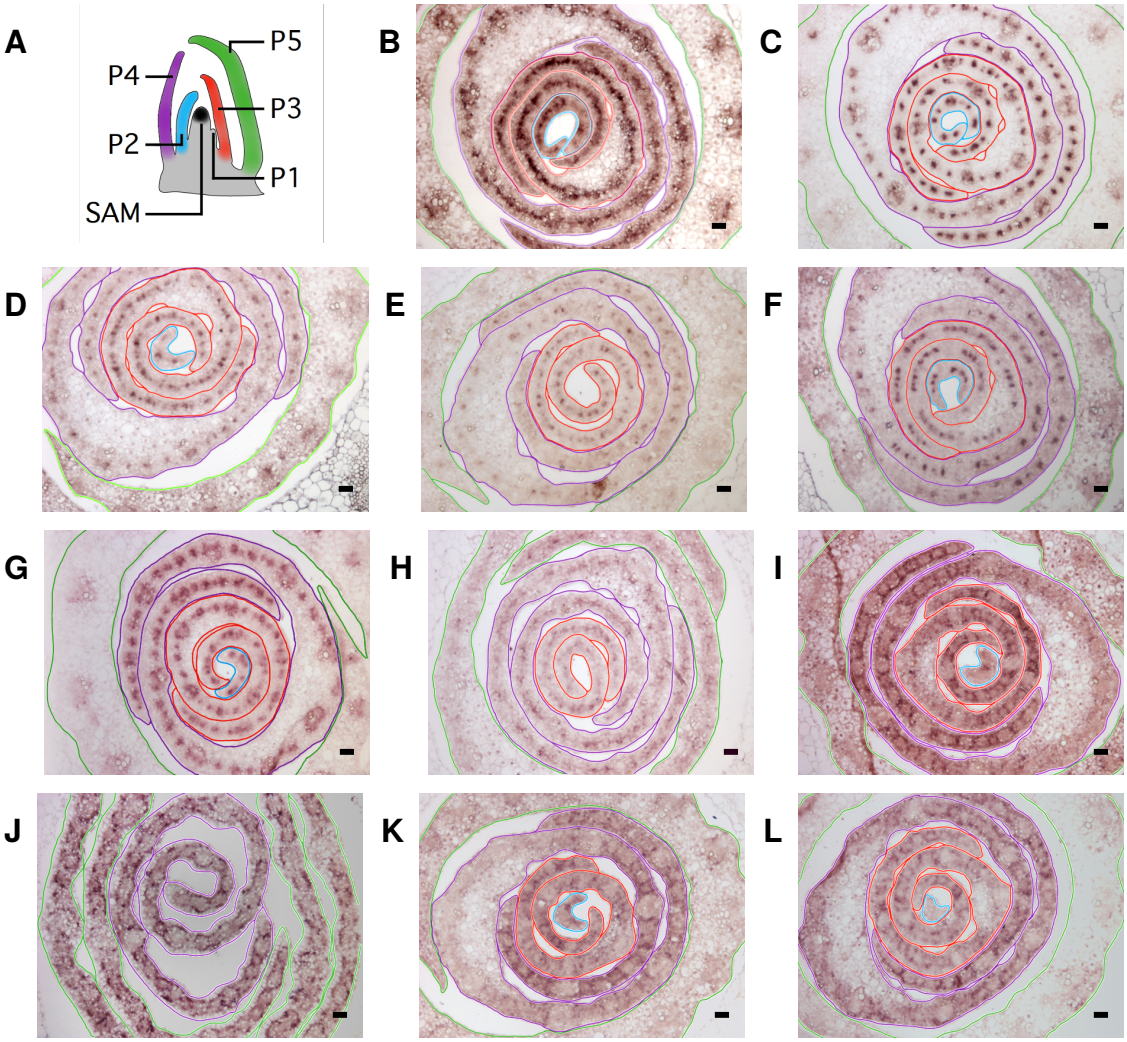


Figure 3.1: Summary of transcript localisation patterns for 11 candidate Kranz regulators. A) Diagram showing longitudinal cross section of monocot shoot apex, shoot apical meristem (SAM) and plastochrons 1 to 5 (P1-5) are colour coded. **B-L)** mRNA accumulation in 4 day old maize seedlings, visualised using *in situ* hybridisation with gene-specific anti-sense RNA probes, images overlaid with colour coded (as in A) plastochron outlines. Genes with mRNA accumulation in ground tissue and procambium- *ZmMIL2* (B), *ZmSHR2* (C), *ZmANT1* (D), *ZmMIL1* (E), *ZmDOT1* (F). Genes with transcript localisation in differentiating bundle sheath, mesophyll or vascular cells- *ZmSHR1* (G), *ZmRVN1* (H), *ZmNKD1* (I), *ZmSCR1* (J), *ZmSCR1H* (K), *ZmCRF1* (L). Scale bar- 50 µm. (Higher magnification images are presented in later figures)

3.2.2 *ZmMIL2* transcript accumulation marks the leaf cell layer that gives rise to vascular tissue

During maize leaf development, the innermost cell layer gives rise to all the vasculature and bundle sheath cells, as well as the mesophyll cells in the central leaf layer (Sharman, 1942; Esau, 1943; Langdale *et al.*, 1989). Strikingly, *ZmMIL2* mRNA was localised to this layer before and during the initiation of Kranz anatomy (fig. 3.2; p.70). This is the first gene to be identified as a marker for cells of this cell lineage, and the spatio-temporal pattern of transcript accumulation overlaps with cells that are competent to form procambium. After procambium initiation, *ZmMIL2* RNA becomes restricted to the bundle sheath precursor cells (fig. 3.2, E).

3.2.3 Procambial centres are initiated in a single cell in ground tissue

3.2.3.1 *ZmSHR2*, *ZmANT1* and *ZmMIL1* mRNA accumulates in the procambium mother cell

Transcripts encoding three transcription factors were detected at the earliest stage of maize procambium development. *ZmSHR2*, *ZmANT1* and *ZmMIL1* mRNA accumulated in a single cell (referred to here as the procambium mother cell - PMC) in the ground tissue layer. The mother cell was always two ground tissue cells away from a more mature vascular bundle (fig. 3.3; p.72).

3.2.3.2 Inhibition of procambium initiation correlates with the presence of *ZmDOT1* mRNA

The accumulation of *ZmDOT1* transcripts was also first detected in a single ground tissue cell, however, this was not the PMC. Interestingly, *ZmDOT1* mRNA first accumulates in the third ground tissue cell away from a more mature vascular

bundle (fig. 3.4; p.74). The lineage of this cell was investigated using the longitudinal developmental gradient that can be seen by examining sequential sections of the shoot apex. From this analysis, it was observed that the PMC undergoes an asymmetric cell division and that *ZmDOT1* transcripts accumulate in only one of the daughter cells (fig. 3.5; p.76). *ZmDOT1* mRNA is detected only at early stages of procambium development, and new PMCs are seen next to veins that do not contain detectable *ZmDOT1* mRNA. These results suggest that *ZmDOT1* could act to inhibit a procambium initiating signal.

3.2.4 Transcripts encoding distinct combinations of transcription factors accumulate in mesophyll versus bundle sheath cells during specification

3.2.4.1 *SHORTROOT* transcripts are differentially expressed in bundle sheath cells

Transcripts of both *SHR*-related genes (*ZmSHR1* and *ZmSHR2*) were detected in vascular bundles, with *ZmSHR1* being expressed later in development than *ZmSHR2* (fig. 3.6; p.78). The localisation of *ZmSHR2* mRNA in the developing vascular bundle is similar to *AtSHR* expression during arabidopsis leaf development (Cui *et al.*, 2014). However, *ZmSHR1* mRNA was also observed in the BS cells. Given the known role of AtSHR in radial tissue patterning, the ZmSHR proteins may control differentiation of distinct cell types during Kranz anatomy development.

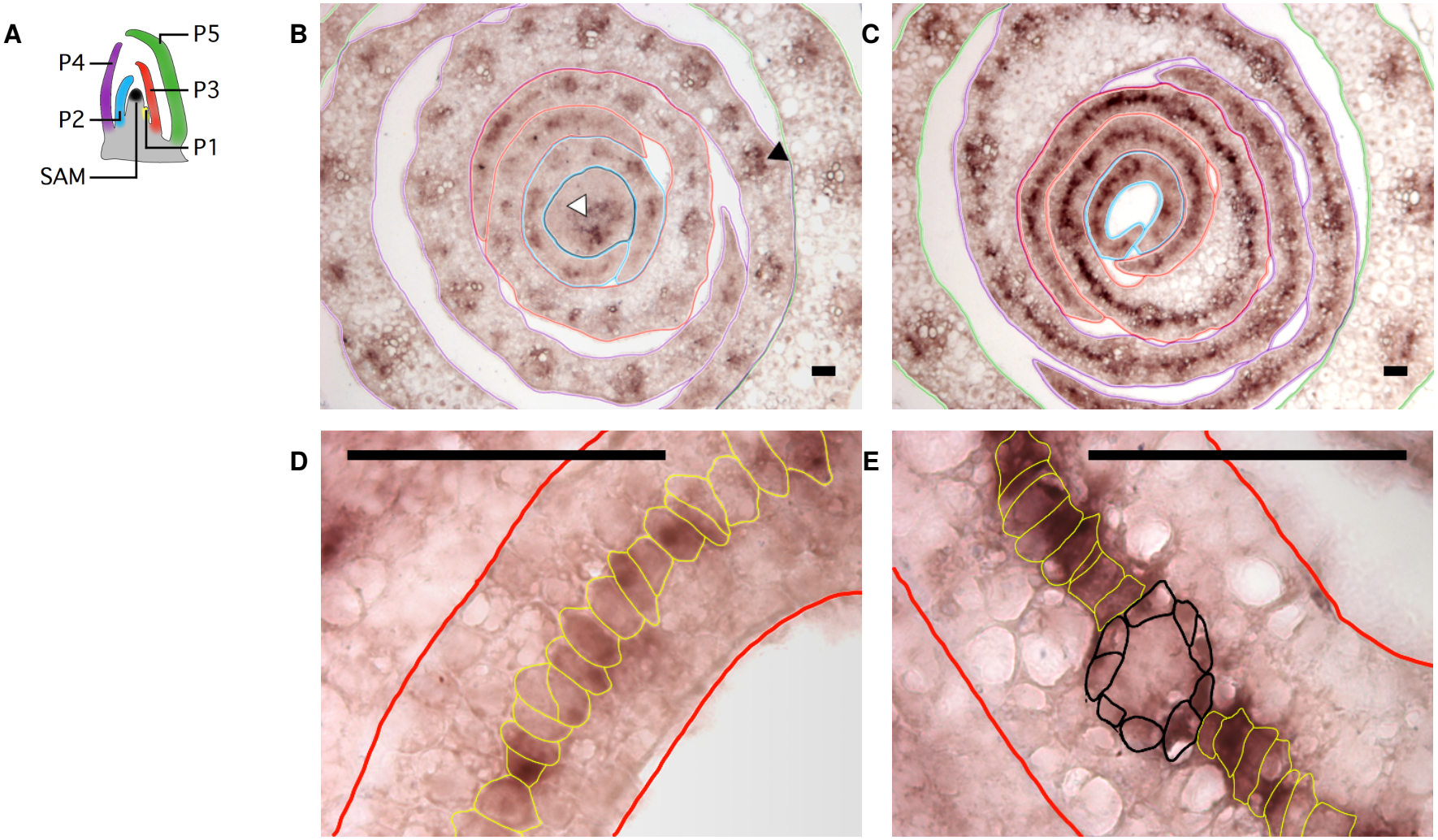


Figure 3.2: *ZmMIL2* mRNA localisation during maize foliar development. Figure legend is on p.71.

Figure 3.2: *ZmMIL2* mRNA localisation during maize foliar development

A) Diagram showing longitudinal cross section of monocot shoot apex with colour coded shoot apical meristem (SAM) and plastochrons 1 to 5 (P1-5). **B-E)** *ZmMIL2* mRNA accumulation in 4 day old maize seedlings, visualised using *in situ* hybridisation, images overlaid with colour coded plastochron outlines. Shoot apex sections in (B) showing - SAM, P2 and P3 (blade), P4 and P5 (sheath), mRNA is in SAM (white arrow) and in the leaf sheath vascular bundles (black arrow). Sections in (C) showing - P2, P3 and P4 (blade), P5 (sheath) with transcript accumulation in the innermost cell layer of blade tissue. P3 blade tissue is magnified to show *ZmMIL2* mRNA in undifferentiated ground tissue cells in the central cell layer (yellow) (D); and in a developing vascular bundle with bundle sheath cells (black) in addition to the undifferentiated ground tissue (yellow) (E). Scale bar- 50 µm.

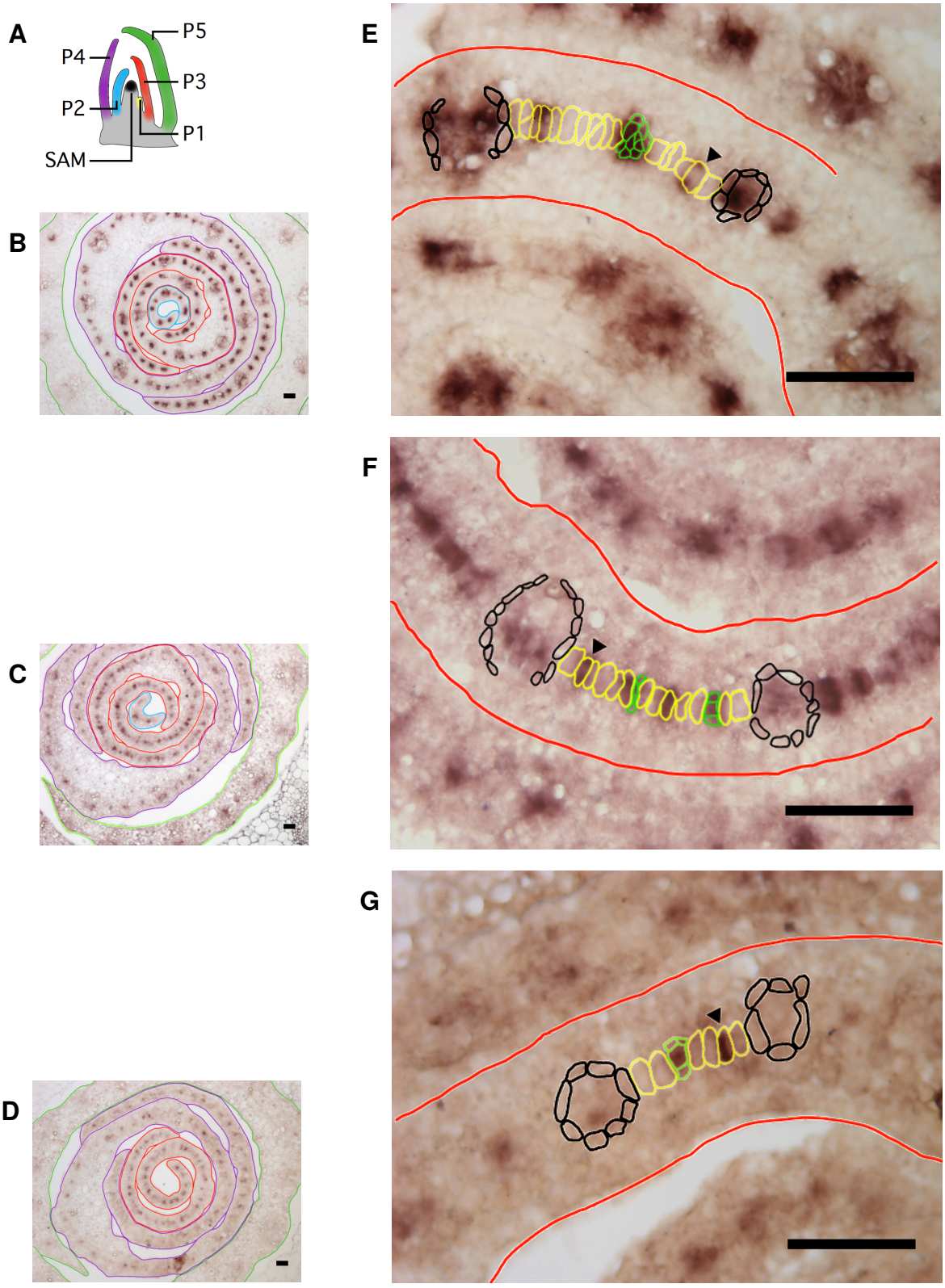


Figure 3.3: *ZmSHR2*, *ZmANT1* and *ZmMIL1* transcripts localise to the procambium mother cell during Kranz development. Figure legend is on p.73.

Figure 3.3: *ZmSHR2*, *ZmANT1* and *ZmMIL1* transcripts localise to the procambium mother cell during Kranz development. A) Diagram showing longitudinal cross section of monocot shoot apex with colour coded shoot apical meristem (SAM) and plastochrons 1 to 5 (P1-5). **B-G)** mRNA accumulation in 4 day old maize seedlings, visualised using *in situ* hybridisation and with *ZmSHR2* (B), *ZmANT1* (C) and *ZmMPL1* (D) images overlaid with colour coded plastochron outlines. Shoot apex sections showing transcript localisation in developing foliar tissues- P2, P3 and P4 (blade) and P5 (sheath). Magnification of P3 blade tissue showing mRNA localised to bundle sheath cells (black), procambium (green) and undifferentiated tissue (yellow) for *ZmSHR2* (E), *ZmANT1* (F) and *ZmMPL1* (G). All three transcripts are first detected in the procambium mother cell (arrow) which is always the second ground tissue cell away from a mature vascular bundle. Scale bar- 50 µm.

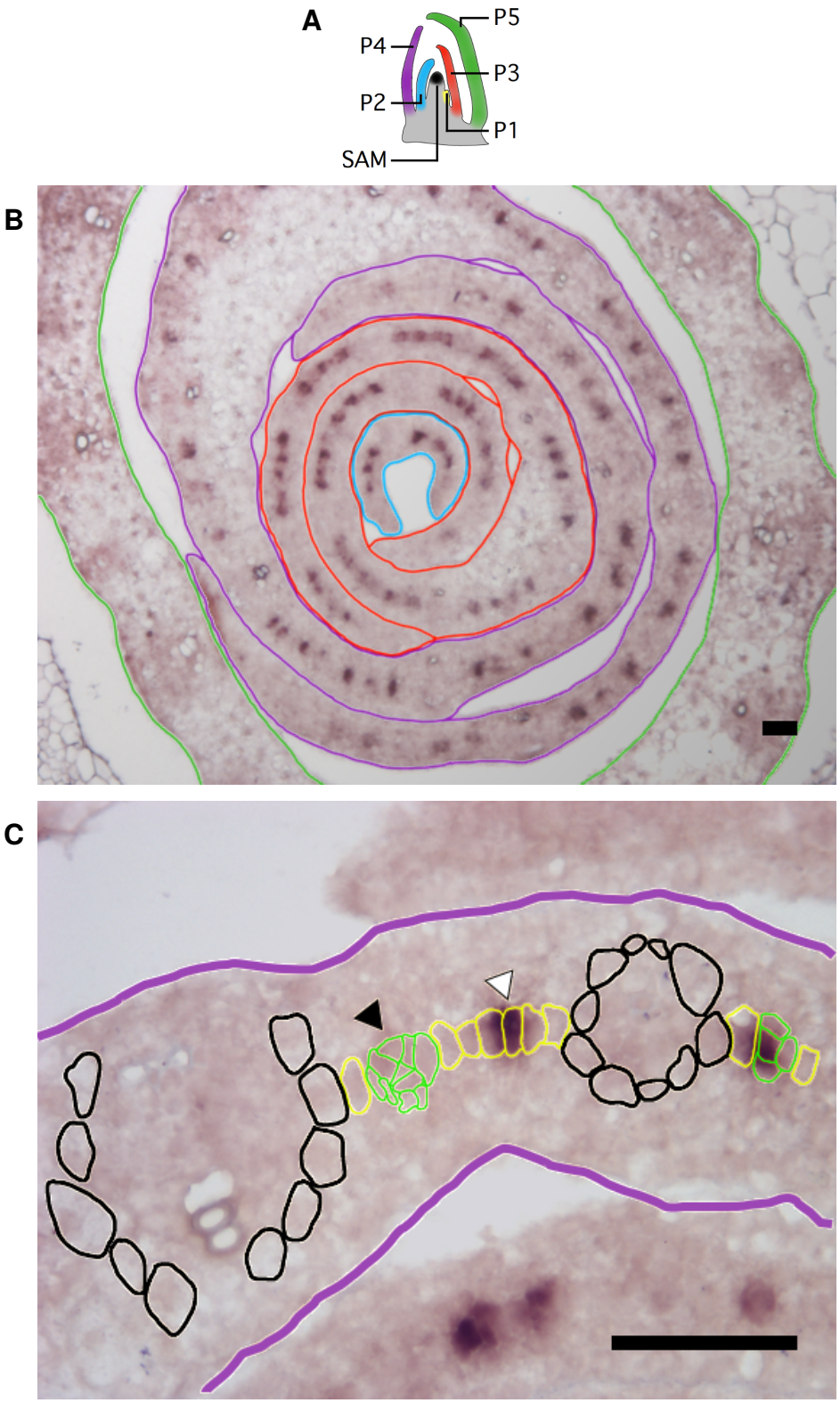


Figure 3.4: Presence of *ZmDOT1* transcript is negatively correlated with procambium initiation in adjacent cells. Figure legend on p.75.

Figure 3.4: Presence of *ZmDOT1* transcript is negatively correlated with procambium initiation in adjacent cells. **A)** Diagram showing longitudinal cross section of monocot shoot apex, with shoot apical meristem (SAM) and plastochrons 1 to 5 (P1-5) colour coded. **B-C)** *ZmDOT1* mRNA accumulation in 4 day old maize seedlings, visualised using *in situ* hybridisation, and with images overlaid with colour coded plastochron outlines. Transcript is localised to procambium in P2-P4 (blade tissue) (B). Magnification of P4 blade tissue shows bundle sheath cells (black) do not contain *ZmDOT1* mRNA when procambium initiation (white arrow) is occurring in adjacent ground tissue (yellow). *ZmDOT1* mRNA accumulates in procambium tissue (green) but levels decrease as veins develop (black arrow). Scale bar- 50 µm.

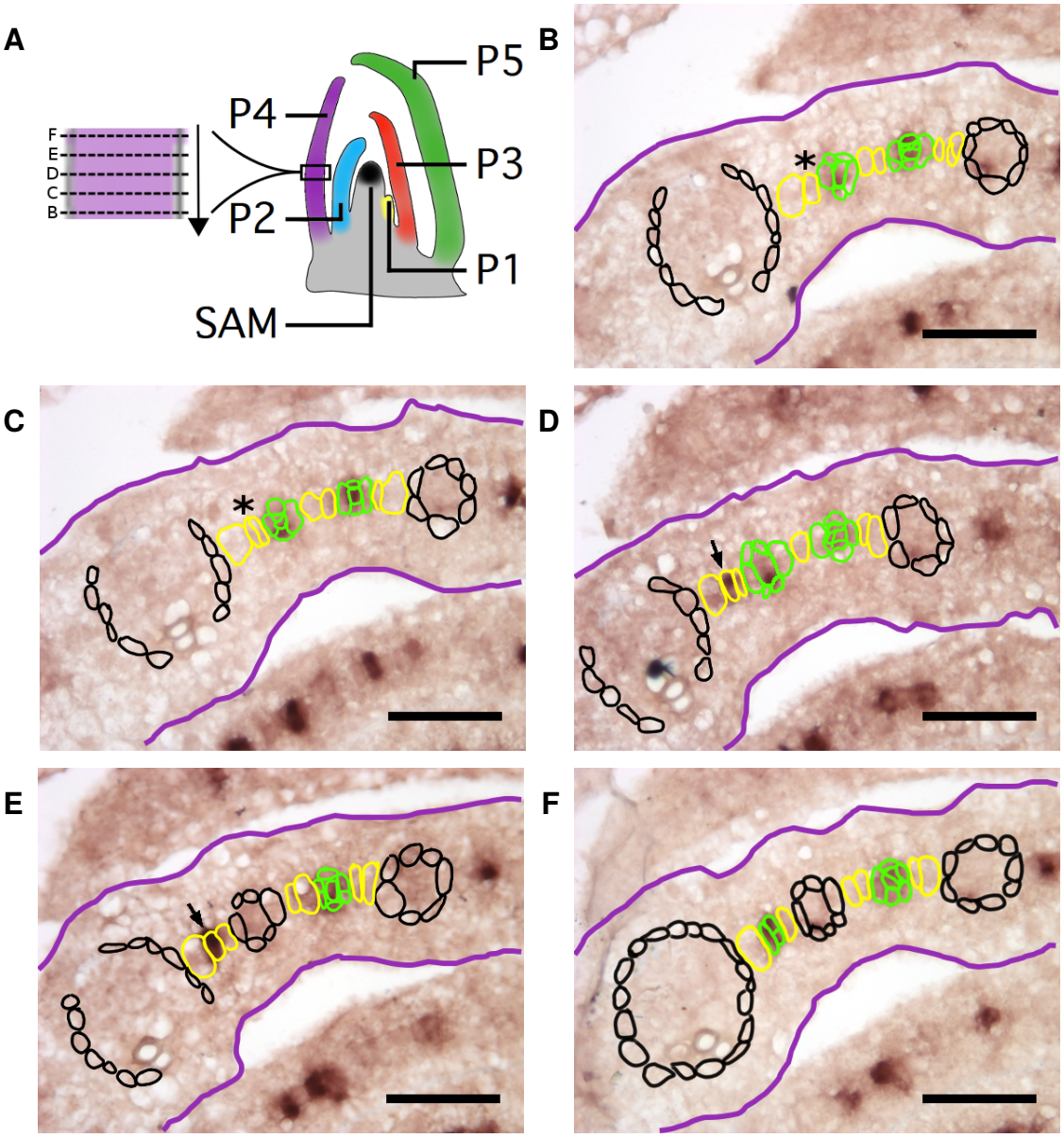


Figure 3.5: *ZmDOT1* transcript accumulates in procambium after an asymmetric cell division near to a mature vein. Figure legend on p.77.

Figure 3.5: *ZmDOT1* transcript accumulates in procambium after an asymmetric cell division near to a mature vein. A) Diagram showing longitudinal cross section of monocot shoot apex, with shoot apical meristem (SAM) and plastochrons 1 to 5 (P1-5) colour coded. Call-out shows position of sequential transverse sections (B-F), illustrating the development of an intermediate vein from tip to base of P4 (direction of vein growth indicated by arrow). **B-F)** *ZmDOT1* mRNA accumulation in a 4 day old maize seedling, visualised using *in situ* hybridisation, and with images overlaid with colour coded plastochron outline. Following an asymmetric cell division in a procambium mother cell (asterisk) (B-C), *ZmDOT1* transcript accumulates in one of the daughter cells (arrow) (D-E). Cell division occurs to form procambium (F). Differentiated bundle sheath cells- black, ground tissue- yellow, procambium- green. Scale bar- 50 µm.

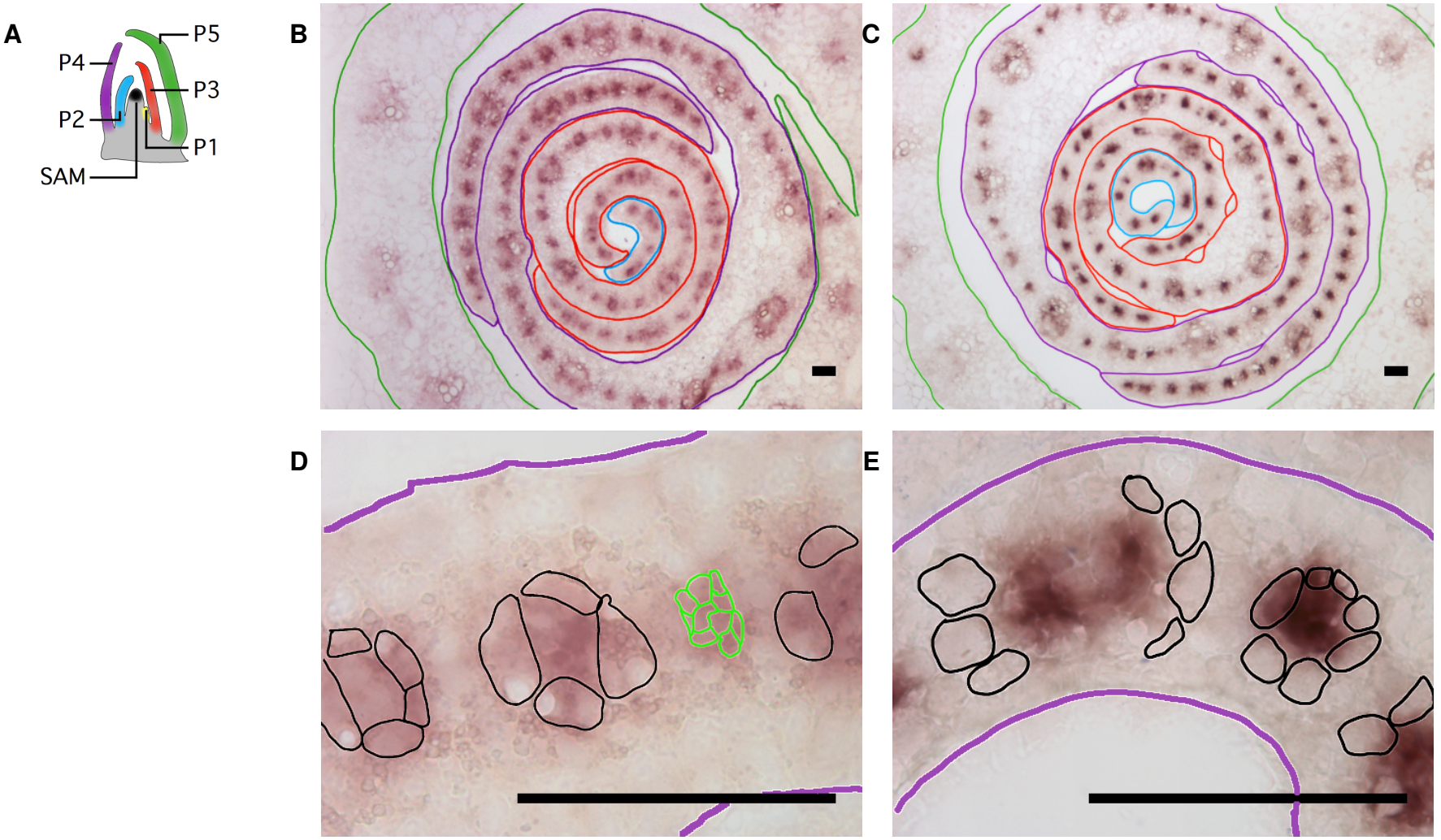


Figure 3.6: *ZmSHR1* mRNA is present in an additional cell layer compared to *ZmSHR2* during Kranz cell differentiation.

Figure legend is on p.79.

Figure 3.6: *ZmSHR1* mRNA is present in an additional cell layer compared to *ZmSHR2* during Kranz cell differentiation

A) Diagram showing longitudinal cross section of monocot shoot apex with colour coded shoot apical meristem (SAM) and plastochrons 1 to 5 (P1-5). **B-E)** mRNA accumulation in 4 day old maize seedlings, visualised using *in situ* hybridisation and with *ZmSHR1* (B) and *ZmSHR2* (C) images overlaid with colour coded plastochron outlines. Localisation of transcripts in developing foliar tissues; P2, P3 and P4 (blade) and P5 (sheath). Magnification of P4 blade tissue showing *ZmSHR1* (D) and *ZmSHR2* (E) mRNA localised to procambium (green) and bundle sheath cells (black). Scale bar- 50 µm.

3.2.4.2 The *INDETERMINATE DOMAIN* genes *ZmNKD1* and *ZmRVN1* are differentially expressed in mesophyll and bundle sheath cells

Transcripts of the two *INDETERMINATE DOMAIN* (IDD) genes that were tested showed different accumulation patterns; *ZmNKD1* transcripts were localised in M cells, whereas *ZmRVN1* mRNA was detected in BS cells (fig. 3.7; p.82). In arabidopsis, the endoderm-specific expression of two IDD genes (*AtMGP* and *AtJKD*) leads to root tissue patterning through the *SHR-SCR* pathway (Welch *et al.*, 2007; Ogasawara *et al.*, 2011). Similarly, the cell-specific mRNA accumulation patterns of *ZmNKD1* and *ZmRVN1* may have a role in BS and M cell differentiation during maize leaf development.

3.2.4.3 *ZmSCR* transcript accumulation patterns change during Kranz cell differentiation

ZmSCR1 and *ZmSCR1H* transcript accumulation patterns differed depending on the developmental stage of the foliar tissue. During BS cell differentiation and procambium initiation, mRNA was localised to BS cells, and to the cell layer that was adjacent to the cell file that accumulated *ZmMIL2* mRNA (fig. 3.8, B-E; p.84). Later in Kranz development, mRNA was detected in the M cell layer with the strongest signal being observed in M precursor cells prior to cell division (fig. 3.8, F-I; p.84). The different *ZmSCR* mRNA accumulation patterns may indicate that the ZmSCR proteins are involved in both BS and M cell-specification events. If this is the case, the proteins would not be sufficient to pattern the two cell types, but may act in concert with other genes which are differentially expressed in the BS and M precursor cells during Kranz development.

3.2.4.4 Expression of a cytokinin-responsive gene in differentiated mesophyll cells

Transcripts of an AP2 gene (*ZmCRF1*), that has homology to arabidopsis cytokinin-responsive genes, were detected in mesophyll cells (fig. 3.9; p.86). These genes have been implicated in facilitating cell expansion during leaf development in a cytokinin-dependent manner (Rashotte *et al.*, 2006). Localisation of *ZmCRF1* mRNA in differentiated mesophyll cells could therefore reflect a role in promoting cell expansion in the leaf blade.

3.2.5 Candidate regulator ortholog expression reveals distinct differences between rice and maize leaf development

3.2.5.1 Candidate regulator ortholog expression is significantly lower during rice leaf development compared to maize foliar leaves

To investigate whether differences in the localisation of candidate Kranz regulator genes were correlated with leaf anatomy differences between rice and maize, the mRNA localisation patterns of orthologous rice genes was tested using ISH on sections from rice shoot apices (see 3.2.5.2 - 3.2.5.5). Orthologs were inferred using maximum-likelihood phylogenetic analysis in which amino acid sequences from seven eudicot and seven monocot species were used. *Amborella trichopoda* was included as a species that diverged before the monocot-eudicot divergence, and *Physcomitrella patens* was used as an outgroup. To resolve the trees and increase branch length accuracy, basal monocot and eudicot species were also included; *Acquilegia coerulea* diverged after the monocot-eudicot divergence (at the base of the eudicots), whilst *Spirodela polyrhiza* diverged at the base of the monocots. All the candidate maize transcription factor encoding genes had closely related rice orthologs (table 3.2; p.87) (see appendix A for phylogenetic trees).

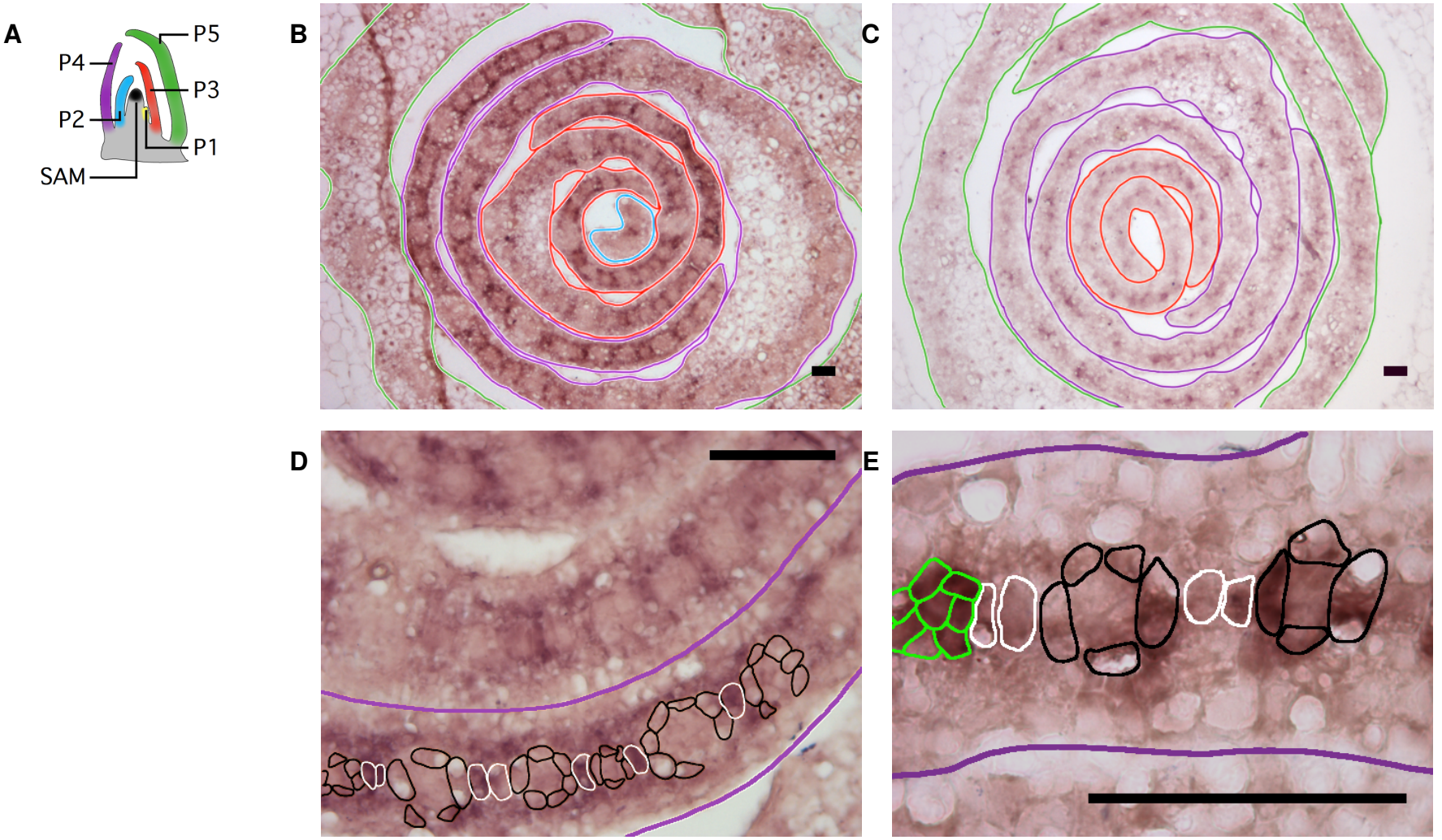


Figure 3.7: *INDETERMINATE DOMAIN* genes *ZmNKD1* and *ZmRVN1* show distinct transcript localisation patterns during Kranz cell differentiation. Figure legend is on p.83.

Figure 3.7: *INDETERMINATE DOMAIN* genes *ZmNKD1* and *ZmRVN1* show distinct transcript localisation patterns during Kranz cell differentiation. A) Diagram showing longitudinal cross section of monocot shoot apex with colour coded shoot apical meristem (SAM) and plastochrons 1 to 5 (P1-5). **B-E)** mRNA accumulation in 4 day old maize seedlings, visualised using *in situ* hybridisation and with *ZmNKD1* (B) and *ZmRVN1* (C) images overlaid with colour coded plastochron outlines. Transcripts were detected in developing foliar tissues- P2, P3 and P4 (blade) and P5 (sheath). Magnification of P4 blade tissue shows differential mRNA localisation patterns- to mesophyll cells (white) for *ZmNKD1* (D) or bundle sheath cells (black) and procambium (green) for *ZmRVN1* (E). Scale bar- 50 µm.

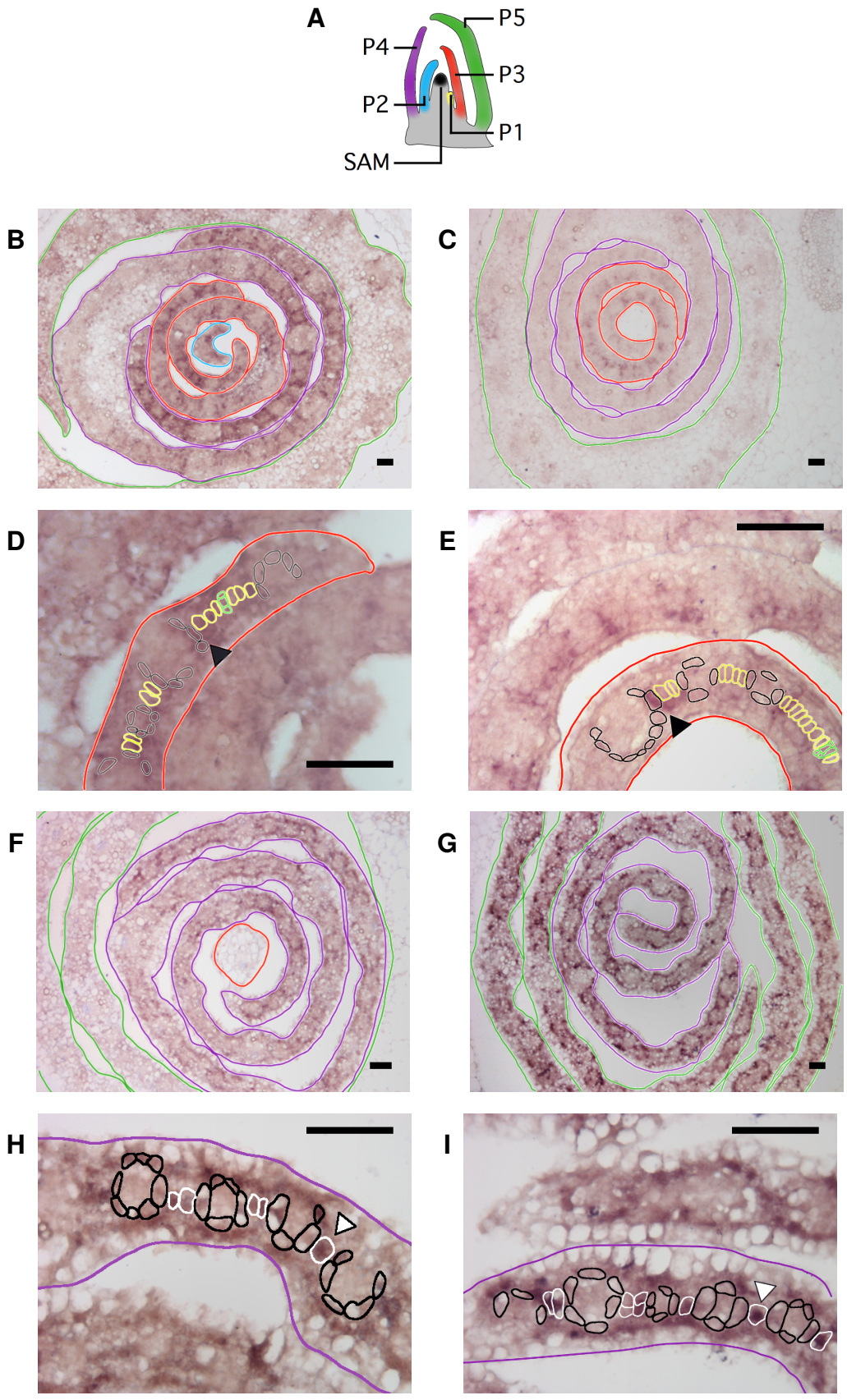


Figure 3.8: mRNA localisation of the homeologs *SCR1* and *SCR1H* differs during Kranz development. Figure legend is on p.85.

Figure 3.8: The mRNA localisation of the homeologs *SCR1* and *SCR1H* differs during Kranz development. A) Diagram showing longitudinal cross section of monocot shoot apex with colour coded shoot apical meristem (SAM) and plastochrons 1 to 5 (P1-5). **B-E)** Images showing mRNA accumulation in 4 day old maize seedlings, visualised using *in situ* hybridisation, and with *ZmSCR1* (B) and *ZmSCR1H* (C) images overlaid with colour coded plastochron outlines. Magnification of P3 blade tissue shows *ZmSCR1* (D) and *ZmSCR1H* (E) mRNA localisation in bundle sheath cells (black) and around the ground tissue (black arrows near yellow cells) but not around procambium (green) **F-I)** mRNA localisation in P4 (blade) and P5 (sheath) foliar tissue undergoing late stages of bundle sheath and mesophyll cell differentiation for *ZmSCR1* (F) and *ZmSCR1H* (G). Magnification of P4 blade tissue highlighting bundle sheath (black) and mesophyll cell (white) differentiation with *ZmSCR1* (H) and *ZmSCR1H* (I) mRNA localised to mesophyll cells with signal being particularly strong in cells prior to division (white arrow). Scale bar- 50 µm.

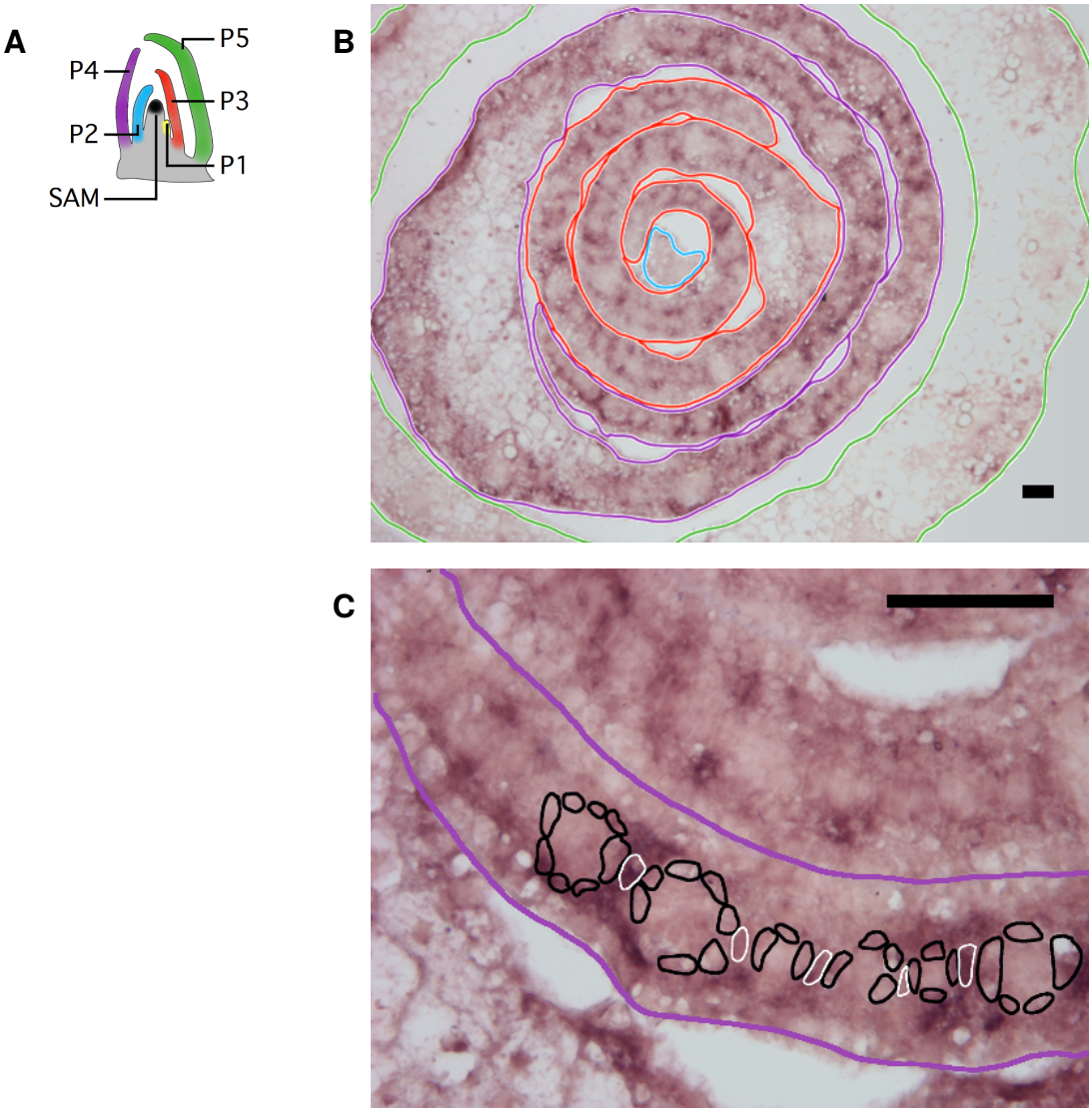


Figure 3.9: *ZmCRF1* mRNA localised to mesophyll cells during Kranz cell differentiation. A) Diagram showing longitudinal cross section of monocot shoot apex, shoot apical meristem (SAM) and plastochrons 1 to 5 (P1-5) are colour coded. **B-C)** *ZmCRF1* mRNA accumulation in 4 day old maize seedlings, visualised using *in situ* hybridisation and with images overlaid with colour coded plastochron outlines. Transcript is localised to mesophyll cells in P3-P4 (blade) (B). Magnification of P4 blade tissue shows mRNA accumulates in mesophyll cells (intervein, white) but not bundle sheath cells (black). Scale bar- 50 µm.

In addition to ISH analysis, the transcriptome of developing rice leaves (van Campen *et al.*, 2016) was used to investigate mRNA abundance of the candidate Kranz regulator orthologs and it was found that in all cases, expression levels of the rice genes were lower than the orthologous genes in maize (table 3.2; below). Strikingly, *OsRVN1* was not expressed during rice leaf development. These data imply that the regulation of procambium development and cell specification in rice leaves may be significantly different from C_4 leaf development in maize.

Gene ID	TF Family	Gene clade	Gene name	Maize ortholog	P3	P4	P5
LOC Os03g31880	GRAS	SHR	*OsSHR2*	*ZmSHR1*	12	3	0
LOC Os07g39820	GRAS	SHR	*OsSHR1*	*ZmSHR2*	35	4	0
LOC Os11g03110	GRAS	SCR	*OsSCR1*	*ZmSCR1* *ZmSCR1H*	33	4	1
LOC Os12g02870	GRAS	SCR	*OsSCR2*	*ZmSCR1* *ZmSCR1H*	41	6	1
LOC Os11g06840	C2H2	MRPI	*OsMIL1*	*ZmMIL1*	3	1	0
LOC Os04g59380	C2H2	MRPI	*OsMIL2*	*ZmMIL2*	11	5	0
LOC Os09g13680	C2H2	DOT5	*OsDOT1*	*ZmDOT1*	10	1	0
LOC Os04g47860	C2H2	IDD	*OsNKD1*	*ZmNKD1*	7	3	1
LOC Os01g09850	C2H2	IDD	*OsRVN1*	*ZmRVN1*	0	0	0
LOC Os01g46870	AP2	CRF	*OsCRF1*	*ZmCRF1*	20	5	0
LOC Os03g12950	AP2	ANT	*OsANT1*	*ZmANT1*	20	1	0
LOC Os06g46770	*N/A*	UBI	*OsUBI*	*ZmUBI1*	394	182	213

Table 3.2: Putative Kranz regulator rice orthologs. Rice orthologs for transcription factor encoding genes identified as putative Kranz regulators in Wang *et al.* (2013b) and Fouracre *et al.* (2014). Orthologs were inferred using maximum-likelihood phylogentic analysis and annotated with gene family and clade. A ubiquitously expressed gene (*OsUBI*) is shown as a reference for high expression levels. mRNA levels are shown in transcripts/million in rice leaf primordia; plastochron (P)- 3, 4 and 5 (van Campen *et al.*, 2016).

These differences in abundance also account for the high background seen in the ISH experiments (see fig.3.10; p.90 and fig.3.13; p.96). Although general trends can still be distinguished, especially with direct observation of the slides, further optimisation is needed to decrease the non-specific DIG signal. This can be carried out by increasing hybridisation and wash stringency or by designing new probes that have less off-target binding.

3.2.5.2 *MIL2* transcripts do not localise to undifferentiated ground tissue in rice leaves

OsMIL2 transcripts localised to proliferating cells in the mestome sheath and phloem tissue of developing veins in all rice plastochrons (fig. 3.10; p.90). Unlike *ZmMIL2* in maize, no signal was detected in procambium tissue during rice leaf development. However, the transcript accumulation pattern was similar to that seen in the C_3-like leaf sheath tissue of maize (see fig. 3.2, B; p.70). These patterns may indicate that MIL2 has been recruited from a role in C_3 vascular cell proliferation to a role in ground tissue and procambium cell proliferation in C_4 tissues.

3.2.5.3 Procambial initiation in rice is confined to regions adjacent to mature veins

The difference in *MIL2* mRNA accumulation patterns in rice and maize suggests that the regulation of procambial initiation may differ between C_3 and C_4 tissues. To visualise the order of procambium initiation in developing rice leaves, *OsDOT1* transcript accumulation patterns were examined (fig. 3.11; p.92). As in maize, OsDOT1 transcripts localised to procambial cells that were situated next to more mature vascular bundles. However, unlike in maize, ground tissue cell proliferation appeared to be confined to the regions between a mature vein and recently initiated procambium (i.e. no proliferation occurred between two developing procambium centres). The small zone of ground tissue proliferation in rice may explain the differences between C_3 and C_4 leaf vein density; in rice it is possible that only cells in this proliferating region are competent to initiate procambium.

3.2.5.4 *OsDOT1* transcripts are detected later in rice development than *ZmDOT1* in maize

The earliest stages of *OsDOT1* transcript accumulation in rice were similar to *ZmDOT1* in maize in that mRNA was detected in a single cell near to a mature vein that did not have detectable *OsDOT1* transcripts (fig. 3.12; p.94). The exact position of the *OsDOT1* marked cell relative to the adjacent vascular bundle could not be confirmed because rice leaf primordia grow at an angle relative to the overall shoot growth axis. This skew resulted in oblique sectioning through vasculature and meant that cell patterning during leaf development was not as clear as in maize transverse sections. However, it was clear that *OsDOT1* mRNA was present throughout procambium development and that transcripts were detected at much later stages of vein development as compared to *ZmDOT1* in maize. These differences in the timing of *DOT1* RNA accumulation may be related to the differences in procambium initiation in maize and rice. It is possible that DOT1 may act as an inhibitor of procambium initiation in monocots. If this were the case, it would be predicted that the shorter window of *DOT1* expression in maize would increase the rate of procambium initiation, leading to increased vein density, and therefore enabling the formation of Kranz anatomy. This hypothesis could be tested by altering *DOT1* expression timing, a shorter window of *DOT1* expression would be predicted to increase procambium initiation rates.

3.2.5.5 Gene expression in procambium differs between maize and rice

Transcripts of *OsANT1*, like *ZmANT1* in maize, were detected in rice procambium cells (fig. 3.13; p.96). However, *OsMIL1* and *OsSHR2* (the *ZmSHR1* ortholog) transcripts were only detected in more developed veins. This suggests that the roles of OsMIL1/ZmMIL1 and OsSHR1/ZmSHR2 are not conserved between rice and maize leaf development.

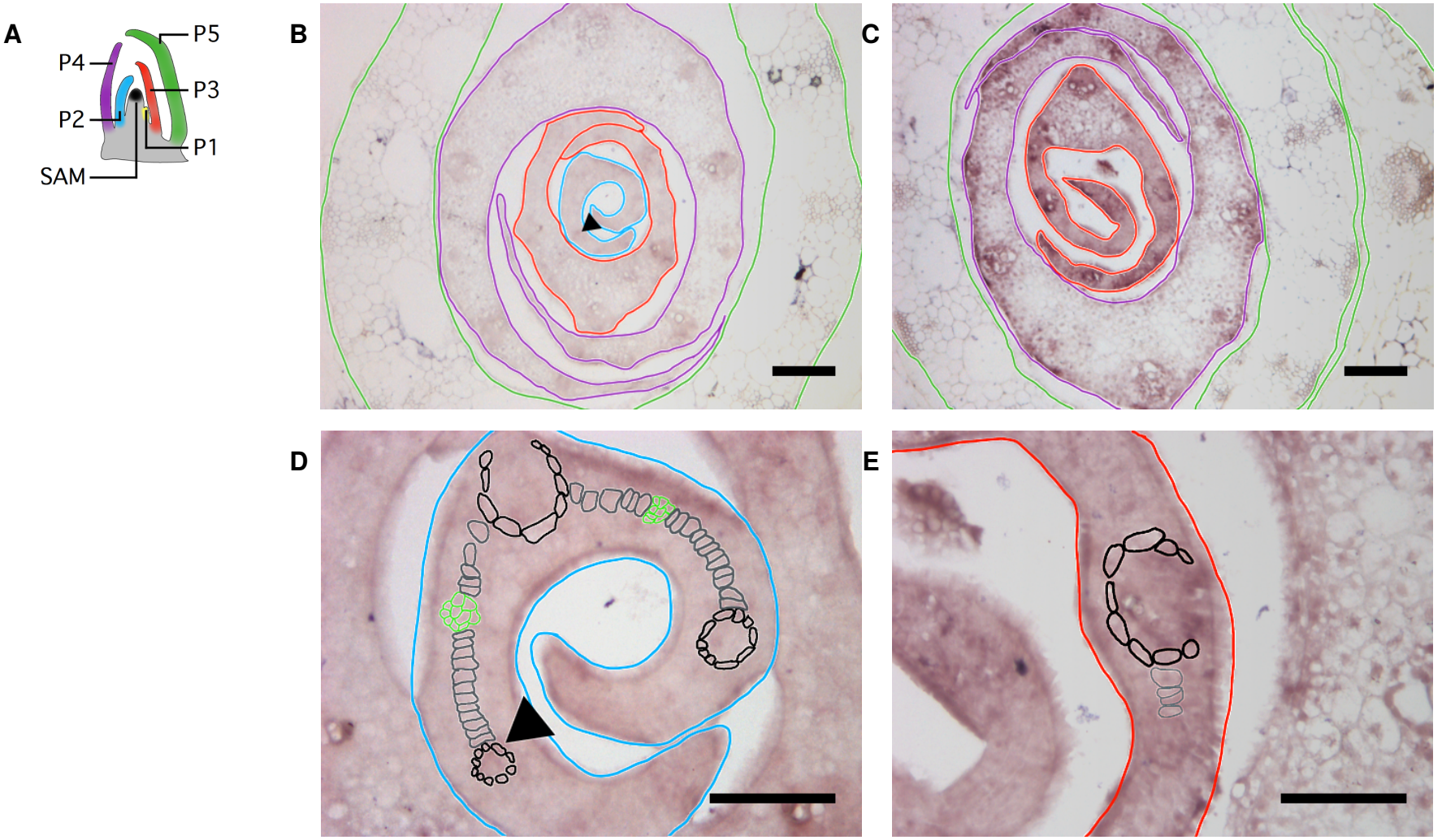


Figure 3.10: *OsMIL2* mRNA localisation during rice leaf development. Figure legend is on p.91.

Figure 3.10: *OsMIL2* mRNA localisation during rice leaf development. mRNA accumulation in 17 day old rice plants visualised using *in situ* hybridisation and with images overlaid with colour coded plastochron outlines. **A)** Diagram showing longitudinal cross section of monocot shoot apex with colour coded shoot apical meristem (SAM) and plastochrons 1 to 5 (P1-5). **B-C)** mRNA localisation in developing foliar tissues; P2, P3 and P4 (blade) and P5 ,(sheath) (B) and more developmentally advanced P3 and P4 (blade) and P5 (sheath) (C). **D-E)** Cellular resolution of blade tissue showing transcript is only detectable in vascular bundles (black arrow) and mRNA does not localise to bundle sheath cells (black), the inner layer of cells (grey) or young procambium (green)- in the P2 tip (D) or in P3 (E). Scale bar- 50 µm.

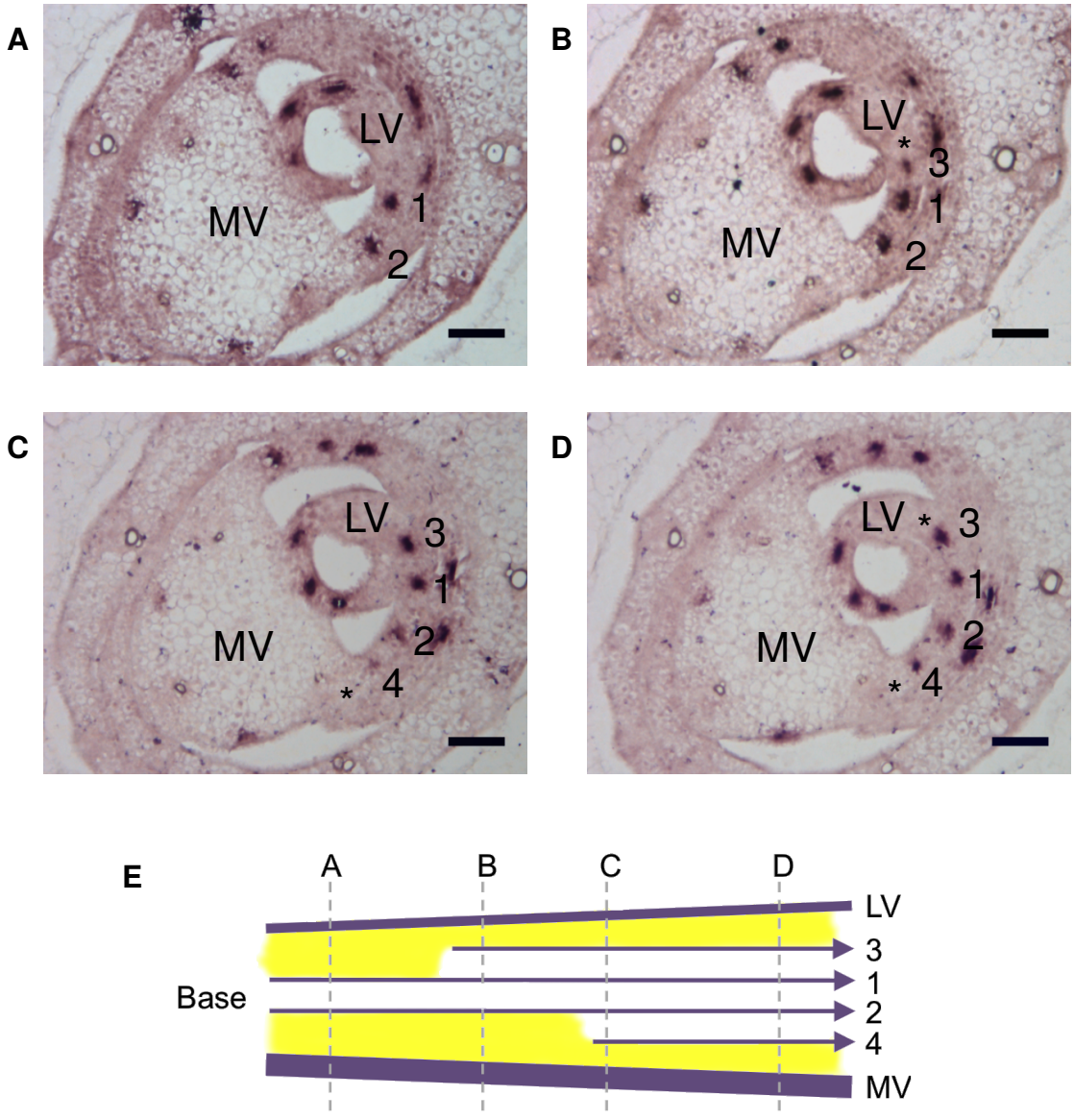


Figure 3.11: Ground tissue cell division in rice foliar leaf predominantly occurs next to mature veins. Figure legend is on p.93.

Figure 3.11: Ground tissue cell division in rice foliar leaf predominantly occurs next to mature veins. A-D) Sequential transverse sections along the longitudinal axis of a rice shoot apex, with *in situ* hybridisation to the procambium marker *OsDOT1* used to visualise developing vasculature. Section shows plastochron 2 with two vascular bundles (1 & 2) between the mid-vein (MV) and lateral vein (LV) (A). Further towards the tip, most leaf growth has occurred near the LV (asterisk) and vein 3 has initiated in this region (B). Moving further towards the tip, most leaf growth has occurred near the MV (asterisk) and vein 4 has initiated in this region (C), closer to the tip, shows more mature veins (D). Scale bar: 50 µm **E)** Longitudinal diagram of veins highlighting position of transverse sections (grey dashed lines) and areas of leaf growth (yellow) (not to scale).

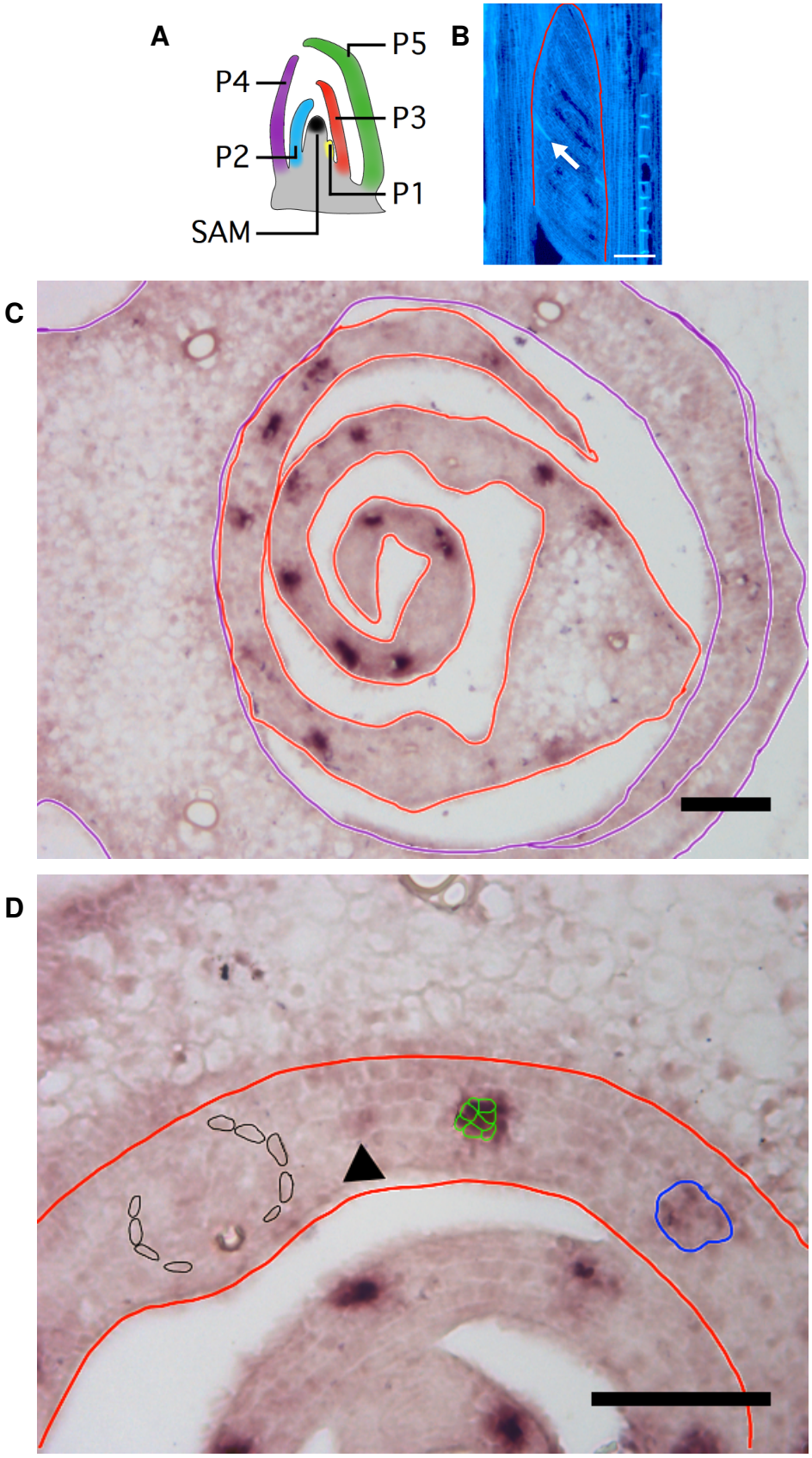


Figure 3.12: *OsDOT1* mRNA accumulation in developing rice leaves continues from early procambium to late stages of procambium development.

Figure legend on p.95

Figure 3.12: *OsDOT1* mRNA accumulation in developing rice leaves continues from early procambium to late stages of procambium development. A) Diagram showing longitudinal cross section of monocot shoot apex with colour coded shoot apical meristem (SAM) and plastochrons 1 to 5 (P1-5). **B)** UV image of rice shoot apex longitudinal section stained with calcofluor-white, showing that P3 (red outline) grows at an angle (white arrow) to the overall shoot growth axis. **C-D)** *OsDOT1* mRNA accumulation in 17 day old rice plants, visualised using *in situ* hybridisation and with images overlaid with colour coded plastochron outlines. Section showing developing foliar tissues- P3 and P4 (C). Magnification of P3 blade tissue (D), mRNA accumulation begins in ground tissue (black arrow) near to developed bundle sheath cells (black) and is also present in procambium (green). Transcript levels are still detectable at late stages of procambium development (blue). Scale bar- 50 µm.

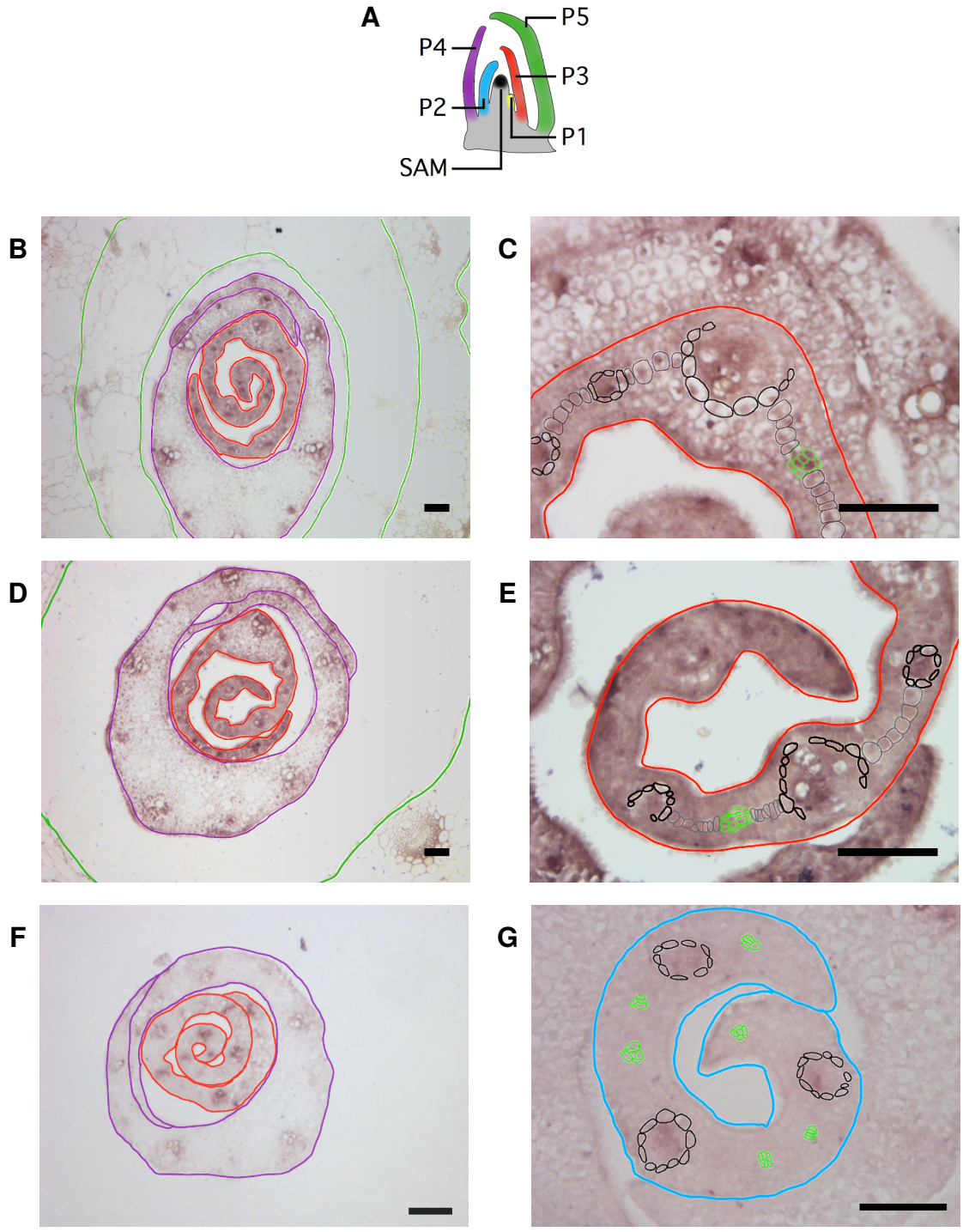


Figure 3.13: *OsANT*, *OsMIL1* and *OsSHR2* mRNA accumulates in vasculature. Figure legend is on p.97.

Figure 3.13: *OsANT*, *OsMIL1* and *OsSHR2* mRNA accumulates in vasculature. A) Diagram showing longitudinal cross section of monocot shoot apex with colour coded shoot apical meristem (SAM) and plastochrons 1 to 5 (P1-5). **B-G)** mRNA accumulation in 17 day old rice plants, visualised using *in situ* hybridisation and with images overlaid with colour coded plastochron outlines. *OsANT1* mRNA localisation in developing foliar tissues- P3 and P4 (blade) and P5 (sheath)(B). Magnified view of P3 (C) showing signal is localised to mature vascular bundles (ringed by bundle sheath cells, black) and procambium (green). No expression is seen in other cells between veins (grey). *OsMPL1* (D, E) and *OsSHR2* (F, G) transcripts are localised to mature vascular bundles (ringed by bundle sheath cells, black) and no detectable signal is seen in procambium (green) or other cells between veins (grey) in P3 and P4 (blade) and P5 (sheath) (D, F), magnified view of P3 (E), magnified view of P2 (G). Scale bar- 50 µm.

3.3 Discussion

3.3.1 Model of Kranz anatomy development in maize

3.3.1.1 Competence of ground tissue to initiate procambium

The maize *in situ* results identified developmental markers for key stages of Kranz development (fig.3.14; p.100), allowing refined models for procambium initiation and cell-type specification during Kranz development to be proposed. In the Kranz procambium initiation model (fig. 3.15; p.103) it is proposed that veins initiate from ground tissues that are marked by *ZmMIL2* expression. Procambium initiation is characterised by the co-expression of *ZmANT1*, *ZmMIL1* and *ZmSHR2* in a single ground tissue cell that is two cells away from a vascular bundle which is no longer expressing *ZmDOT1*. It is interesting to note that in arabidopsis, procambium initiation is marked by *ATHB8* and *SHR* expression in cells directly adjacent to existing procambium (Scarpella *et al.*, 2004; Gardiner *et al.*, 2011). This is not the case with monocot procambium initiation and may explain the difference in veination patterning between eudicots and monocots. In maize, the procambium initiation model invokes a signal from a mature vein that is required to initiate procambium and that travels several cell layers. Given the expression profile, ZmDOT1 could act to inhibit this signal, or it may inhibit procambium initiation in neighbouring cells using a more direct mechanism such as intercellular protein movement.

3.3.1.2 SHORTROOT-SCARECROW pathway in mesophyll and bundle sheath cell patterning

The tissue types in Kranz anatomy (vein, BS and M) are specified at different stages of maize leaf development, and *SHR-SCR* pathway related genes (*ZmSHR1*, *ZmSHR2*, *ZmSCR1*, *ZmSCR1H*, *ZmRVN1* and *ZmNKD1*) are expressed during these stages in tissue-specific patterns. Given that the *SHR-SCR* pathway

is known to regulate cell-type specification of endodermal tissue in plant roots, the Kranz cell specification model predicts that the maize genes play a role in BS and M cell specification during Kranz development (fig. 3.16; p.105).

BS differentiation occurs before M cell specification during Kranz development and gene expression patterns suggest that the *SHR-SCR* pathway may act to specify this cell type. ISH data indicates that the *IDD* gene *ZmRVN1* is expressed in cells adjacent to a vascular bundle (the BS cell layer). Given that the IDD proteins AtMGP and AtJKD are known to activate *AtSCR* gene expression, the Kranz cell-specification model (fig. 3.16, A) proposes that the ZmRVN1 protein activates the expression of *ZmSCR1* and *ZmSCR1H* genes in the BS precursor cell. ISH shows that the two *ZmSHR* genes are expressed in vascular bundles during leaf development and because SHR proteins are mobile, the model proposes that SCR proteins which are present in the BS layer act to stop SHR moving more than one cell away from the vein during early Kranz development. The SHR-SCR complex, alongside other transcriptional regulators which are expressed in the developing BS cell, thus act to specify BS cell fate.

The Kranz cell specification model also proposes that the SHR-SCR complex is involved in M cell differentiation (fig. 3.16, B). The *IDD* gene *ZmNKD1* transcripts localise to the cells between mature BS cells (the M cell precursors) and therefore the ZmNKD1 protein may act to activate the transcription of *ZmSCR* genes in this tissue type. After BS cells have differentiated, *ZmSHR1* transcripts are detected in the BS cells, and *ZmSHR2* mRNA is present in the vascular bundles. Therefore the model proposes that the SCR protein present in the M cell precursors forms a complex with the SHR proteins that have moved from the vein and BS cells. This SHR-SCR complex, alongside other transcriptional regulators which are expressed in the developing M cell, thereby act to specify M cell fate.

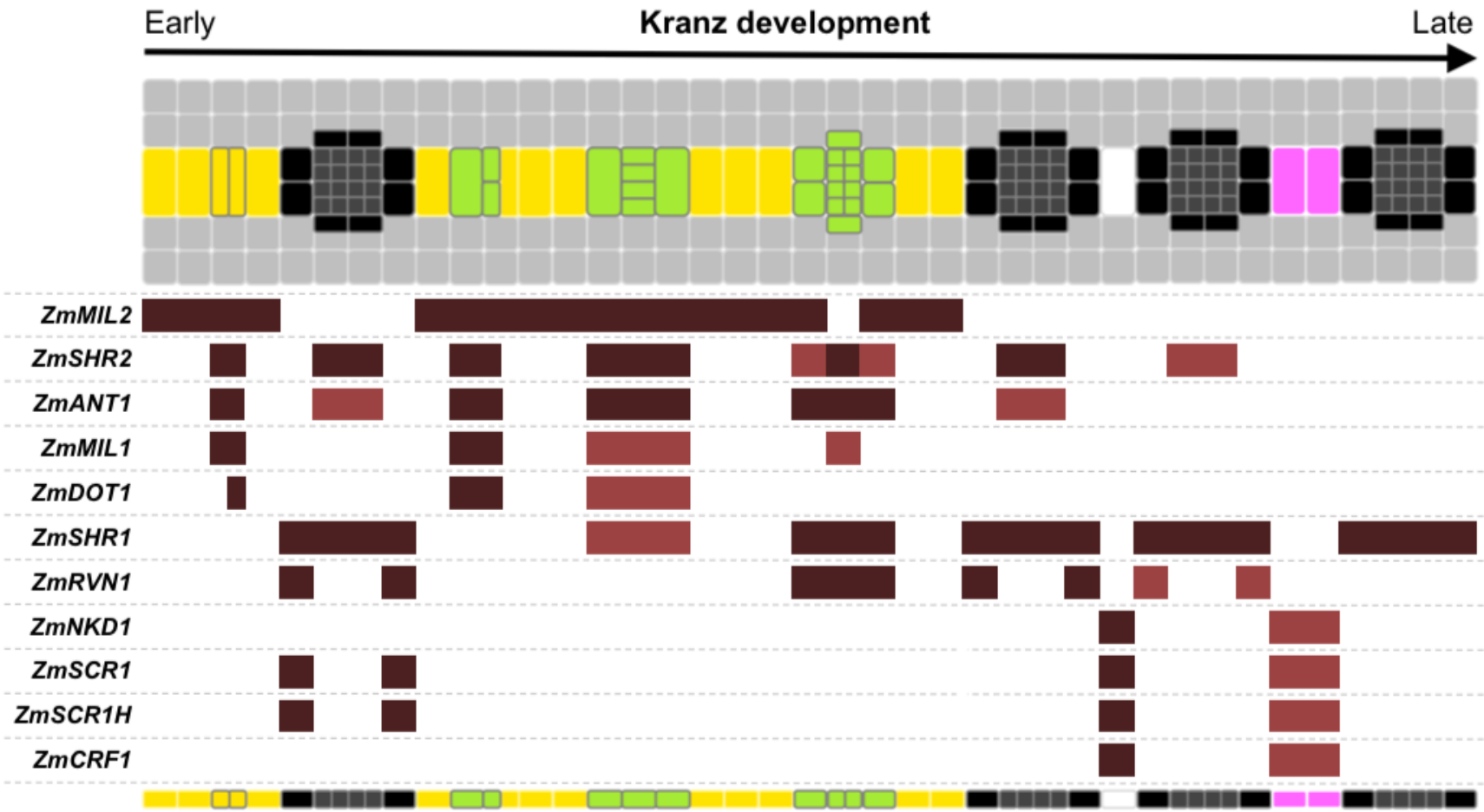


Figure 3.14: Summary of mRNA accumulation of candidate genes during Kranz development in maize.

Figure legend on p.101.

Figure 3.14: Summary of mRNA accumulation of candidate genes during Kranz development in maize. Diagram showing key stages of Kranz development, from ground tissue (yellow) to initial cell division of procambium mother cell (yellow with grey outline), to procambium development (green) and differentiation of bundle sheath (black) and mesophyll cells (precursor- white, mesophyll- pink). Location of mRNA accumulation of 11 candidate genes at each developmental stage is shown (horizontal brown bars). Relative *in situ* hybridisation signal strength for each gene is shown using dark (higher signal strength) or light (lower signal strength) coloured bars.

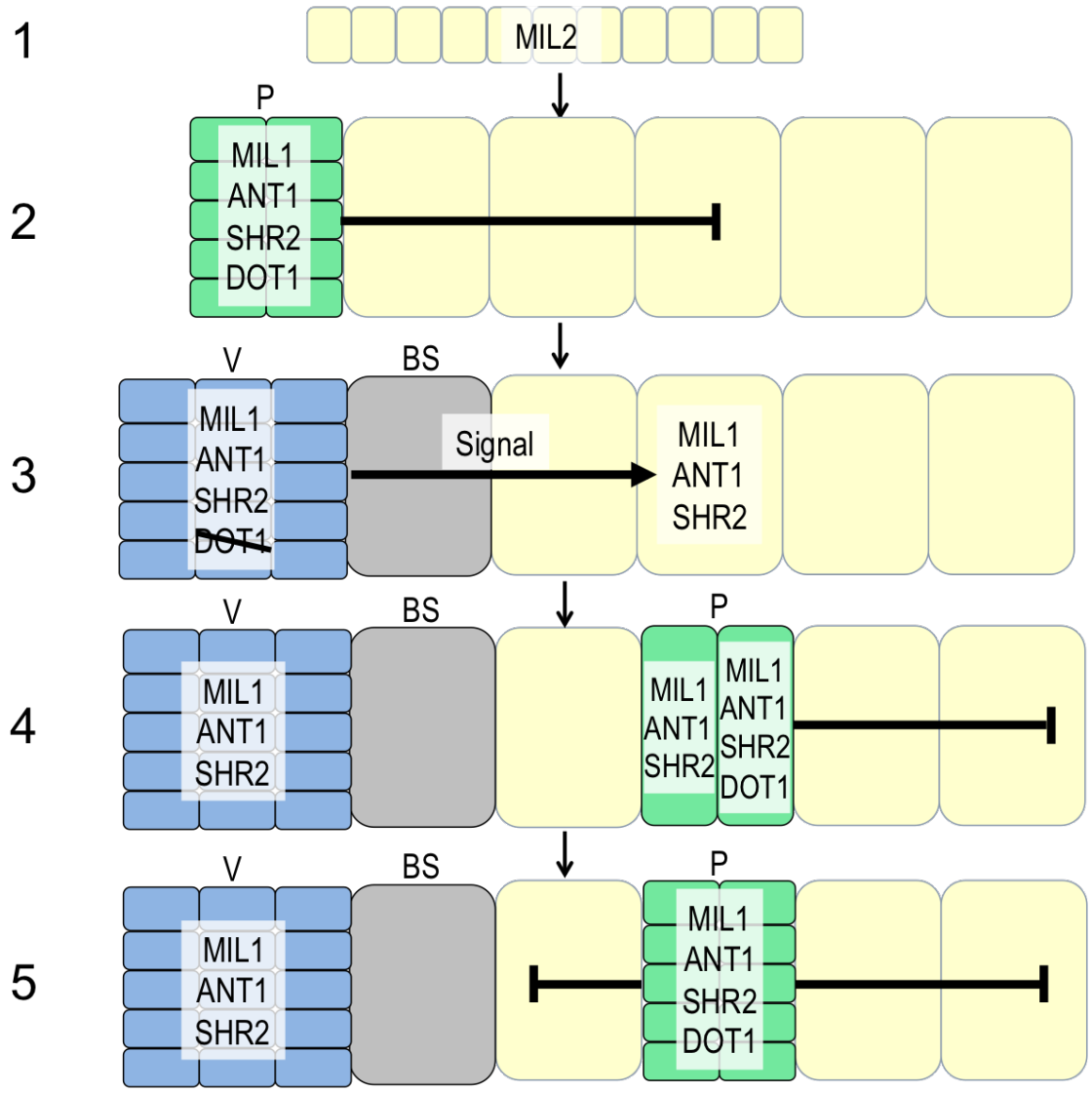


Figure 3.15: Model for vein initiation during Kranz anatomy development.

Figure legend on p.103.

Figure 3.15: Model for vein initiation during Kranz anatomy development. 1) MRPI-LIKE 2 (MIL2) is localised to the innermost cell layer in a developing leaf, where the protein delimits the undifferentiated cells (yellow) that have competency to develop into vasculature during Kranz development. **2)** Procambium is initiated from one of the undifferentiated cells, and as it divides it is characterised by *MIL1*, *AINTEGUMENTA 1* (*ANT1*), *SHORTROOT 2* (*SHR2*) and *DEFECTIVELY ORGANIZED TRIBUTARIES 1* (*DOT1*) accumulation. DOT1 inhibits procambium initiation in adjacent cells. **3)** The procambium differentiates into vasculature (blue), and bundle sheath cells (grey) differentiate around the vein. At this stage DOT1 transcripts are no longer detectable allowing signal from the vein/BS cells to induce *MIL1*, *ANT1* and *SHR2* expression in the second undifferentiated cell away from the bundle sheath cell. **4)** This procambium initial cell undergoes an asymmetric division, and one of the daughter cells accumulates *DOT1* mRNA. **5)** The procambial cells divide and the *DOT1* expression domain expands to all procambial cells, inhibiting any procambium initiation in adjacent cells, as in (2).

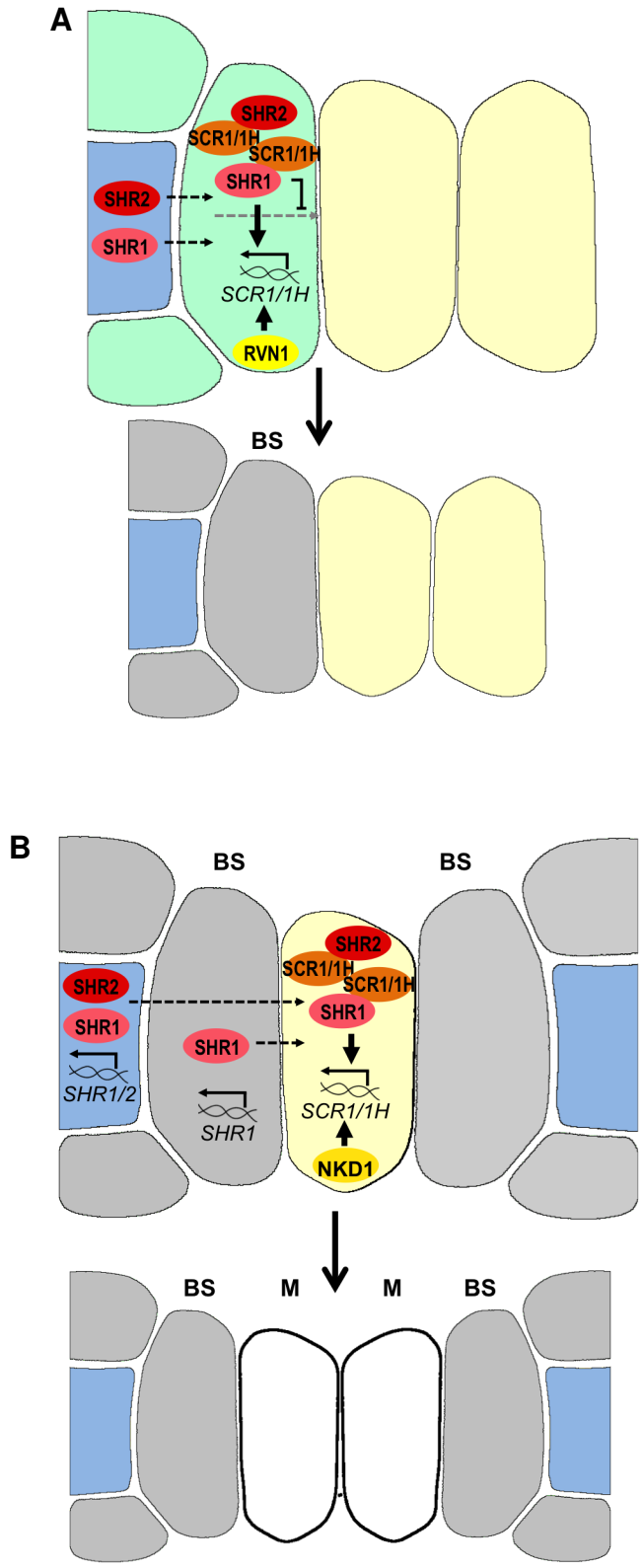


Figure 3.16: Model for bundle sheath and mesophyll cell specification through the SHORTROOT-SCARECROW developmental pathway.

Figure legend on p.105.

Figure 3.16: Model for bundle sheath and mesophyll cell specification through the SHORTROOT-SCARECROW developmental pathway A-B) SHORTROOT and SCARECROW proteins specify both bundle sheath (BS) and mesophyll (M) cells through temporal separation of the differentiation events. A) BS cell specification occurs earlier than M differentiation around a developing vascular bundle, the RAVEN 1 (RVN1) protein (light yellow) in the BS precursor cell (light green) induces *SCARECROW 1* (SCR) (light orange) and *SCARECROW HOMEOLOG 1* (SCR1H) (dark orange) expression in that cell. The SCR1 and SCR1H proteins then limit the movement of SHORTROOT 1 (SHR1) (light red) and SHORTROOT 2 (SHR2) (dark red) proteins from the vein (blue). The SHR-SCR complex, along with other unidentified proteins, induces BS cell differentiation. B) M cell differentiation occurs later in leaf development around a vascular bundle with differentiated BS cells. NAKED 1 (NKD1) (dark yellow) is localised to ground tissue cells that are between differentiated BS cells (grey) and the protein activates the transcription of *SCR1* and *SCR1H* in these cells. *SHR1* is expressed in veins and BS cells, and there is weak vascular expression of *SHR2*. These proteins both move into the M precursor cell and interact with the SCR proteins. The SHR-SCR complex, along with other unidentified proteins, then causes expression of M cell specific genes, and the division of single cells to form the two M cell (white) Kranz pattern (B).

3.3.2 Understanding differences between maize and rice gene expression may enable Kranz anatomy engineering in a C_3 context

3.3.2.1 Ground tissue proliferation is restricted in rice leaf development

Ground tissue cell proliferation is observed throughout the maize foliar blade in young leaf primordia. However, in rice leaves, undifferentiated cell proliferation is restricted to sites that are located between a mature vein and a young procambial centre. Given that rice vein initiation only occurs in regions with proliferating ground tissue, only cells in this region may be competent to initiate procambium. Therefore, to engineer Kranz anatomy in a C_3 context, a developmental mechanism is required that expands the ground tissue domain to include regions between developing procambial centres.

3.3.2.2 Key developmental regulators are present during procambium development in maize

Orthologs of known arabidopsis vascular development regulators (*SHR* and *ANT*) were revealed to be markers of procambium development in Kranz anatomy, and two genes encoding uncharacterised transcription factors (*ZmMIL1* and *ZmMIL2*) had overlapping expression domains with these known regulators. In rice, only *OsANT1* was localised to procambium, with the rice orthologs of *ZmMIL1*, *ZmMIL2* and *ZmSHR1* not expressed until late in vein development. Given that procambium initiation in rice occurs next to vasculature that is expressing *OsMIL1*, *OsMIL2* and *OsSHR1*, it is possible that the expression of the maize genes during Kranz development acts as a positive signal to promote procambium initiation competence in the adjacent ground tissue cells. This knowledge could be used to engineer Kranz anatomy in a C_3 context; expression of these genes in rice procambium may increase the area of tissue that is competent for procambium initiation.

3.3.2.3 Inhibition of procambium initiation differs in maize and rice

DOT1 expression is negatively correlated with procambium initiation in both rice and maize. In the C_3 developmental context, ground tissue proliferation and procambium initiation only occur adjacent to mature vasculature with prolonged *OsDOT1* expression in developing procambial centres presumably inhibiting new vein initiation between such sites. That is, if DOT1 is an inhibitor of vein initiation, the prolonged expression in rice may control vein spacing patterns by decreasing the frequency of vein initiation and thus increasing the number of interveinal mesophyll cells, relative to the maize leaf. The decreased period of DOT1 activity in maize could therefore be the cause of increased vein density, allowing procambium initiation to occur between developing, as opposed to mature, vascular centres. If this is the case, it is likely that *OsDOT1* expression will need to be modulated to allow Kranz engineering in rice.

3.4 Future Perspectives

3.4.1 MIL protein interaction partners could be key Kranz regulators

MIL genes are expressed at key points during Kranz development. However, there is no published functional information for any of the closest orthologs in other species, and constitutive expression of *ZmMIL1* or *ZmMIL2* in rice did not condition any Kranz-like cell patterning traits (P. Wang, personal communication). The function of a pair of maize genes from a neighbouring gene clade has been reported, and these genes were shown to enhance the activity of a Myb transcription factor (Royo *et al.*, 2009). The *MIL* genes may therefore act in transcription factor complexes, enhancing the activity of an interacting partner. A partner protein for *MIL2* may

play an integral role in maintaining competence for procambium initiation during Kranz development. This partner could be identified using pull-down assays with proteins extracted from developing Kranz tissues, and protein expression of the partner in rice may increase the window of competence for procambium initiation.

3.4.2 Downregulation of *DOT1* expression could increase vein density in rice

DOT1 transcript accumulation negatively correlates with procambium initiation in both maize and rice. Rice calli transformed with a construct to constitutively express *ZmDOT1* produce green shoots but do not regenerate (M. Schuler & D. Vlad, personal communication). Given that the differentiation of shoots was not suppressed by ZmDOT1, it may be that the protein does not inhibit regeneration per se, but that regeneration does not occur because vein formation is suppressed. If *DOT1* is an inhibitor of procambium initiation, *OsDOT1* knock-down lines should show increased vein density. Modulating *OsDOT1* expression patterns during rice leaf development may increase vein density and allow the development of Kranz-like leaf anatomy in the C_3 developmental context.

4

The role of *AINTEGUMENTA-like* genes in Kranz development

4.1 Introduction

4.1.1 AINTEGUMENTA-like proteins identified as putative Kranz anatomy regulators

The Kranz development transcriptomics study conducted by Wang *et al.* (2013b) is an ideal platform from which to select putative Kranz regulator genes for further experimentation; transcription factor families that are over-represented in the foliar leaf dataset compared to the husk leaf dataset are of particular interest. A large number of plant transcription factor gene families (Jin *et al.*, 2014) were over-represented in the foliar dataset, including the *AP2-EREBP* family. In maize, this family consists of 127 genes (Zhuang *et al.*, 2010), which makes it the 7th largest transcription factor family (using data from Jin *et al.* (2014)). In the Kranz leaf transcriptomics dataset, this family was the 4th largest. The over-representation of *AP2-EREBP* gene expression during maize foliar leaf development makes this family an interesting target for further work to explore the

role of these genes in Kranz development.

The *AP2-EREBP* family includes a clade of genes related to the *AINTEGUMENTA* (*ANT*) gene (Kim *et al.*, 2006), and four of these *AINTEGUMENTA-like* (*AIL*) genes were identified as putative Kranz development regulators by Wang *et al.* (2013b). This is striking because in arabidopsis and rice, the *AIL* clade consists of only eight genes for each species (Kim *et al.*, 2006), and thus a large proportion of the maize orthologs are implicated in Kranz development. Three AIL proteins were identified as putative positive regulators of Kranz anatomy (gene expression upregulated in C_4 foliar development, compared to C_3-like husk), and one as a putative negative regulator (gene expression downregulated in foliar development compared to husk) (see table 4.1). The prevalence of *AIL* genes in the Kranz regulator analysis suggests that these genes are promising targets for engineering Kranz anatomy into rice.

4.1.2 AIL proteins have developmental functions that are consistent with a role in Kranz development

AIL proteins are implicated in a variety of developmental roles and many of these functions are consistent with the developmental processes that would enable Kranz patterning. The development of Kranz anatomy is characterised by persistent vein initiation during leaf lamina growth in both eudicot and monocot C_4 species (reviewed in Nelson & Langdale, 1989). Vasculature develops from procambial cells that are derived from ground meristematic tissue (Esau, 1965; Foster, 1952), and this meristematic activity must be maintained within the leaf to allow Kranz patterning. AIL proteins, including ANT, have been shown to maintain stem cell activity in the arabidopsis shoot apical meristem (Galinha *et al.*, 2007; Mudunkothge & Krizek, 2012) and two *AIL* genes (*PLETHORA1*, *PLT1* and *PLETHORA2*, *PLT2* , in concert with the *SHORTROOT* pathway, perform a similar

Gene ID	Kranz regulator	FP 1	FP 3&4	FP 5	FI	FE	HP 1	HP 3&4	HP 5	HI	HE	BS	M
GRMZM2G146688	Positive	1336	2065	864	1	1	392	690	249	0	0	0	0
GRMZM2G021573	Positive	1351	1873	922	14	2	425	671	452	6	1	0	0
GRMZM2G399072	Positive	1420	1364	203	0	0	272	12	1	0	1	0	0
GRMZM2G086573	Negative	17	229	454	7	0	743	743	404	4	1	0	0

Table 4.1: *AINTEGUMENTA*-like genes identified as putative Kranz regulators in maize. mRNA expression in relative transcript abundance in- foliar (F) and husk (H) plastochron 1 (P1), 3&4 (P3&4) and 5 (P5), immature (I) and expanded (E) leaves (Wang *et al.*, 2013b), and in mature bundle sheath (BS) and mesophyll (M) cells (Li *et al.*, 2010) .

function in the arabidopsis root stem cell niche (Aida *et al.*, 2004). ANT also functions to maintain meristematic competence in developing arabidopsis leaves (Mizukami & Fischer, 2000; Kang *et al.*, 2007). Given that the maintenance of ground meristem tissue in a developing C_4 leaf is integral to Kranz patterning, and that AIL proteins have been shown to maintain stem cell identity, the *ZmAIL* candidate regulators may fulfil this role in maize leaves during Kranz development.

Cell proliferation is another key factor in Kranz development in that ground meristem tissue must divide to produce procambium, and vascular cells must proliferate to give cell-dense tissues such as phloem (Elo *et al.*, 2009). All characterised AIL proteins have been implicated in promoting cell proliferation (reviewed in Horstman *et al.*, 2014), and over-expression of *AtANT* in arabidopsis has been shown to increase general cell proliferation rates in the leaf, as well as final vascular cell number (Mizukami & Fischer, 2000; Kang *et al.*, 2007). The *ZmAIL* Kranz regulator candidates may thus control cellular proliferation associated with vein development.

4.1.3 AIL proteins interact with auxin and cytokinin signalling networks

Auxin is a key hormonal regulator of vascular patterning and it is therefore expected that auxin localisation and signalling dynamics differ between leaf development in C_3 and C_4 species (Nelson, 2011). In the arabidopsis root apical meristem, *AtPLT* genes have been implicated as downstream targets of auxin signalling through the action of MONOPTEROS (Aida *et al.*, 2004), and *AtANT* and *AtAIL6* have been implicated in auxin signalling in shoot and flower development (Krizek, 2011). In monocots, the expression of the rice *ANT* ortholog *CROWN ROOTLESS 5* (*CRL5*) was induced by auxin treatment during crown root development, potentially through transcriptional activation via an AUXIN RESPONSE FACTOR (ARF)

protein (Kitomi *et al.*, 2011). These studies suggest that *ZmAIL* genes could also be activated by auxin, and may play a role as transducers of auxin signalling during Kranz development.

The cytokinin signalling network is antagonistic to the auxin pathway, and the interplay between the two networks is integral to maintaining organ and tissue development throughout plant growth, including during vascular development (Coenen & Lomax, 1997; Bishopp *et al.*, 2011). Notably, *AIL* genes have been implicated in the cytokinin signalling network. As may be expected from transcription factors that are upregulated by auxin, application of cytokinin downregulates *JcANT* expression in the inflorescence meristem of the eudicot *Jatropha curcas* (Pan *et al.*, 2014), possibly due to auxin signalling inhibition. Direct interactions between *ANT* genes and the cytokinin network have also been shown; in rice root development, *OsCRL5* upregulates expression of a cytokinin negative regulator *Type-A Arabidopsis Response Regulator* (*ARR*) gene, inhibiting cytokinin signalling (Kitomi *et al.*, 2011). Surprisingly, in arabidopsis and poplar roots, cytokinin application appears to upregulate *AtANT* and *PtAIL1* expression during cambial proliferation (Randall *et al.*, 2015a). The contrasting ways that *AIL* gene expression interacts with cytokinin signalling suggests that these genes can act to modulate auxin and cytokinin responses in different ways, dependent on the developmental context. Given increased procambium initiation in the Kranz developmental context, compared to most C_3 species, *AIL* genes could have therefore been key players in the evolution of Kranz anatomy.

4.1.4 Functional domains in AIL proteins

Transcription factors characteristically consist of distinct amino acid domains that mediate DNA binding and transcriptional activation (Liu *et al.*, 1999). Identification of these domains can highlight whether mRNA transcript variants code for func-

tional proteins and also reveal potential functional differences between related proteins. AtANT functional domains consist of a four amino acid sequence in the middle of the protein that is required for nuclear localisation (Krizek & Sulli, 2006), an 80 amino acid transcriptional activation domain in the N' terminal half of the protein (Krizek & Sulli, 2006) and two ca. 70 amino acid AP2 DNA binding domains in the C' terminal half. The two DNA binding domain sequences are highly conserved between rice and arabidopsis AIL proteins (Nole-Wilson *et al.*, 2005; Kim *et al.*, 2006) and are joined by a linker that is critical for correct DNA binding (Krizek, 2003) (fig.4.1; p.115). Nineteen amino acids within the DNA binding domains have been shown to be crucial for binding to the DNA consensus sequence 5′-gCAC(A/G)N(A/T)TcCC(a/g)ANG(c/t)-3′ (Krizek, 2003; Nole-Wilson & Krizek, 2000). The characterisation of ANT protein functional domains allows comparison of the C_4 AIL candidate proteins with C_3 orthologs and differences in the DNA binding or transcriptional activation domains may indicate a regulatory change that played a part in enabling Kranz anatomy development.

4.1.5 Kranz regulators are predicted to change rice vein patterning

The ZmAIL proteins appear to be good candidates for regulators of Kranz patterning, but this must be functionally validated. Wang *et al.* (2013b) define positive Kranz regulators as genes that are more highly expressed in maize C_4 tissue compared to C_3-like tissue. If these regulators are sufficient to induce Kranz, ectopically expressing these genes in a C_3 developmental background should give leaves with C_4-like traits. Conversely, negative Kranz regulators are defined as genes that are down-regulated during Kranz development. This suggests that these proteins are active in C_3 development, and that suppression of gene function should enable the development of C_4-like traits. The aim of this chapter was to

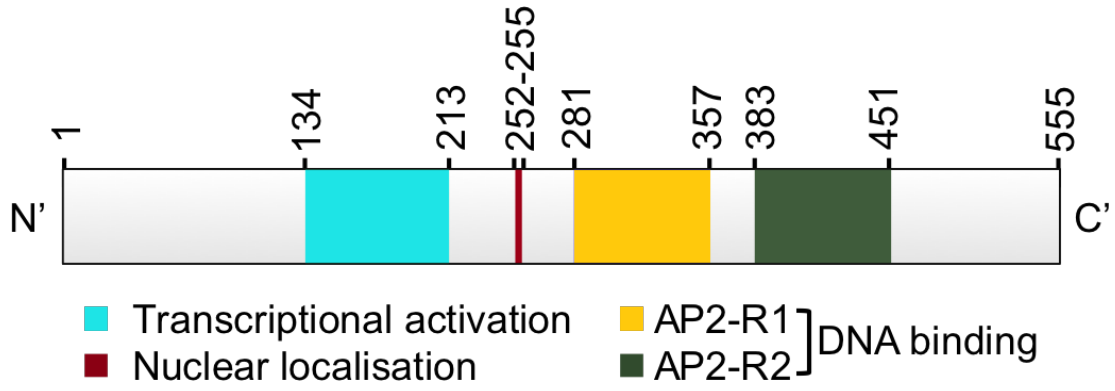


Figure 4.1: Functional domains in the AtANT protein. AtANT protein displayed from amino terminus (N') amino acid $AtANT_1$, to carboxyl terminus (C') $AtANT_{555}$. Transcriptional activation domain is between $AtANT_{134}$-$AtANT_{213}$, nuclear localisation signal between $AtANT_{252}$-$AtANT_{255}$, two AP2 domains are between $AtANT_{281}$-$AtANT_{357}$ and $AtANT_{383}$-$AtANT_{451}$. The amino acid linker between the AP2 domains ($AtANT_{358}$-$AtANT_{382}$) is also required for DNA binding (Nole-Wilson & Krizek, 2000; Krizek, 2003; Krizek & Sulli, 2006).

test the effect of modifying putative Kranz regulator expression on vein patterning in a C_3 background. Rice has been shown to survive significant changes in vein patterning (Scarpella, 2003; Smillie *et al.*, 2012; Ma *et al.*, 2016) and was therefore chosen as the model system. If the *ZmAIL* genes identified by Wang *et al.* (2013b) are Kranz regulators, vein patterns in transgenic rice could be modified.

4.2 Results

4.2.1 AtANT orthologs are putative positive Kranz regulators that underwent gene duplication in monocots

To understand the evolutionary relationships of the four putative AIL Kranz regulators (table 4.1), a maximum-likelihood bootstrap phylogenetic tree was inferred using amino acid sequences from seven eudicot and eight monocot species and RAxML software (Stamatakis, 2014). *Amborella trichopoda* was included as a species that diverged before the monocot-eudicot divergence, and *Physcomitrella patens* was used as an outgroup. To resolve the tree and increase branch length accuracy, basal monocot and eudicot species were also included; *Acquilegia coerulea* diverged after the monocot-eudicot divergence (at the base of the eudicots), whilst *Musa acuminata* and *Spirodela polyrhiza* both diverged at the base of the monocots. The tree shows that all three putative positive Kranz regulators are closely related to the *AtANT* and *AtAIL1* genes, whilst the negative regulator is in the *AIL5* clade (fig.4.2; p.117). To name the *ZmAIL* candidates, similarity to the arabidopsis and rice orthologs was used; *GRMZM2G021573* has already been assigned to *ZmANT1* (see section 3.1.1; p.64), *GRMZM2G146688* will be referred to as *ZmCRL5*, *GRMZM2G399072* as *ZmAIL1* and *GRMZM2G086573* as *ZmAIL5*.

Phylogenetic analysis allows inference of gene duplication events. Gene duplication has been linked with the evolution of novel phenotypes (reviewed in

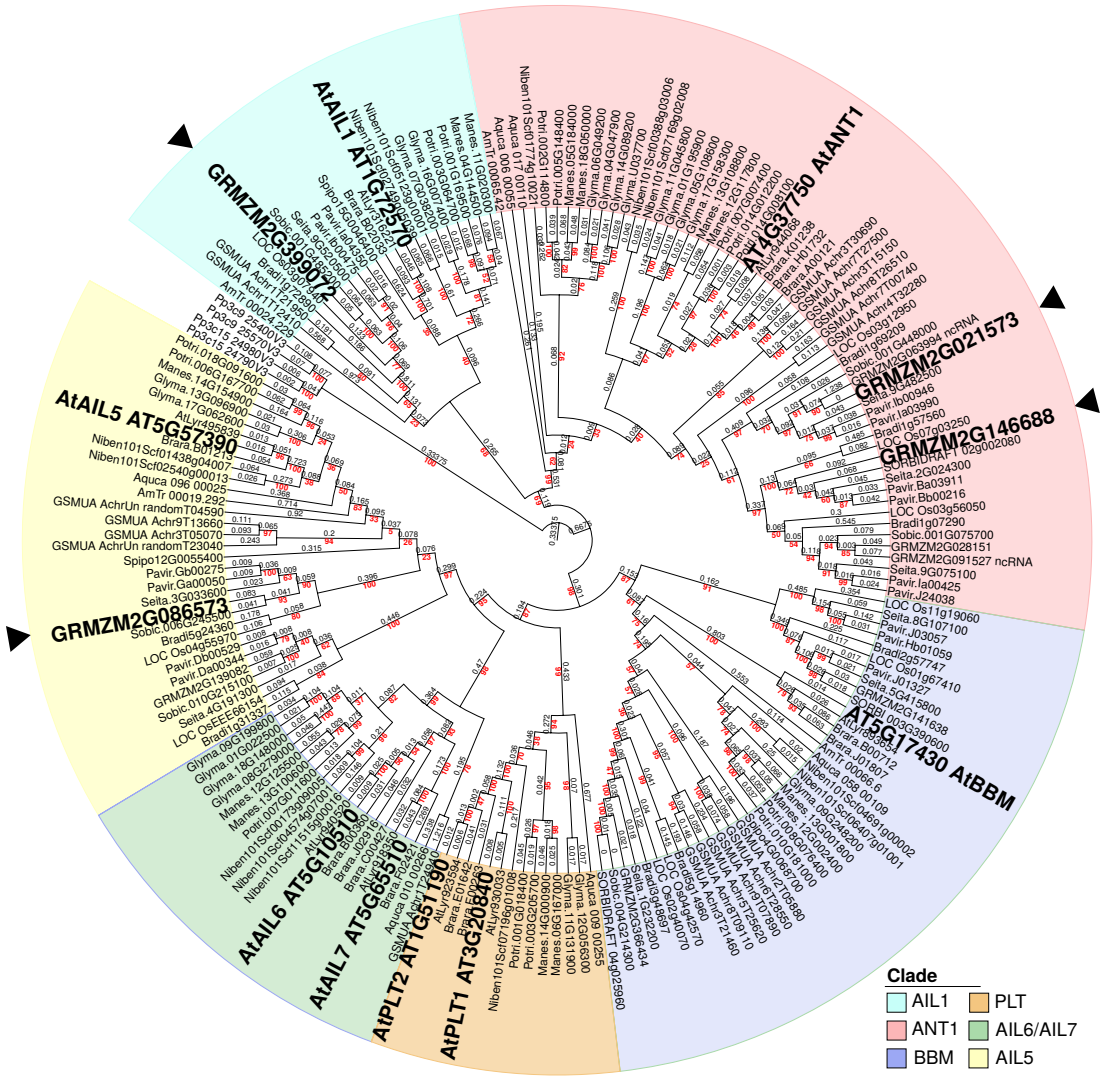


Figure 4.2: AINTEGUMENTA-like protein phylogenetic tree showing putative maize Kranz regulators. Maximum-likelihood inference of phylogentic relationships between the AINTEGUMENTA-like (AIL) proteins in monocot and eudicot species. Branch length measured in number of amino acid substitutions per site (black branch labels), bootstrap values shown for each node (red branch labels). Gene families are colour-coded by relationship to arabidopsis AIL proteins ("At" bold labels). Putative Kranz candidates are highlighted ("GRMZM" bold labels and arrows).

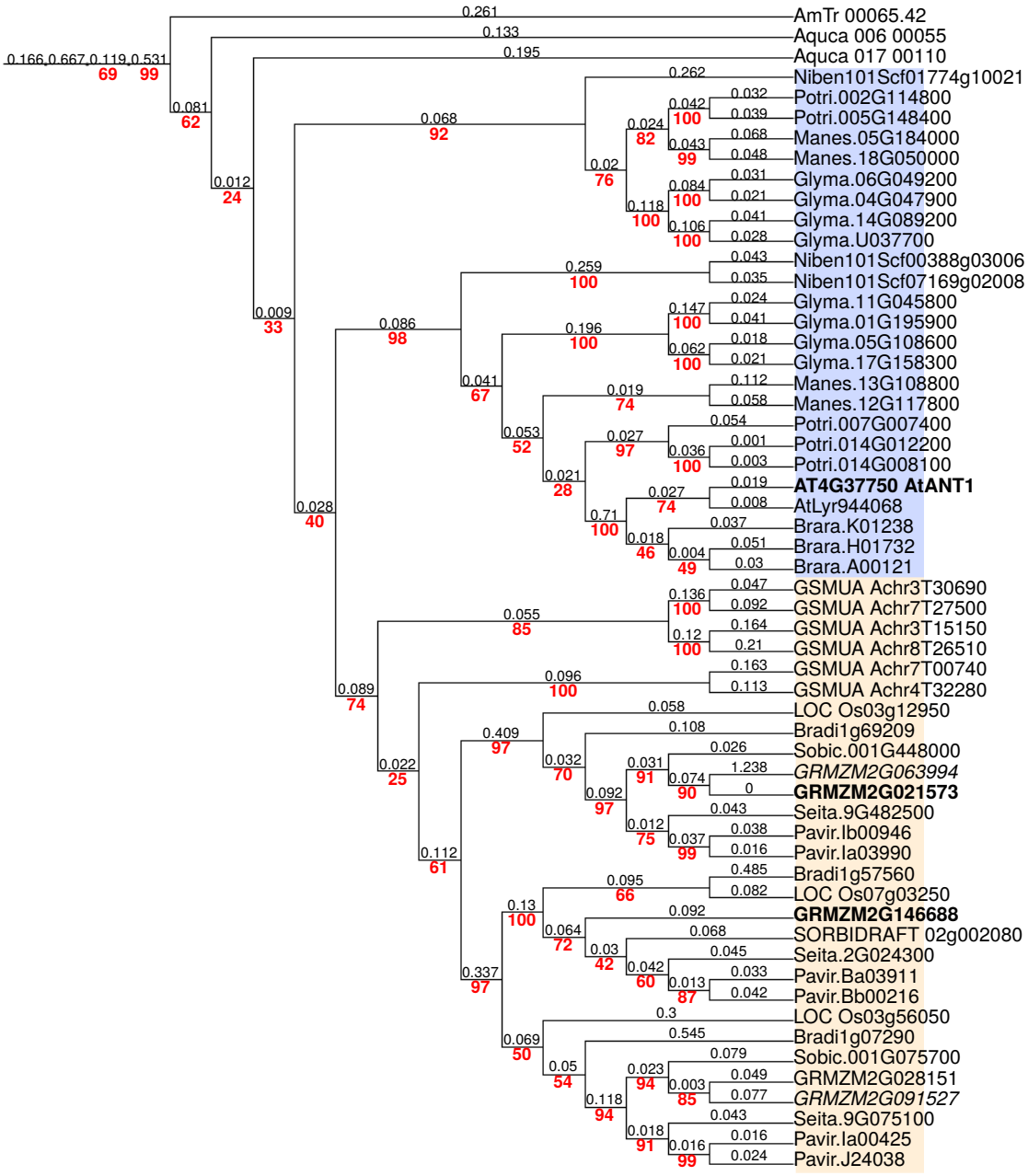


Figure 4.3: Pruned AINTEGUMENTA-like protein phylogenetic tree to highlight AINTEGUMENTA clade. Maximum-likelihood inference of phylogentic relationships between *ANT* genes, pruned from fig.4.2. Branch length measured in number of amino acid substitutions per site (black branch labels), bootstrap values shown for each node (red branch labels). Arabidopsis ("AT" bold label), putative Kranz regulators ("GRMZM" bold labels) and putative non-coding RNA (italic labels) are highlighted. Blue- eudicots, cream- monocots.

Moore & Purugganan, 2005), and may have been crucial for the evolution of Kranz anatomy (Monson, 2003). From the *AIL* clades that feature putative Kranz regulators, the *ANT* clade in particular shows gene expansion. Two monocot specific gene duplications are inferred (fig.4.3; p.118), both before the PACMAD/BEP clade divergence with one more recent than the other. These duplication events mean that the resulting genes (*ZmANT1* and *ZmCRL5*) could have undergone sub- or neo-functionalisation, altering protein activity during leaf development and enabling C_4 evolution. *ZmANT1* and *ZmCRL5* were therefore taken forward for further analysis.

4.2.2 Verification and correction of monocot *ANT* gene models using RNAseq data and RT-PCR

Protein alignments enable analysis and annotation of functional domains in related proteins, however, this approach requires accurate amino acid sequences. *In silico* analysis of genomic sequences can be used to predict gene models and therefore amino acid sequences (Ouyang *et al.*, 2007; Kersey *et al.*, 2016), however, because this process is automated, genes can be mis-annotated. Transcriptomics data for maize (Wang *et al.*, 2013b) were used to verify the accuracy of the *ZmANT* gene models available in the Phytozome database (Goodstein *et al.*, 2012). Three transcripts were predicted for *ZmANT1*, one full-length (transcript one - T1, consisting of eight relatively short exons, including a 9 bp exon (Klucher *et al.*, 1996)) and two truncated (transcript two- T2 and transcript three- T3). The annotation of the 5′ untranslated region (UTR) of T2 was altered between different versions of genome annotation (fig.4.4, A; p.120), however, there was no evidence for this transcript in the RNAseq data (fig.4.4, B). There was evidence for an un-annotated transcript variant (transcript 4 - T4) in the RNAseq data, which was confirmed by RT-PCR analysis (fig.4.5, A & B; p.121), however, only T1

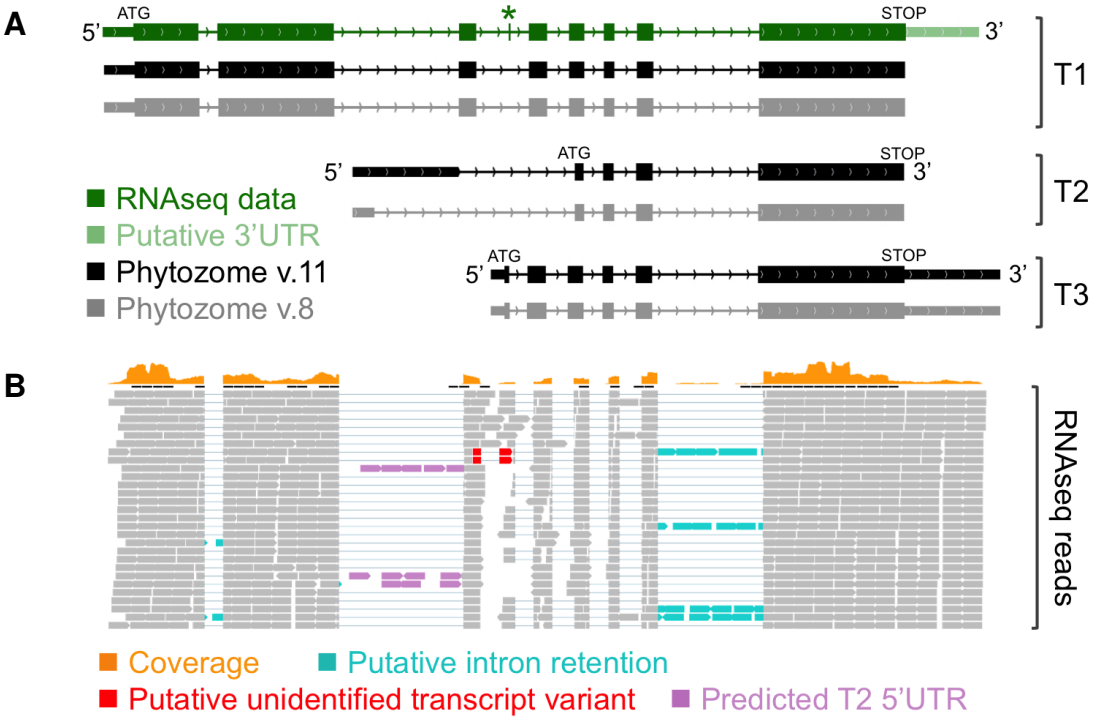


Figure 4.4: Mis-annotation of *ZmANT1* gene models in Phytozome.

A) *ZmANT1* gene models for transcript- 1 (T1), 2 (T2) and 3 (T3), taken from two versions of Phytozome- v.11 (black) and v.8 (gray), compared to full length transcript from RNAseq data (green). Translation start ("ATG") and stop ("STOP") sites are indicated. Thin horizontal line- intron, medium horizontal line- untranslated region (UTR), thick horizontal line- exon. Green asterisk- 9 bp exon (missing from all predicted gene models). **B)** Wang *et al.* (2013b) RNAseq transcript reads from maize plastochrons 3 & 4 aligned to genome models in (A), RNA reads mapped to genome using TopHat (personal communication, S. Kelly; Kim *et al.*, 2013), coverage of genome by reads (orange), RNA reads (gray), putative intron retention (blue), predicted 5′ UTR for transcript 2 (pink), putative unidentified transcript (red).

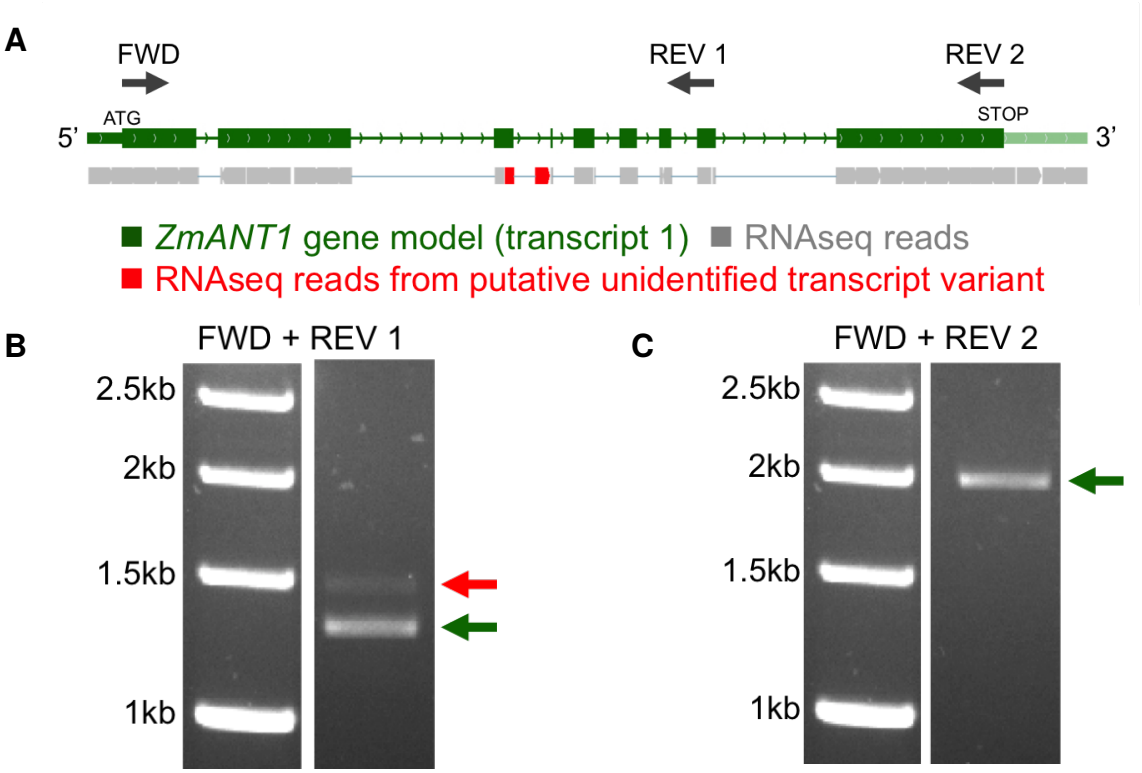


Figure 4.5: Confirmation of *ZmANT1* unidentified transcript variant using RT-PCR. A) Primers "FWD" and "REV2" amplify the full length *ZmANT1* transcript 1 (T1) cDNA (green). Primers "FWD" and "REV1" amplify a section encompassing an expressed region unique to a putative unidentified transcript (red). Green medium horizontal line- untranslated region (UTR), thick horizontal line- exon. Green and gray thin horizontal line- intron. Gray thick horizontal line- RNAseq reads. **B-C)** Agarose gels showing RT-PCR using maize foliar primordia cDNA and primer pairs- "FWD" and "REV1" (B) and "FWD" and "REV2" (C). Predicted T1 fragment (green arrow), transcript variant (red arrow).

was identified with a full length RT-PCR amplification, so T4 is truncated in the 3′ region (fig.4.5, C). RNAseq data, RT-PCR amplification and sequencing showed that the T1 gene model was missing a 9 bp exon and had a mis-annotated splice site that introduced an extra serine residue into the predicted amino acid sequence. Overall, verification of *ZmANT1* gene models using transcriptomic data was a successful approach to correct errors in the gene model, and identified several mis-annotations in the full-length transcript which were corrected before protein sequence alignment.

Transcriptomics data have also been produced for rice primordia (van Campen *et al.*, 2016) and so a similar verification approach was taken for *LOC Os03g12950* (*OsANT1*). The rice gene model predicted a truncated gene product (Ouyang *et al.*, 2007) due to mis-annotated splice sites. RNAseq data was used to annotate the gene model correctly (fig.4.6; p.123). Only one transcript was predicted, however, the high level of putative intron retention, especially between exons 3 and 4 and exons 7 and 8 may indicate the presence of truncated transcripts similar to those predicted in the *ZmANT1* locus (fig.4.6 & fig.4.4). The correct maize and rice *ANT* gene model annotations were used to verify the other monocot gene models in the *ANT* clade. *GRMZM2G146688*, *GRMZM2G3990772 Bradi1g69209*, *Bradi1g07290*, *LOC Os07g03250* (*OsCRL5*) and *LOC Os03g56050* were all found to have mis-annotated splice sites; these were corrected using the *ZmANT1* and *OsANT1* models prior to protein sequence alignment.

4.2.3 Functional domain differences are seen between C_3 and C_4 monocot ANT1 protein sequences

To investigate how protein sequences have diverged in the monocot ANT clade, protein sequence alignments using three C_3 and three C_4 monocot species were generated using MergeAlign (Collingridge & Kelly, 2012) and *ZmANT1* was aligned

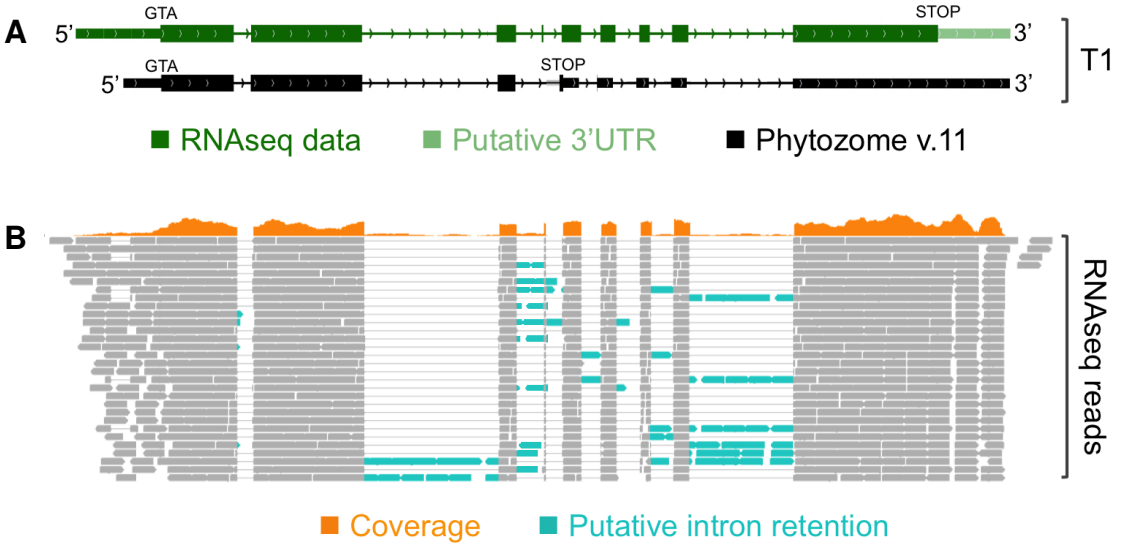


Figure 4.6: Mis-annotation of *OsANT1* gene model in Phytozome.

A) *OsANT1* gene model (one transcript, T1) taken from Phytozome v.11 (black), compared to full length transcript from RNAseq data (green). Translation start ("GTA") and stop ("STOP") sites are indicated. Thin horizontal line- intron, medium horizontal line- untranslated region (UTR), thick horizontal line- exon. **B)** van Campen *et al.* (2016) RNAseq transcript reads from rice plastochron 3 aligned to genome model in (A), RNA reads mapped to genome using TopHat (S. Kelly, personal communication; Kim *et al.*, 2013, coverage of genome by reads (orange), RNA reads (gray), putative intron retention (blue).

against *AtANT* to give putative functional domain annotation (fig.4.7, p.125). DNA binding domain residues were highly conserved between arabidopsis and maize and there was a substantial amount of residue conservation in the N' terminal half of the protein (including the transcriptional activation domain, upstream of the DNA binding domains). However, beyond the DNA binding domains there was a high degree of amino acid sequence divergence in the C' terminal half of the proteins, and the nuclear localisation site was not conserved in ZmANT1. Since the monocot-eudicot divergence, there thus appears to have been strong selective pressure in arabidopsis and maize to maintain both AP2 DNA binding domains and the N' terminal sequence that includes the transcriptional activation site. These observations suggest that ANT protein tertiary structure is highly constrained around these domains and that protein function is essential to plant fitness.

To test for C_3 versus C_4 differences in monocot ANT1 protein sequence, amino acid alignments for rice, *Brachypodium distachyon*, *Hordeum vulgare* (all C_3), maize, *Sorghum bicolor* and *Setaria italica* (all C_4) were analysed (fig.4.8, p.126). All the monocot AP2 DNA binding domains showed a high degree of sequence similarity, as did the N' terminal halves of the proteins. However, there were amino acid residues that were consistently different between the C_3 and C_4 species both in the putative transcriptional activation domain and in the DNA binding domains. In the transcriptional activation domain, C_4 species were missing one hydrophobic residue in the middle of the domain, and had an additional methionine residue towards the end of the region. The proteins also contained a DYRQ motif consisting of four polar residues that was not present in C_3 sequences. Significantly, there were also consistent amino acid differences in the DNA binding domains; a region that appears to be under high purifying selection. The differences between C_3 and C_4 ANT1 protein sequence are consistent with the hypothesis that ZmANT1 is a Kranz regulator, particularly given that the species examined represented two independent origins of C_4 (Grass Phylogeny Working Group II, 2012). Residue

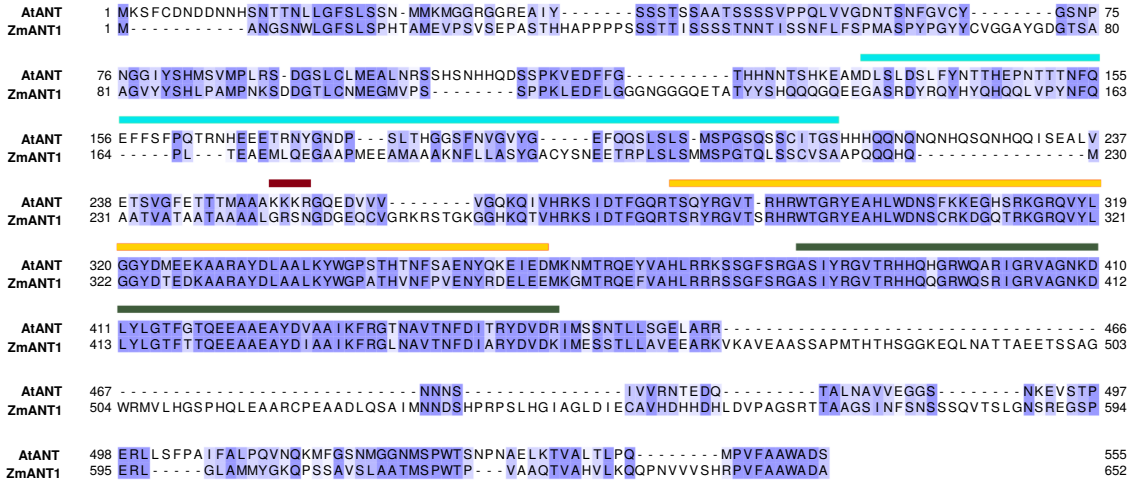


Figure 4.7: Functional domain annotation of ZmANT1 using AtANT1 amino acid sequence alignment. ANT1 sequence alignments of *Zea mays* (Zm) and *Arabidopsis thaliana* (At) proteins. N' terminal amino acid (a.a.)- $AtANT_1$/$ZmANT1_1$, C' terminal a.a.- $AtANT_{555}$/$ZmANT1_{652}$. AtANT functional domains annotated by horizontal lines- transcriptional activation (blue), nuclear localisation (red), AP2-R1 DNA binding (orange), AP2-R2 DNA binding (green). Degree of a.a. residue conservation shown by colour intensity proportional to Blocks Substitution Matrix (BLOSUM62) values.



Figure 4.8: ANT1 amino acid sequence differences between C_3 and C_4 monocots are present in functional domains. ANT1 protein sequences alignment from monocot species; three C_4- *Zea mays* (Zm), *Sorghum bicolor* (Sb) and *Setaria italica* (Si) (cream background) and three C_3- *Oryza sativa* (Os), *Brachypodium distachyon* (Bd) and *Hordeum vulgare* (Hv) (white background). N' terminal amino acid (a.a.)- ZmANT1$_1$, C' terminal a.a.- ZmANT1$_{652}$. Functional domains annotated by horizontal lines- putative transcriptional activation (blue), AP2-R1 DNA binding (orange), AP2-R2 DNA binding (green). Conserved differences in putative functional domains between C_3 and C_4 species (red arrows). Degree of a.a. residue conservation shown by colour intensity proportional to Blocks Substitution Matrix (BLOSUM62) values.

differences in the DNA binding domain may attenuate or change the ability of ANT1 to target gene loci in C_4 species, and differences in the putative transcriptional activation domain may change the expression levels of target genes, enabling Kranz anatomy development.

Protein alignments were also created for the other putative Kranz regulator, ZmCRL5 (fig.4.9, p.128). Unlike with the ANT1 sequences, there were no consistent C_3 - C_4 functional domain residue differences in CRL5 monocot orthologs. CRL5 sequences did display some conserved differences near to the putative transcriptional activation domain, including an alanine rich hydrophobic motif which appears to be specific to CRL5 proteins in C_4 species (it was not seen in any of the ANT1 sequences). This observation suggests that there could be C_4 specific changes in expression levels of CRL5 target genes. However, as there were no consistent C_3 - C_4 differences in the functional domains of CRL5 proteins, ZmANT1 appears to be a better candidate for Kranz regulation and for functional analysis in a C_3 developmental context.

4.2.4 *ZmANT1* transcripts encode both a full-length and a truncated protein with possible regulatory functions

Transcription factor activity is regulated on several levels; transcriptional, translational and post-translational controls have all been documented (Gonzalez, 2015). The regulation of ZmANT1 during leaf development is likely to occur at all these levels. Transcriptomics data and RT-PCR revealed the presence of a transcript variant (transcript 3; fig.4.4; p.120) that encodes a truncated ZmANT1 protein (ΔZmANT1). The protein lacks the transcriptional activation domain, but contains almost all the amino acid residues shown to be necessary for DNA binding in arabidopsis (fig. 4.10; p.129). As such, this protein may act as a dominant negative regulator of ZmANT1 function by competitively binding with target gene

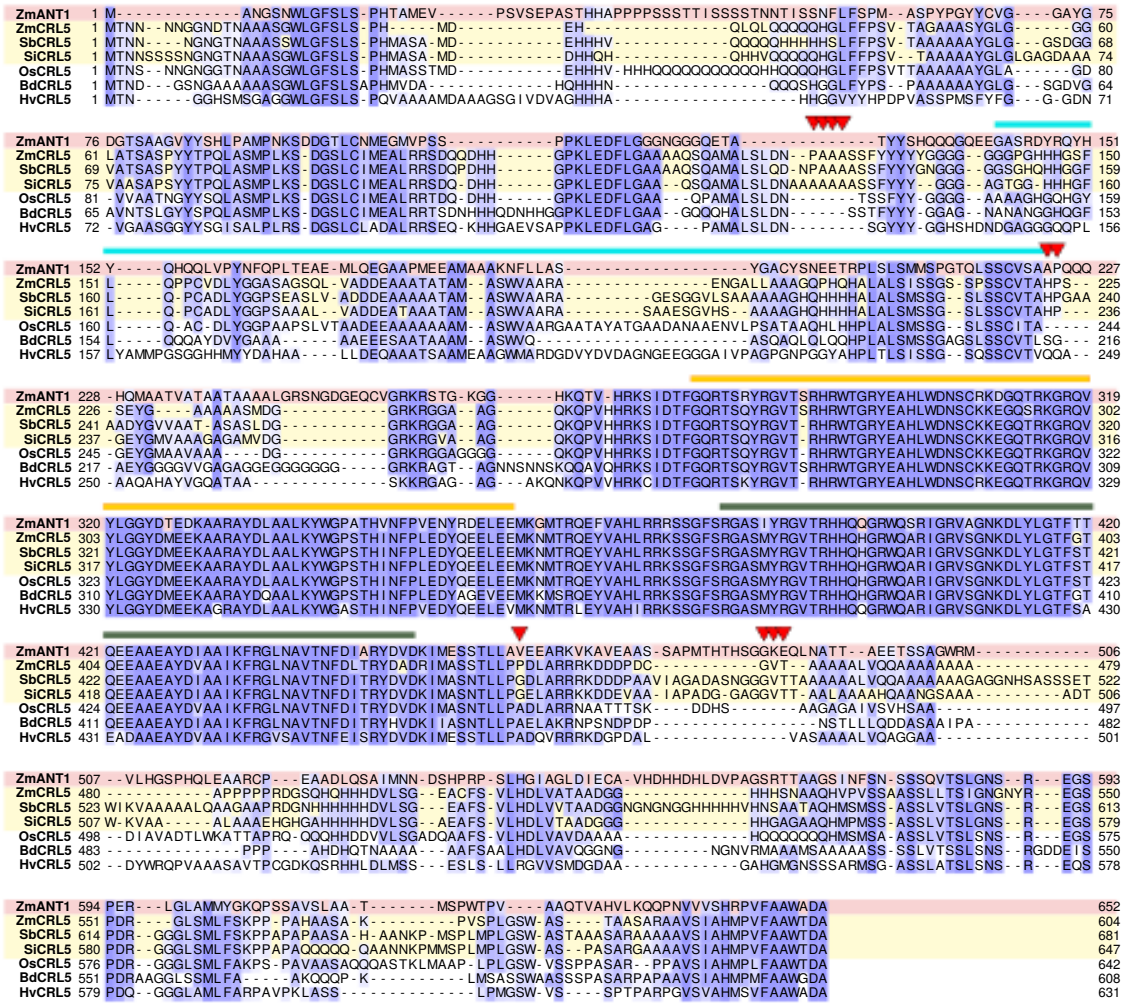


Figure 4.9: CRL5 amino acid sequence differences between C_3 and C_4 monocots are outside the annotated functional domains. CRL5 protein sequences alignment from monocot species; three C_4- *Zea mays* (Zm), *Sorghum bicolor* (Sb) and *Setaria italica* (Si) (cream background) and three C_3- *Oryza sativa* (Os), *Brachypodium distachyon* (Bd) and *Hordeum vulgare* (Hv) (white background), ZmANT1 protein sequence used as reference (pink background). N' terminal amino acid (a.a.)- ZmANT1$_1$, C' terminal a.a.- ZmANT1$_{652}$. Functional domains annotated by horizontal lines- putative transcriptional activation (blue), AP2-R1 DNA binding (orange), AP2-R2 DNA binding (green). Conserved differences between C_3 and C_4 species (red arrows). Degree of a.a. residue conservation shown by colour intensity proportional to Blocks Substitution Matrix (BLOSUM62) values.



Figure 4.10: *ZmANT1* transcript 3 encodes a truncated protein variant (△ZmANT1) that lacks a transcriptional activation domain. ZmANT1 full length (T1) and truncated (T3) protein sequence alignments. Functional domains annotated by horizontal lines- putative transcriptional activation (blue), AP2-R1 DNA binding (orange), AP2-R2 DNA binding (green). Amino acid residues- necessary for DNA binding (black arrows) and that attenuate DNA binding (grey arrows) in arabidopsis (Krizek, 2003).

promoter sequences but failing to activate transcription. If ΔZmANT1 is a negative regulator, ectopic expression in a C_3 developmental context would be predicted to generate an ANT1 knock-down phenotype.

4.2.5 Developmental consequences of *ZmANT1* overexpression in rice

4.2.5.1 Transformation of *ZmANT1* into rice led to low rates of regeneration and a high frequency of gene silencing

To test the activity of ZmANT1 in a C_3 developmental context, a *$35S_{pro}$::HYG-$ZmUBI_{pro}$::ZmANT1* construct was transformed into rice callus using agrobacterium-mediated gene transfer. T_0 plants were regenerated on cytokinin-rich media but regeneration rates were poor and only three transgenic T_0 lines set seed (table 4.2; p.131). T1 seed was grown from lines 3, 4 & 5, and tested for transgene insertion using PCR (fig.4.11, p.132). The lines did not show a significant deviation from a 3:1 segregation of the transgene, tested using a chi-square test of independence (*$ZmUBI_{pro}$::ZmANT1-3*: $\chi^2(1, N = 14) = 1.17$, $p>0.05$; *$ZmUBI_{pro}$::ZmANT1-4*: $\chi^2(1, N = 14) = 1.17$, $p>0.05$; *$ZmUBI_{pro}$::ZmANT1-5*: $\chi^2(1, N = 8) = 0.13$, $p>0.05$), indicating single insertion events. Expression analysis on a subset of the individuals showed transgene silencing in all of the *$ZmUBI_{pro}$::ZmANT1-5* individuals and half of the *$ZmUBI_{pro}$::ZmANT1-3* seedlings. These data suggest that ubiquitous expression of *ZmANT1* in rice decreases plant fitness and that lines with strong transgene expression fail to regenerate.

Line	Transgene	T_0 survival	T_0 seed set
ZmANT1-1	Yes	Yes	Poor
ZmANT1-2	Yes	No	N/A
ZmANT1-3	Yes	Yes	Good
ZmANT1-4	Yes	Yes	Good
ZmANT1-5	Yes	Yes	Good

Table 4.2: *$ZmUBI_{pro}$::ZmANT1* rice transformation regeneration summary.

4.2.5.2 Transformation of △*ZmANT1* into rice led to successful regeneration and high levels of transgene expression

To test for possible dominant negative effects of the △ZmANT1 protein, a *$35S_{pro}$::HYG-$ZmUBI_{pro}$::△ZmANT1* construct was transformed into wild-type rice callus, with the same procedure as in section 4.2.5.1. Unlike with *ZmANT1*, △*ZmANT1* plants regenerated well and seed set in T_0 plants was good (table 4.3). Segregation patterns in T_1 lines 1, 4 and 5 were analysed. The lines did not show a significant deviation from a 3:1 segregation of the transgene, tested using a chi-square test of independence (*$ZmUBI_{pro}$::ZmANT1-1*: $\chi^2(1, N = 6) = 0.06$, $p > 0.05$; *$ZmUBI_{pro}$::ZmANT1-3*: $\chi^2(1, N = 10) = 0.03$, $p > 0.05$; *$ZmUBI_{pro}$::ZmANT1-4*: $\chi^2(1, N = 9) = 1.12$, $p > 0.05$), indicating single insertion events. T_2 individuals were subsequently tested for transgene insertion and expression (fig.4.12, p.133) and there was strong transgene expression with no gene silencing.

Line	Transgene	T_0 survival	T_0 seed set
△*ZmANT1-1*	Yes	Yes	Good
△*ZmANT1-2*	Yes	Yes	Good
△*ZmANT1-3*	Yes	Yes	Good
△*ZmANT1-4*	Yes	Yes	Good
△*ZmANT1-5*	Yes	Yes	Good
△*ZmANT1-6*	Yes	Yes	Good
△*ZmANT1-7*	Yes	Yes	Good

Table 4.3: *$ZmUBI_{pro}$::△ZmANT1* transformation regeneration summary.

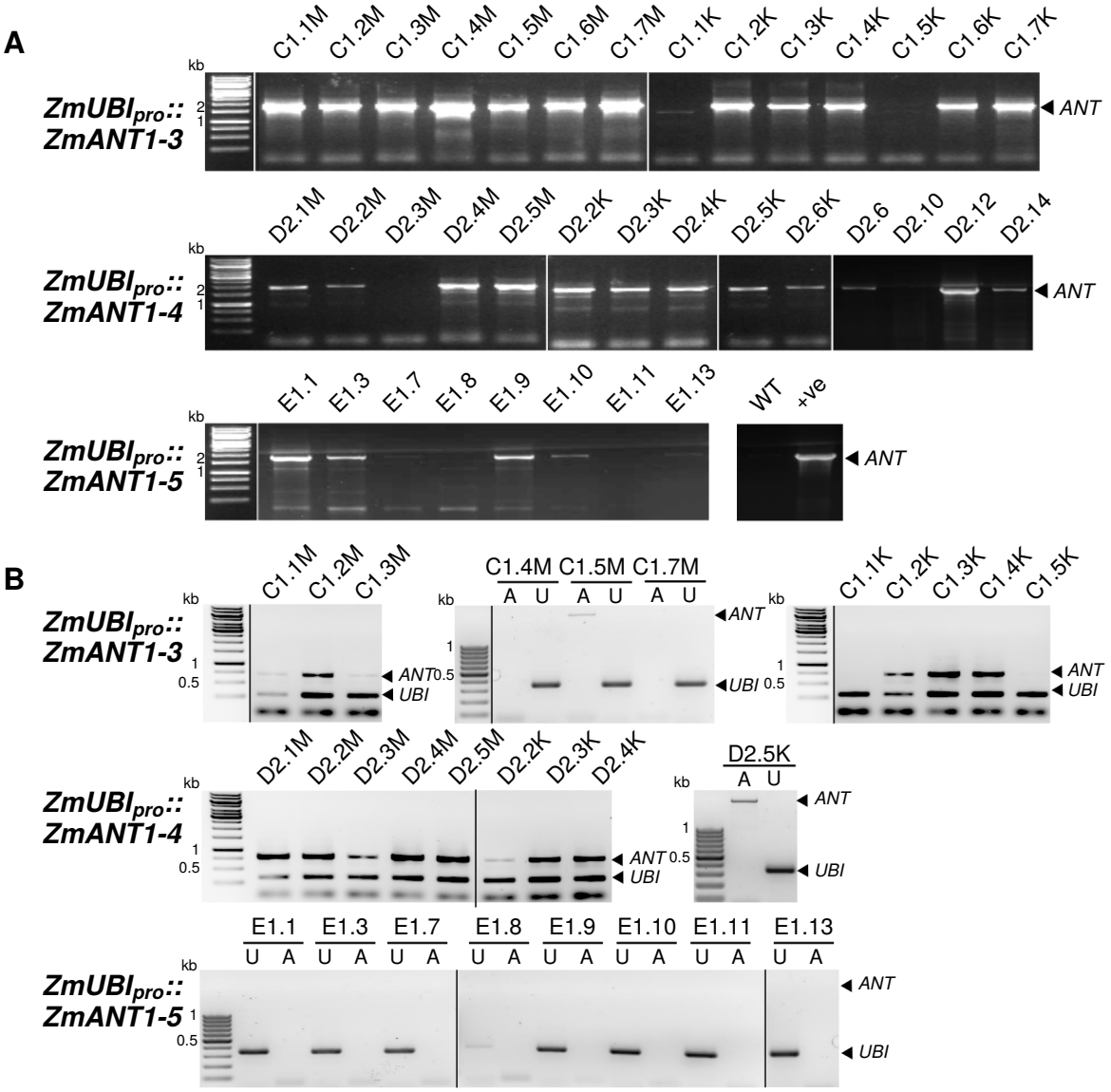


Figure 4.11: Composite of gel images showing *ZmUBI$_{pro}$::ZmANT1* genotyping and expression analysis. T_1 individuals from three *ZmUBI$_{pro}$::ZmANT1* lines were analysed for **A)** Transgene insertion using PCR with genomic DNA. Diluted plasmid was used as a positive control ("+ve") and non-transgenic Kitaake rice DNA as a negative control ("WT") . **B)** Transgene expression using RT-PCR. *OsUBI* amplification was used as a positive control, two sets of gels were run, one where both *ANT* (718 bp) and *UBI* (340 bp) PCR amplification products were loaded into one lane, and the other where *ANT* (1930 bp) and *UBI* (340 bp) were loaded in separate wells ("A" and "U" respectively). Each PCR well ID is a unique identifier for one individual which corresponds with the RT-PCR ID. ID code refers to independent transformation events (C1/D2/E1), T_1 individuals within each family (.#), and kinetin or MS growth condition (K/M).

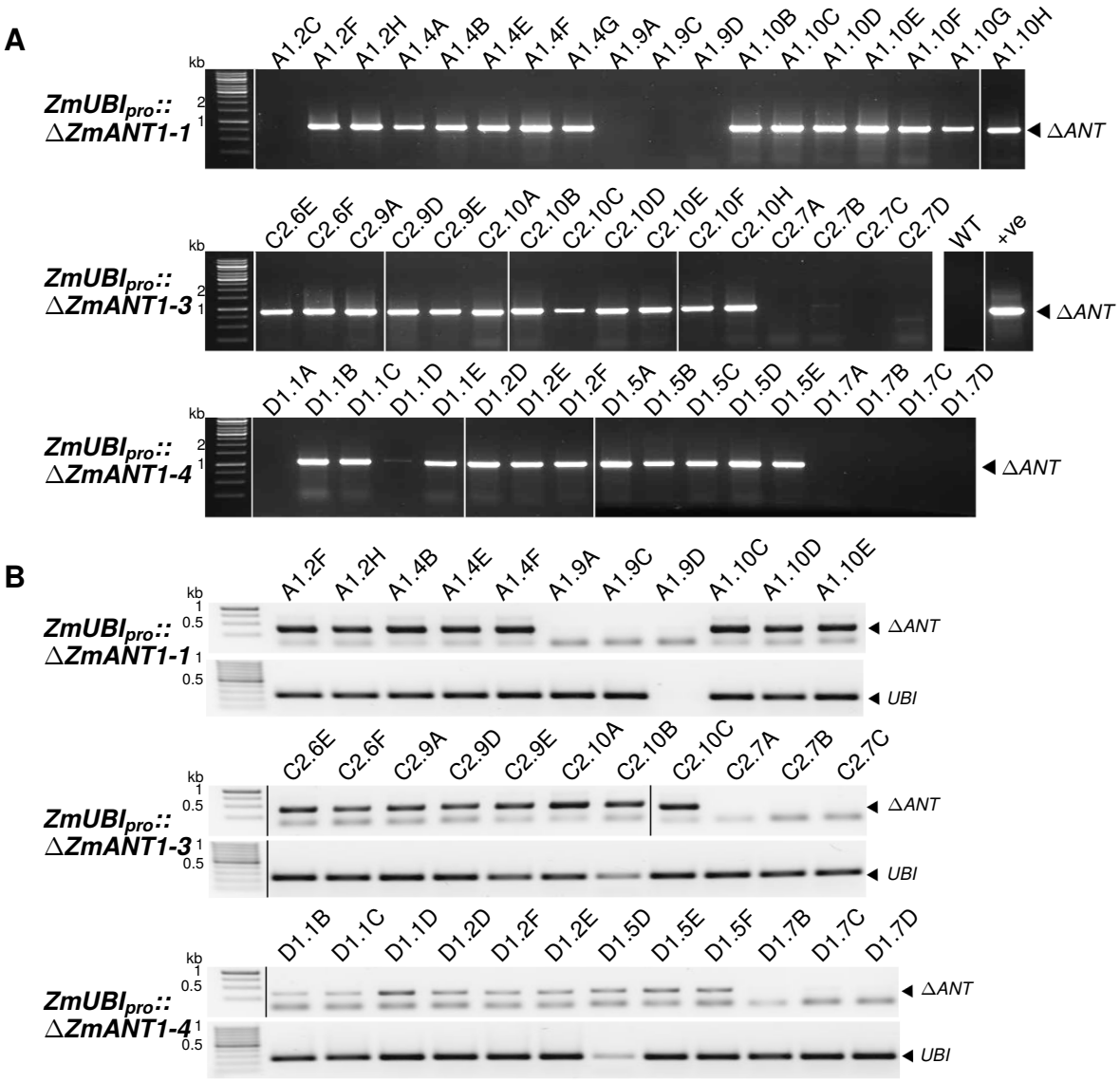


Figure 4.12: Composite of gel images showing *ZmUBI$_{pro}$::ΔZmANT1* genotyping and expression analysis. T_2 individuals from three *ZmUBI$_{pro}$::ΔZmANT1* lines were analysed for **A)** Transgene insertion using PCR with genomic DNA. Diluted plasmid was used as a positive control ("+ve") and non-transgenic Kitaake rice DNA as a negative control ("WT"). **B)** Transgene expression using RT-PCR. *OsUBI* amplification was used as a positive control. Each PCR well ID is a unique identifier for one individual which corresponds with the RT-PCR ID. ID code refers to independent transformation events (A1/C2/D1), the T_1 family (.#) and the T_2 individual in the family (last letter).

4.2.6 ZmANT1 mediated suppression of leaf blade expansion in rice is dependent on the N' terminal activation domain

4.2.6.1 ZmANT1 decreases leaf width without any effect on vein initiation when the N' terminal transcriptional activation domain is present

To determine whether ubiquitous expression of *ZmANT1* in rice conditions Kranz-like traits, and whether ubiquitous expression of $\triangle$*ZmANT1* elicits a knock-down ANT1 phenotype, cross-sections of transgenic rice leaves were examined for vein number and leaf width. Leaf vein density was compared to wild-type plants (fig.4.13; p.136). There were no statistically significant differences between transgenic and wild-type individuals, however, there was a trend for increased vein density in the *ZmANT1* plant lines that was not seen in the $\triangle$*ZmANT1* individuals (fig.4.13, A & D). This difference was due to a trend for narrower leaves, without a concomitant change in the number of veins (fig.4.13, B & C), indicating that *ZmANT1* overexpression does not influence vein initiation. As these trends were not seen with $\triangle$*ZmANT1*, the N' terminal domain appears to be necessary for the inhibitory trend of ZmANT1 on rice leaf blade expansion.

Interveinal distance, mesophyll cell size and mesophyll cell number were measured to further investigate the cellular basis of the leaf narrowing trend, with measurements being made in the middle of the leaf lamina (between the second and third lateral veins, counting out from the mid-rib) (fig.4.14; p.138). Consistent with the leaf width trend, there was also a trend for decreased interveinal distance in *ZmANT1* transgenic lines (fig.4.14, A). This decrease appears to be driven by a slight reduction in mesophyll cell length and a trend for lower mesophyll cell numbers (fig.4.14, B & C). These trends were not seen in $\triangle$*ZmANT1* lines, again

indicating that the presence of the ZmANT1 N' terminal domain is necessary for the phenotypic trends. Given the small sample sizes, the trends seen were not statistically significant, however, if further experimentation with increased sample sizes gave significant differences, these data may suggest that ZmANT1 suppresses rice leaf blade expansion by suppressing mesophyll cell division and expansion through transcriptional regulation of downstream targets.

4.2.6.2 ZmANT1 increases bundle sheath cell width and vein size

To investigate whether *ZmANT1* expression in rice leaves had an effect on vascular bundle development, a number of vascular parameters were measured in the *ZmUBI$_{pro}$::ZmANT1* transgenic lines. These included BS cell size, lateral vein height, lateral and intermediate vein area and vascular cell number (fig.4.15; p.4.15). Significant increases in BS width were seen in both transgenic lines compared to WT (fig.4.15, A), and intermediate vein area and lateral vein height were significantly larger than WT in the *ZmUBI$_{pro}$::ZmANT1-3* line (*ZmUBI$_{pro}$::ZmANT1-4* showed a similar, non-significant, trend) (fig.4.15, B & D). The lateral vein measurements also showed a trend for increased area (fig.4.15, E). These differences were not seen in $\triangle$*ZmANT1* lines (fig.4.16; p.4.16), indicating that the N' terminal domain is also necessary for these effects on rice leaf development. Together, these data indicate that ZmANT1 may have a contrasting effect on mesophyll and vascular cell development; affecting downstream gene targets to constrain mesophyll cell size and number whilst having a positive effect on BS and vascular cell growth.

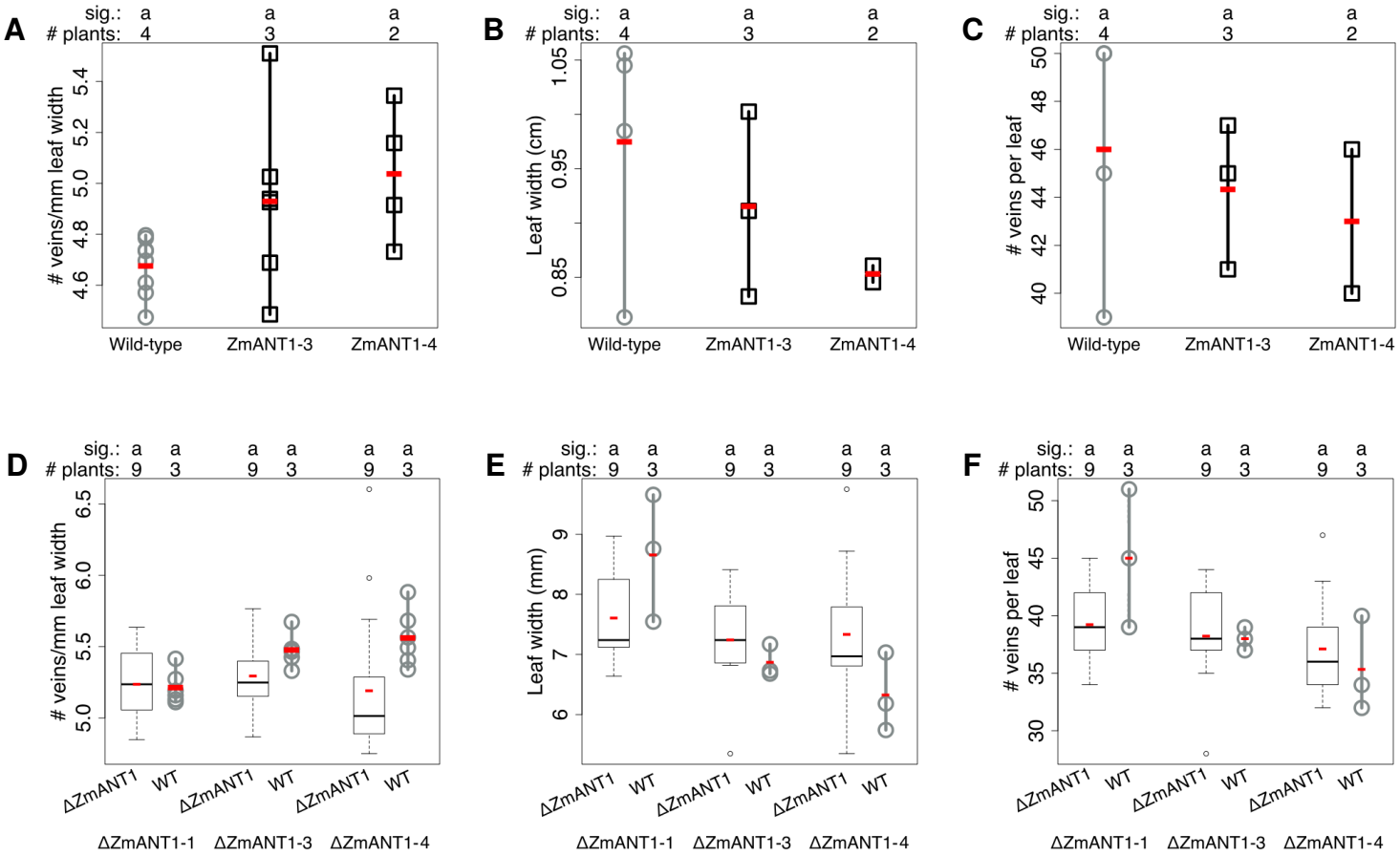


Figure 4.13: Increased vein density in *ZmUBI$_{pro}$::ZmANT1* plants is due to narrower leaves and is dependent on the N' terminal domain. Figure legend is on p.137.

Figure 4.13: Increased vein density in *ZmUBI$_{pro}$::ZmANT1* plants is due to narrower leaves and is dependent on the N' terminal domain. A-C) Strip charts showing - vein density (veins/mm leaf width) (A), leaf width (cm) (B) and number of veins per leaf (C) in wild-type and *ZmUBI$_{pro}$::ZmANT1* siblings. **D-F)** Boxplots showing- vein density (veins/mm leaf width) (D), leaf width (mm) (E) and number of veins per leaf (F) in wild-type and *ZmUBI$_{pro}$::ΔZmANT1* siblings. "Sig."- different letters indicate significant differences in group means ($p<0.05$). # plants- number of individual plants used for analysis. **Strip charts**: Circle- individual data point, vertical line- range of data, red dash- mean. **Boxplots**: Box- data within 25th-75th percentiles, with black horizontal line- median, red dash- mean, whiskers- lowest and highest value still within 1.5 IQR, open circle- value outside 1.5 IQR.

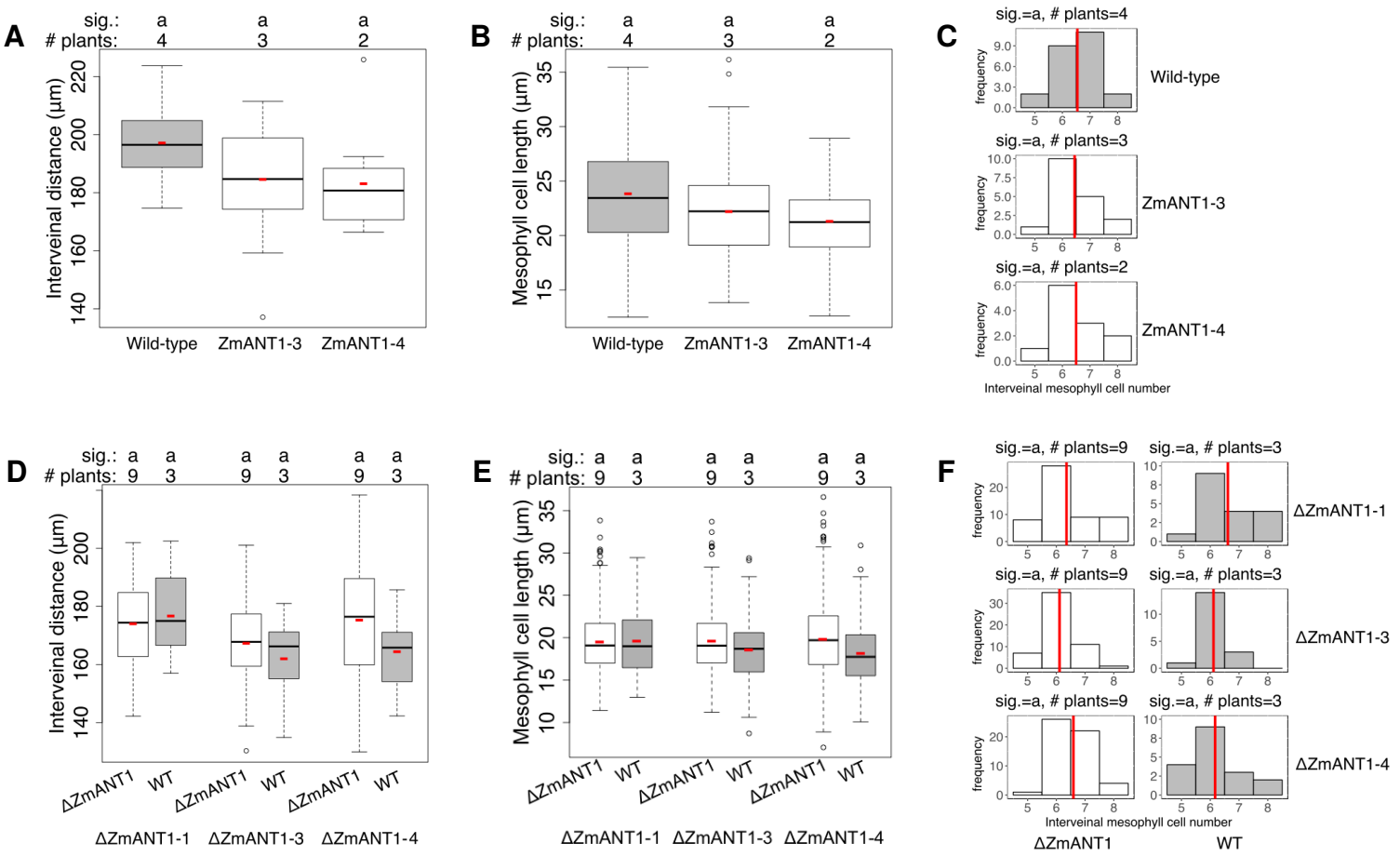


Figure 4.14: *ZmUBI$_{pro}$::ZmANT1* plants show a N' terminal dependent decrease in interveinal distance due to fewer mesophyll cells. Figure legend is on p.139.

Figure 4.14: *ZmUBI*$_{pro}$*::ZmANT1* plants show a N' terminal dependent decrease in interveinal distance due to fewer mesophyll cells. A-C) Boxplots and histograms showing - interveinal distance (µm) (A), mesophyll cell length (µm) (B) and mesophyll cell number between adjacent intermediate veins (C) in wild-type and *ZmUBI*$_{pro}$*::ZmANT1* siblings. **D-F)** Boxplots and histograms showing- interveinal distance (µm) (E), mesophyll cell length (µm) (F) and mesophyll cell number between adjacent intermediate veins (G) in wild-type and *ZmUBI*$_{pro}$*::ΔZmANT1* siblings. "Sig."- different letters indicate significant differences in group means ($p<0.05$). # plants- number of individual plants used for analysis. **Boxplots**: Box- data within 25^{th}-75^{th} percentiles, with black horizontal line- median, red dash- mean, whiskers- lowest and highest value still within 1.5 IQR, open circle- value outside 1.5 IQR. **Histograms**: Red vertical line- mean.

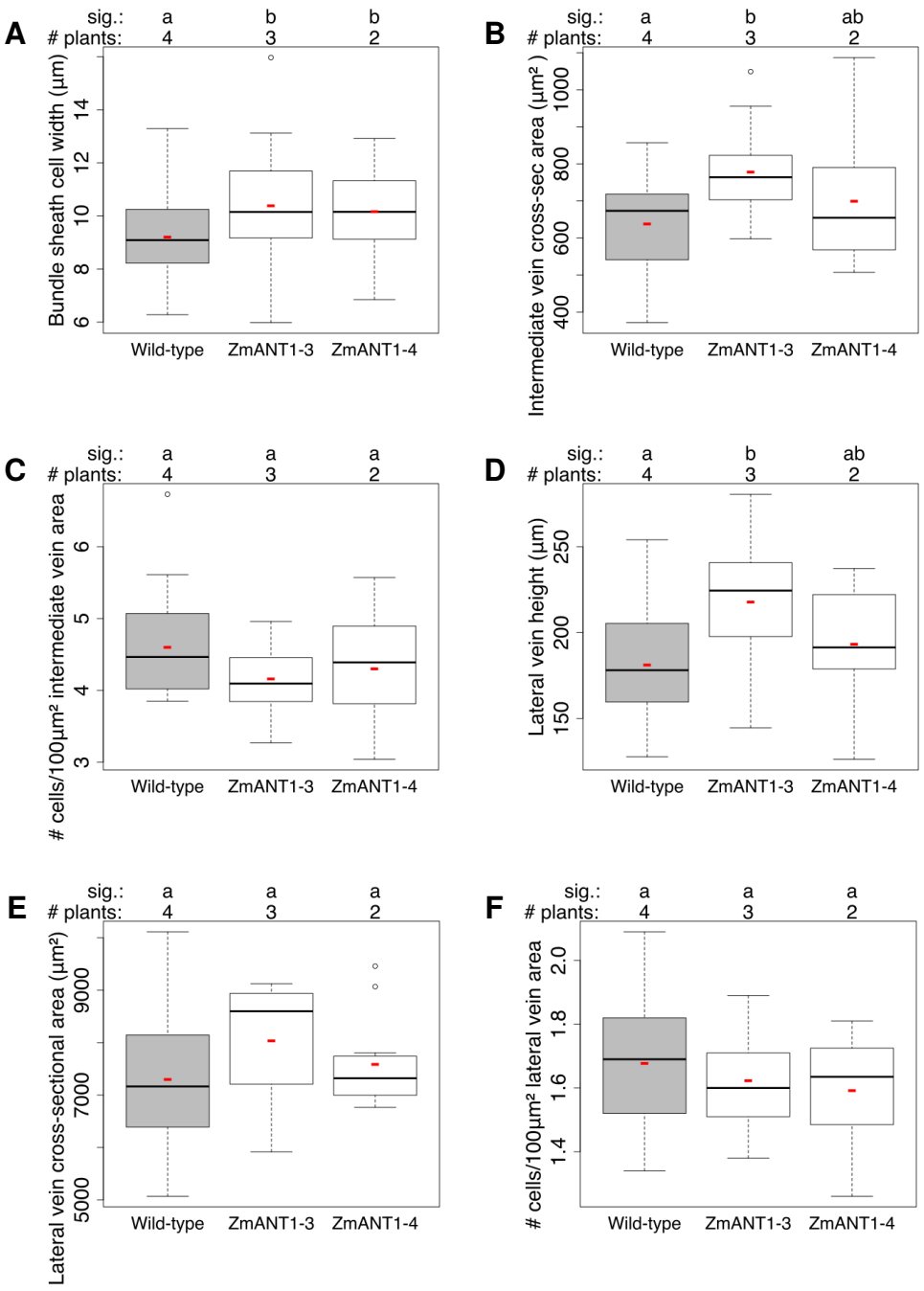


Figure 4.15: *ZmUBI$_{pro}$::ZmANT1* plants show increased bundle sheath width and vein size. Figure legend is on p.141.

Figure 4.15: *ZmUBI$_{pro}$::ZmANT1* plants show increased bundle sheath width and vein size. Leaf measurements taken from wild-type and *ZmUBI$_{pro}$::ZmANT1* lines. **A-C)** Boxplots showing measurements around intermediate veins- bundle sheath cell width (µm) (A), vascular bundle cross-sectional area (µm^2) (B), number of vascular cells per 100µm^2 bundle area (C). **D-F)** Boxplots showing measurements around lateral veins- height (µm) (D), vascular bundle cross-sectional area (µm^2) (E), number of vascular cells per 100µm^2 bundle area (F). "Sig."- different letters indicate significant differences in group means ($p<0.05$). # plants- number of individual plants used for analysis. **Boxplots**: Box- data within 25th-75th percentiles, with black horizontal line- median, red dash- mean, whiskers- lowest and highest value still within 1.5 IQR , open circle- value outside 1.5 IQR.

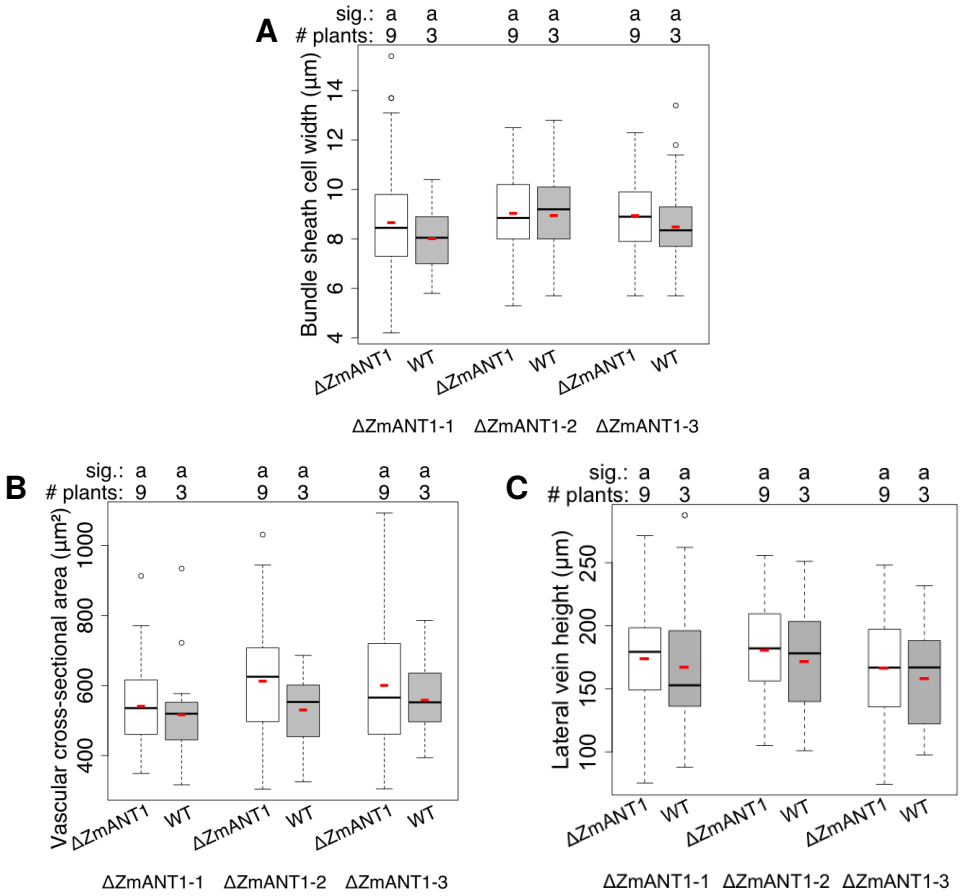


Figure 4.16: *ZmUBI$_{pro}$::ΔZmANT1* plants do not display bundle sheath cell or vein size differences. A-B) Boxplots showing measurements around intermediate veins- bundle sheath cell width (µm) (A), vascular bundle cross-sectional area (µm^2) (B) of wild-type and *ZmUBI$_{pro}$::ΔZmANT1* siblings. **C)** Lateral vein height (µm) of wild-type and *ZmUBI$_{pro}$::ΔZmANT1* siblings. "Sig."- different letters indicate significant differences in group means ($p<0.05$). # plants- number of individual plants used for analysis. **Boxplots**: Box- data within 25th-75th percentiles, with black horizontal line- median, red dash- mean, whiskers- lowest and highest value still within 1.5 IQR, open circle- value outside 1.5 IQR.

4.2.7 AINTEGUMENTA-like target gene roles include cell growth regulation and RNA modification

Identifying putative ZmANT1 target genes in maize leaf development could begin to provide a molecular explanation for the effects of ZmANT1 in rice and could give insights into the role of ZmANT1 during Kranz development. To identify ANT target genes, an *in silico* search for the ANT DNA binding consensus sequence was conducted in 5kb DNA regions upstream of maize open reading frames (ORFs) using PatMatch software (Yan *et al.*, 2005). To identify genes relevant to Kranz development, the putative target ORFs were then filtered using the maize leaf transcriptomics data from Wang *et al.* (2013b) to give genes that are expressed at >5 transcripts per million in plastochron 3 and 4, either during husk (HP3&4) or foliar (FP3&4) development. To indicate whether these genes could have a positive or negative regulatory role in Kranz development, the filtered target gene list was divided into genes that were up-regulated or down-regulated during Kranz development (FP3&4) compared with the husk expression levels, and gene ontology (GO) terms for all the genes were analysed using PANTHER (Protein Analysis Through Evolutionary Relationships) software (Mi *et al.*, 2016). See appendix B (p.233) for gene lists.

Interestingly, many AIL targets genes encode for enzymes involved in metabolism, including purine synthesis. This potentially implicates AIL protein activity in the regulation of M cell development. Several transcription factor genes were also identified as AIL protein targets including a *CVL* gene, two *SCARECROW-like* (*SCL*) GRAS family genes *AC234164.1-FGT004*, *GRMZM2G079470* (both upregulated in Kranz development) and six putative Kranz regulators (table 4.4, p.144) (Wang *et al.*, 2013b). These results give support to the hypothesis that AIL proteins are important regulators at the early stages of Kranz development.

Gene ID	Kranz regulator	FP 1	FP 3&4	FP 5	HP 1	HP 3&4	HP 5	BS	M	Protein type & role
GRMZM2G053338	Positive	1698	1123	246	200	174	88	0	0	**IAA-amido synthetase**: Auxin homeostasis
GRMZM2G112795	Positive	3656	2964	2624	1667	862	1206	65	117	**Unknown**
GRMZM2G171852	Positive	175	439	507	122	224	334	15	0	**DOF protein**: Transcription factor
GRMZM2G150408	Positive	548	763	990	290	218	359	3	1	**Unknown**
GRMZM2G147420	Negative	65	132	226	86	175	307	1	0	**Unknown**
GRMZM2G016480	Negative	1091	1776	2020	1830	2534	2974	2	0	**Receptor kinase**: Serine/threonine kinase

Table 4.4: *AINTEGUMENTA-like* protein target genes encode putative Kranz regulators. AIL protein target genes identified as either putative positive or negative Kranz regulators. mRNA expression in relative transcript abundance in- plastochrons: foliar (FP) 1, 3&4 and 5 and husk (HP) 1, 3&4 and 5 (Wang *et al.*, 2013b), and bundle sheath (BS) and mesophyll (M) cells (Li *et al.*, 2010).

Biological process GO terms provide insight into the general biological roles of the identified AIL targets. The analysis shows that proteins encoded by putative AIL target genes are involved in a large variety of biological processes (fig.4.17, A & B; p.146). AIL targets code for proteins involved in the cell cycle, cell growth and cellular differentiation. Comparison of the AIL targets that are up-regulated or down-regulated during Kranz development shows that almost all of the target genes implicated with a role in developmental processes and cell growth are up-regulated during Kranz development (fig.4.17, C & D). These data indicate that AIL proteins may control cell growth and proliferation during Kranz development.

Molecular function GO terms complement the biological process annotations by providing information about the specific activity of identified proteins. The majority of identified target gene products have catalytic activity (fig.4.18 A & B; p.148), however, there are also many nucleic acid binding proteins. It is interesting to note that the serine-threonine kinase receptors in the list are all up-regulated (rather than down-regulated) during Kranz development (fig.4.18 C & D). These genes all encode pentatricopeptide repeat-containing proteins, a family of proteins that is involved in post-transcriptional RNA-modifying processes (Schmitz-Linneweber & Small, 2008), suggesting that AIL proteins may activate post-transcriptional regulatory networks during Kranz development.

Cytokinin and auxin signalling are vital for leaf development. To investigate whether AIL proteins are involved in hormone signalling networks the target gene lists were searched for GO terms relating to these hormones. Six auxin and six cytokinin related genes were identified (table.4.5; p.150), including two putative Kranz regulator genes identified by Wang *et al.* (2013b). One of these genes is *GRMZM2G053338* (*ZmGH3.8*; Feng *et al.*, 2015) which belongs to a protein family that has been shown to decrease free-auxin levels by conjugating IAA with amino acids. Interestingly, the majority of the hormone related AIL targets are downregulated during Kranz development. This could indicate that AIL proteins act directly

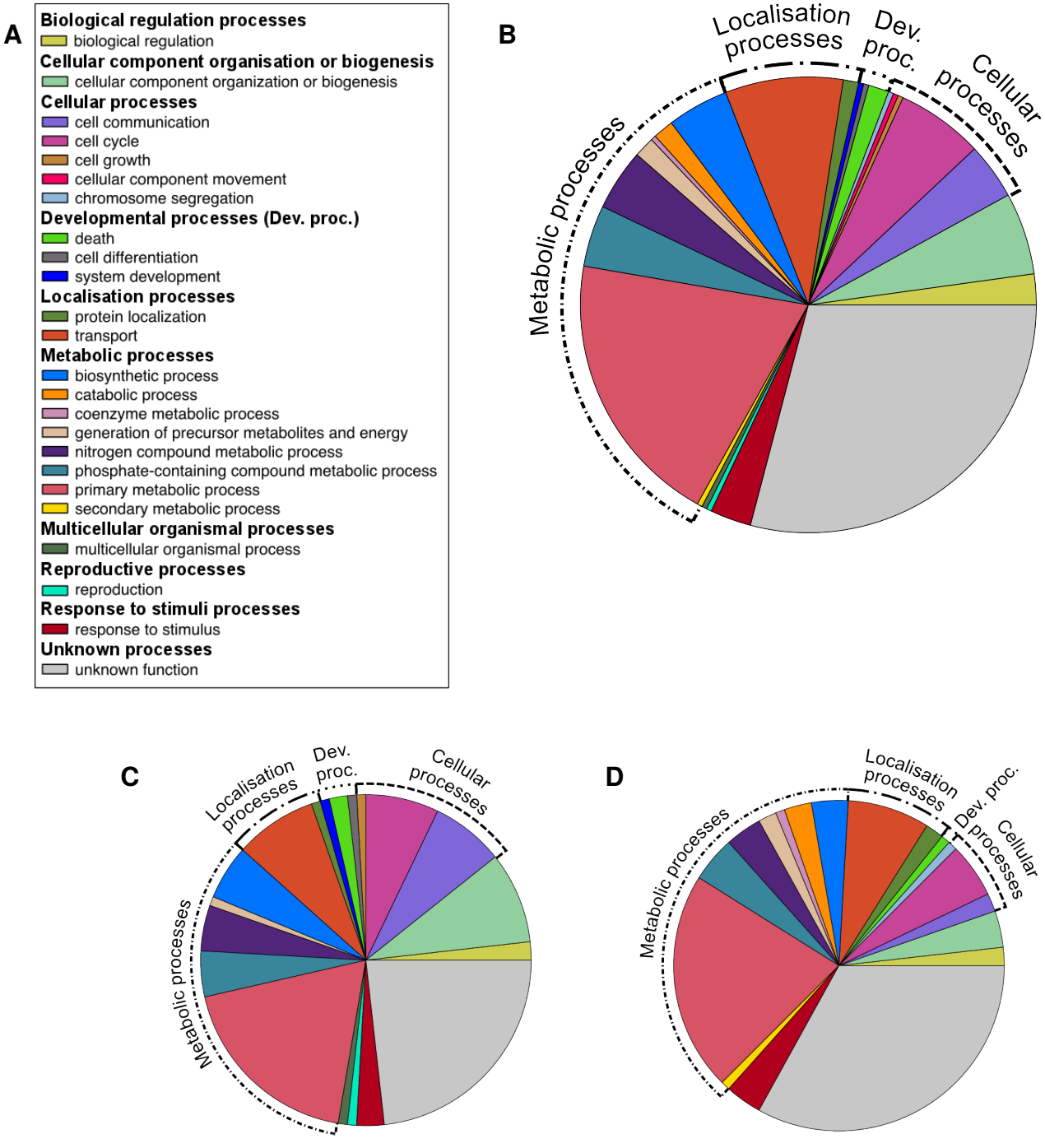


Figure 4.17: Biological process gene ontology of putative ANT target genes expressed during maize leaf development. Figure legend on p.147

Figure 4.17: Biological process gene ontology of putative ANT target genes expressed during maize leaf development. A) Categories of gene ontology (GO) annotations related to biological processes present in the putative ANT target gene list. **B-D)** Pie charts showing biological process gene ontology of putative ANT target genes- all that are expressed during maize leaf development (262 genes) (B), subset that is up-regulated during Kranz development (90 genes) (C) and subset that is down-regulated during Kranz development (135 genes) (D). The majority of the genes are involved in metabolism, including biosynthesis and phosphate metabolism.

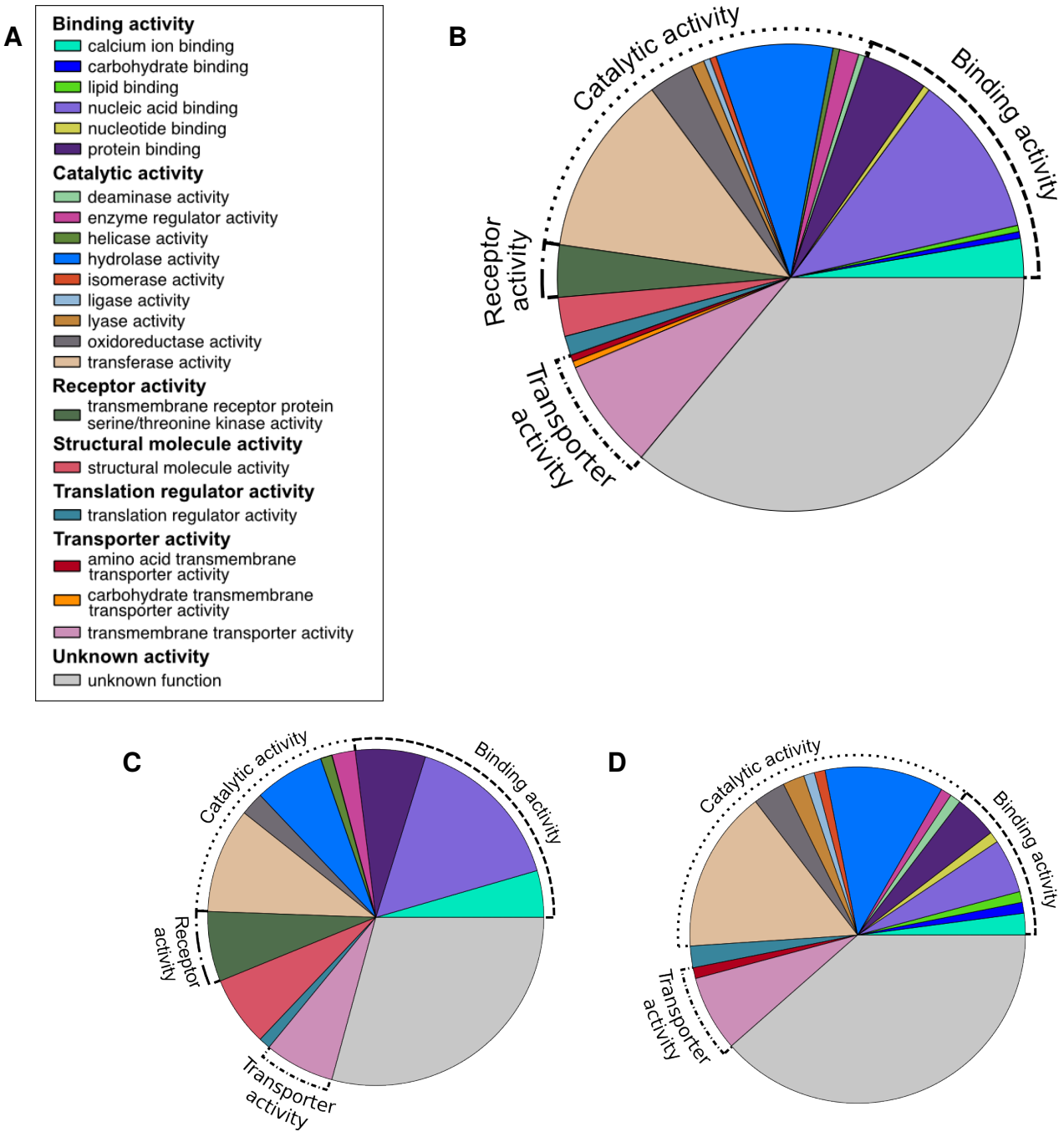


Figure 4.18: Molecular function gene ontology of putative ANT target genes expressed during maize leaf development. Figure legend on p.149

Figure 4.18: Molecular function gene ontology of putative ANT target genes expressed during maize leaf development. A) Categories of gene ontology (GO) annotations related to molecular function of translated gene products present in the putative ANT target gene list. **B-D)** Pie charts showing molecular function gene ontology of putative ANT target genes- all that are expressed during maize leaf development (262 genes) (B), subset that is up-regulated during Kranz development (90 genes) (C) and subset that is down-regulated during Kranz development (135 genes) (D).

	Gene ID	F & H expr.	FP 1	FP 3&4	FP 5	HP 1	HP 3&4	HP 5	BS	M	Protein type & role
Auxin	GRMZM2G053338*	F>H	1698	1123	246	200	174	88	0	0	**IAA-amido synthetase**: Auxin homeostasis
related	GRMZM2G030710	H>F	1400	1266	981	2057	1895	1361	5	3	**ARF24**: KANADI protein interactor, abaxial/adaxial patterning.
	GRMZM2G175812	H>F	0	1	1	43	130	18	0	0	**Chalcone synthase**: Metabolism.
	GRMZM2G167018	H>F	13	8	9	55	39	25	75	26	**NAC protein**: Transcription factor, morphogenesis.
	GRMZM2G156977	H>F	0	0	0	2	13	3	0	0	**NAC protein**: Transcription factor, morphogenesis.
	GRMZM2G070563	H>F	8	8	6	4	19	11	23	4	**Auxin efflux carrier**: Auxin transport.
	AC233850.1_FGT005	F≈H	109	96	84	81	90	87	1	0	**E2F**: Cell cycle control.
Cytokinin	GRMZM2G112795*	F>H	3656	2964	2624	1667	862	1206	65	117	**Unknown**
related	GRMZM2G005036	F>H	1567	1319	1393	1402	1101	845	33	21	**Mitochondrial processing peptidase**: Mitochondrial respiratory chain.
	GRMZM2G111216	H>F	0	2	3	0	3	16	48	20	**Chloroplast stem-loop binding protein**: Chloroplast rRNA synthesis.
	GRMZM2G009344	H>F	1	6	4	139	32	14	58	4	**Purine permease**: Cytokinin transport.
	GRMZM2G167018	H>F	13	8	9	55	39	25	75	26	**NAC protein**:Transcription factor, morphogenesis.
	GRMZM2G156977	H>F	0	0	0	2	13	3	0	0	**NAC protein**: Transcription factor, morphogenesis.

Table 4.5: *AINTEGUMENTA-like* protein target genes expressed during maize leaf development that encode auxin and cytokinin responsive proteins Gene expression during Kranz development- upregulated (F>H) or downregulated (H>F), relative transcript abundance for- plastochrons: foliar (FP) 1, 3&4 and 5 and husk (HP) 1, 3&4 and 5 (Wang *et al.*, 2013b), and bundle sheath (BS) and mesophyll (M) cells (Li *et al.*, 2010). Putative Kranz regulator genes (*) (Wang *et al.*, 2013b).

to repress auxin and cytokinin related gene expression during Kranz development. Alternatively, it is possible that the C_4 specific amino-acid differences in ZmANT1 act to decrease binding affinity, reducing the strength of transcriptional activation thereby reducing gene expression and facilitating Kranz development.

4.2.8 ZmANT1 is antagonistic to cytokinin signalling in rice leaf development

4.2.8.1 ZmANT1 represses kinetin-mediated leaf blade expansion in rice

To investigate the role of ZmANT1 in auxin and cytokinin regulated developmental pathways, a preliminary experiment was conducted to investigate the effect of ZmANT1 on kinetin-regulated leaf development in rice. T1 individuals from the *$35S_{pro}$::HYG-ZmUBI$_{pro}$::ZmANT1* rice lines (see section 4.2.5.1) and a wild-type line were germinated on kinetin-rich media and mature leaves were phenotyped for leaf traits.

Vein density, leaf width and vein number were measured in mature rice leaves. In wild-type rice, kinetin treatment led to a significant decrease in vein density and an increase in interveinal distance compared with ½MS treatment (fig.4.19, A & D; p.152). However, this effect was not seen in *ZmANT1* lines; when these lines were treated with kinetin, the vein density and interveinal distance was similar to that of non-kinetin treated wild-type leaves (fig.4.19, C & F). Overall, these data suggest that cytokinin signalling promotes leaf blade expansion, and that constitutive ZmANT1 expression may inhibit this signalling pathway.

To investigate the cellular cause of the observed trends in leaf blade expansion, M cell number and size, and BS width measurements were made in the middle of the leaf lamina (between the second and third lateral veins, counting out from the mid-rib) (fig.4.20; p.155). In wild-type rice, there was a non-significant trend for increased M cell number when plants were treated with kinetin (fig.4.20, A),

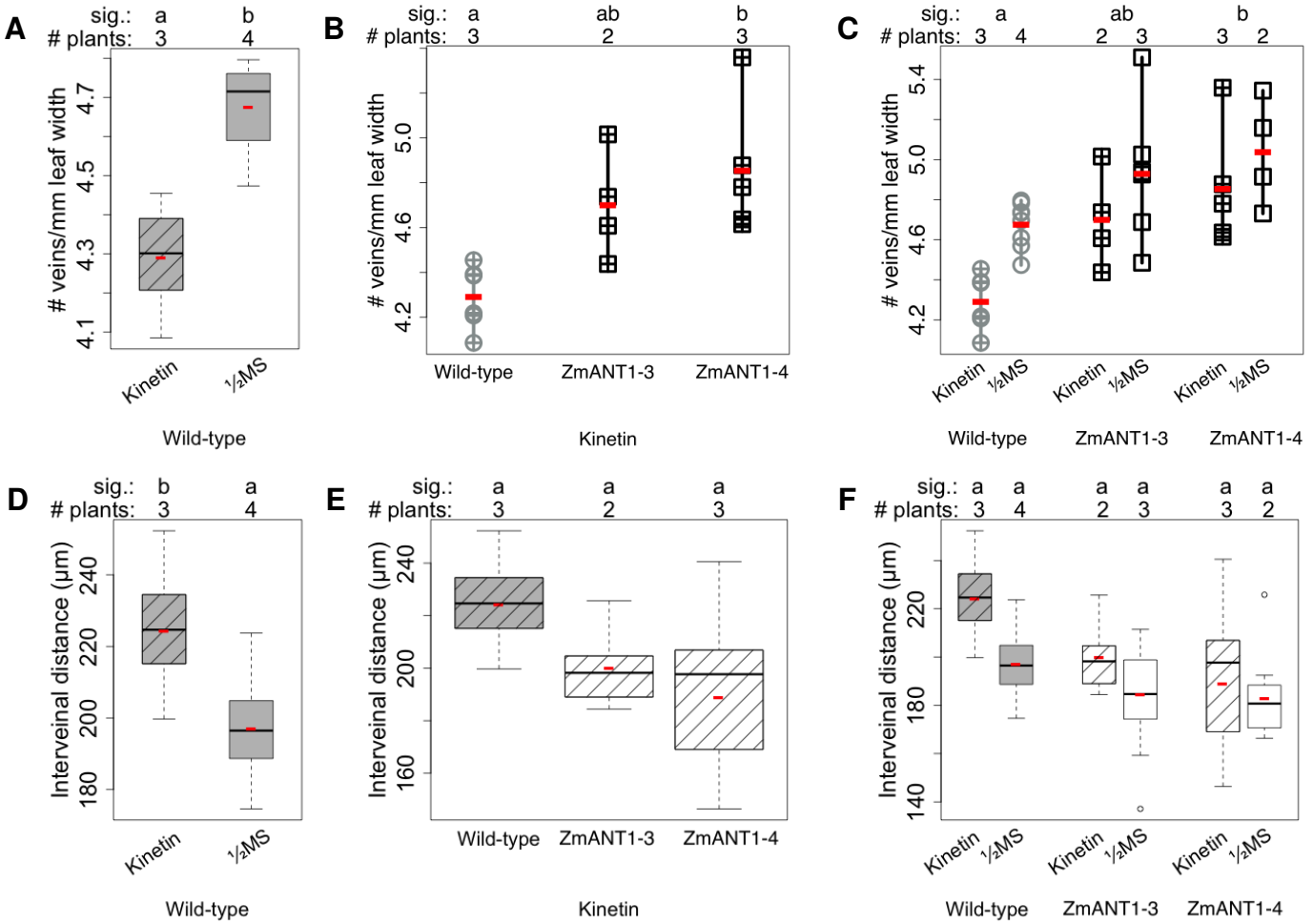


Figure 4.19: ZmANT1 inhibits kinetin effects on vein density and interveinal distance in rice. Figure legend is on p.153.

Figure 4.19: ZmANT1 inhibits kinetin effects on vein density and interveinal distance in rice. A) Boxplots and **B-C)** strip charts showing effect of kinetin on wild-type leaf vein density (veins/mm leaf width) compared with- ½MS control (A), kinetin treated *ZmUBI$_{pro}$::ZmANT1* lines (B) and ½MS control in wild-type and *ZmUBI$_{pro}$::ZmANT1* lines (C). Note wild-type and kinetin data sets are the same across A, B and C. **D)** Boxplot and **E-F)** strip charts showing effect of kinetin on wild-type leaf interveinal distance (µm) compared with- ½MS control (D), kinetin treated wild-type and *ZmUBI$_{pro}$::ZmANT1* lines (E) and ½MS control in wild-type and and *ZmUBI$_{pro}$::ZmANT1* lines (F). Note wild-type and kinetin data sets are the same across D, E and F. **Boxplots**: Box- data within 25th-75th percentiles, with black horizontal line- median, red dash- mean, whiskers- lowest and highest value still within 1.5 IQR, open circle- value outside 1.5 IQR. **Strip charts**: Square/circle- individual data point, vertical line- range of data, red dash- mean. Note 1. All ½MS data has been shown in section 4.2.6.1. Note 2. The results of post-hoc significance testing vary between graphs because statistical tests appropriate to the data were used. No statistical difference between wild-type lines was observed in C & F because two-way nested ANOVAs were used for the statistical analysis, the statistical power of this method is low compared to the one-way ANOVAs used for data in graphs A and D.

but kinetin did not have an effect on M cell size (fig.4.20, B). These data imply that the wider leaf blades observed in kinetin treated wild-type rice may be caused by increased M cell proliferation. Notably, in kinetin treated *ZmANT1* lines, no differences in M cell number or width were observed (fig.4.20, A-B), further suggesting that ZmANT1 may inhibit the effects of kinetin-signalling during rice leaf development.

A non-significant trend for larger BS cells was observed in kinetin treated wild-type plants (fig.4.20, C) and *ZmANT1* overexpression also increased BS size (see section 4.2.6.2). However, when *ZmANT1* lines were treated with kinetin, the effects on BS cell size were not observed, these results indicate that there may be a mutual inhibition between kinetin-signalling and ZmANT1, therefore suggesting that the two pathways could be antagonistic.

4.2.8.2 ZmANT1 and kinetin are antagonistic in rice vein development

ZmANT1 was shown to have a statistically significant effect on vein development in at least one transgenic rice line (see section 4.2.6.2), so to test whether kinetin is antagonistic to ZmANT1 in this stage of leaf development, vascular traits in mature leaves were also measured. Lateral vein height was found to be significantly larger in non-hormone treated *ZmANT1* lines (see section 4.2.6.2), however, kinetin treatment inhibited this increase (fig.4.21, A; p.157). *ZmANT1* lines also displayed a trend for increased intermediate vein area (see section 4.2.6.2). A similar trend was seen in kinetin treated wild-type rice plants. However, these effects were not seen when *ZmANT1* lines were treated with kinetin (fig.4.21, D). If these results were replicated in a larger experiment that showed significant differences in the kinetin and transgene conditions, these data would be consistent with ZmANT1 and kinetin acting as antagonists in rice leaf development. The effects of ZmANT1 and kinetin on leaf development are summarised in table 4.6; p.156.

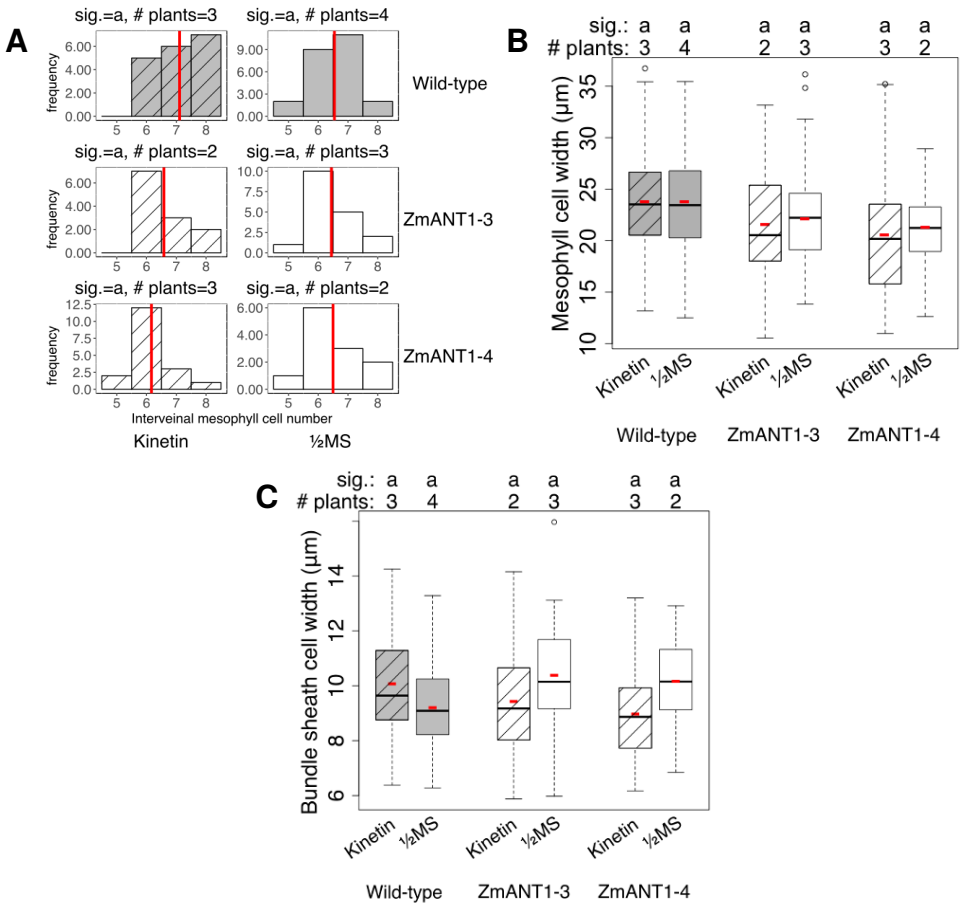


Figure 4.20: Kinetin-induced mesophyll cell proliferation and bundle sheath size increases are not seen in *ZmANT1* lines. Effect of kinetin compared with ½MS control in wild-type and two *ZmANT1* expressing rice lines. **A)** Histograms and **B)** boxplot showing mesophyll- cell number between adjacent intermediate veins (A), and cell widths (µm) (B). **C)** Boxplot showing bundle sheath cell widths around intermediate veins (µm). "Sig."- different letters indicate significant differences in group means ($p<0.05$). # plants- number of individual plants used for analysis. **Boxplots**: Box- data within 25th-75th percentiles, with black horizontal line- median, red dash- mean, whiskers- lowest and highest value still within 1.5 IQR, open circle- value outside 1.5 IQR. **Histograms**: Red vertical line- mean.

	Effect of		
Trait	**ZmANT1**	**Kinetin**	**ZmANT1 + Kinetin**
Vein density	Increase	Decrease	None
Intervein distance	Decrease	Increase	None
Mesophyll cell number	Decrease	Increase	None
Mesophyll cell width	Decrease	None	None
Bundle sheath width	Increase	Increase	None
Lateral vein height	Increase	None	None
Intermediate vein area	Increase	Increase	None

Table 4.6: Effects of ZmANT1 and kinetin on rice leaf development.

Comparison of effect relative to wild-type rice.

4.2.9 The predicted AIL protein targets in arabidopsis leaf development share little orthology with maize targets

The effect of *ZmANT1* expression in rice development is markedly different from *AtANT1* over-expression in arabidopsis, particularly in terms of leaf lamina growth and vein initiation (Krizek, 1999; Kang *et al.*, 2007). To investigate whether this might be due to differences in AIL target gene expression during leaf development, the ANT binding consensus sequence was used to identify putative arabidopsis AIL protein gene targets *in silico* using a similar approach as in section 4.2.7. Gene lists and GO term analysis can be found in appendix B (p.233). Similar classes of biological processes and molecular functions appeared to be regulated by AIL proteins in arabidopsis and maize, however only 11% of the predicted AIL target genes that were expressed during leaf development were orthologous between maize and arabidopsis. This suggests that AIL targets involved in leaf development in eudicots and monocots are not well conserved.

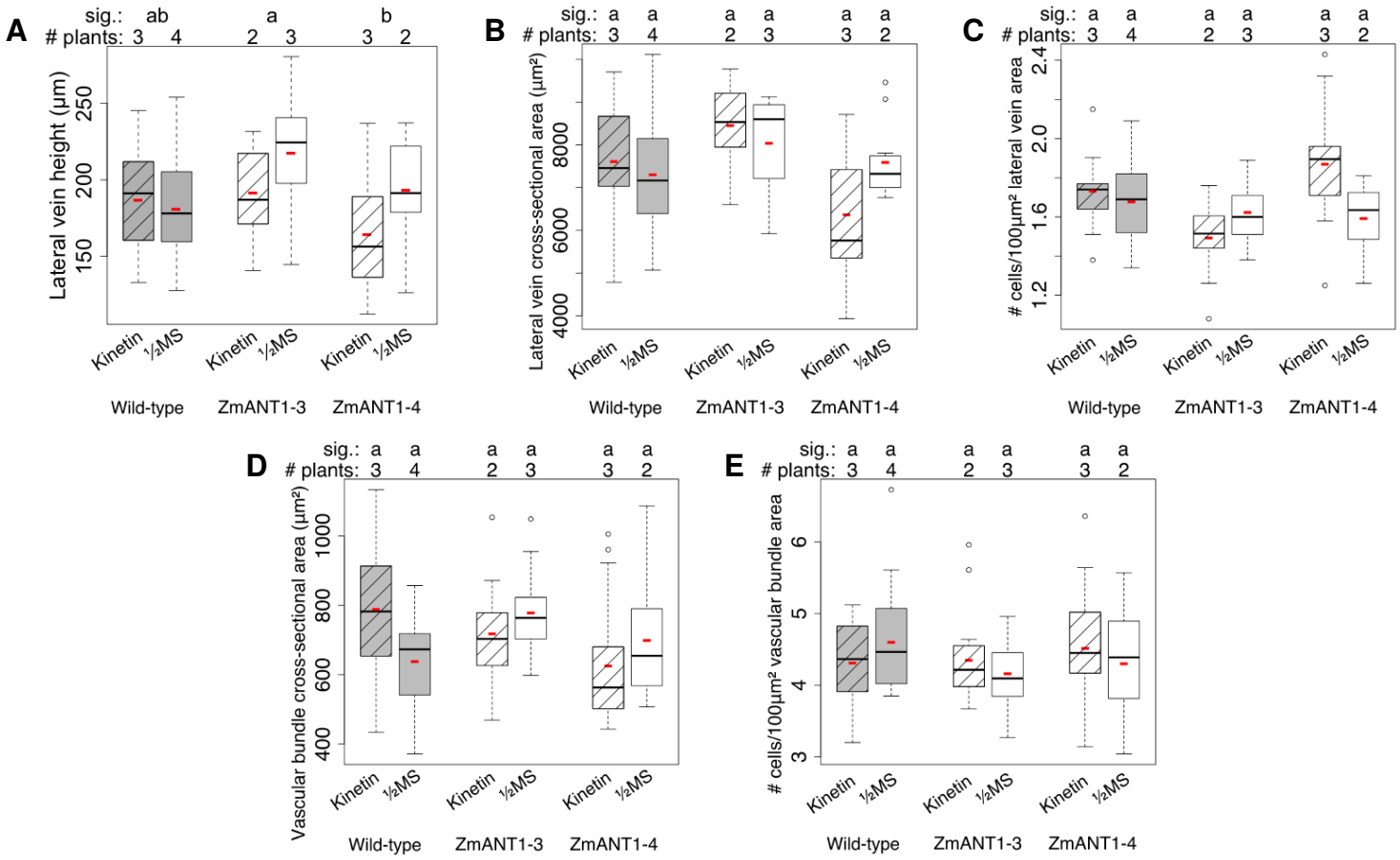


Figure 4.21: Kinetin and ZmANT1 are antagonistic in rice vein development. Boxplots showing effect of kinetin compared with ½MS control in wild-type and *ZmANT1* rice lines. **A-C)** Lateral veins - height (µm) (A), cross-sectional area (µm^2) (B), and cell number per µm^2 cross-sectional area (C). **D-E)** Intermediate veins- cross-sectional area (µm^2) (D), and cell number per µm^2 cross-sectional area (E). **Boxplots**: Box- data within 25th-75th percentiles, with black horizontal line- median, red dash- mean, whiskers- lowest and highest value still within 1.5 IQR, open circle- value outside 1.5 IQR.

To investigate whether there were likely to be differences in AIL regulation of auxin and cytokinin related genes between eudicots and monocots, predicted AIL target genes expressed during arabidopsis leaf development were analysed for auxin and cytokinin related GO terms (table 4.7; p.159). More auxin and cytokinin related targets were identified in arabidopsis than in maize (compare tables 4.5; p.150 and 4.7). There were some similarities in the target gene lists, such as genes involved in abaxial/adaxial patterning and auxin responsive factors (ARFs), however these genes were not orthologous and the roles of most of the other genes did not overlap between maize and arabidopsis. Interestingly, proteins with known roles as negative regulators of auxin (*IAA18* and *TOPLESS*) and cytokinin (*ARR7*) were identified as putative arabidopsis AIL targets.

To further investigate whether there were likely to be differences in the regulation of hormone-responsive genes between monocots and eudicots during leaf development, arabidopsis orthologs to the predicted hormone responsive AIL targets in maize (table 4.5; p.150) were analysed *in silico* for expression during leaf development and for ANT binding regions upstream of the ORFs. The genes were categorised into one of three groups: group 1 contained genes that were expressed during leaf development and had AIL target motifs in both arabidopsis and maize; group 2 contained genes that were expressed during maize and arabidopsis leaf development but where the arabidopsis ortholog was not predicted to be regulated by AIL; and group 3 contained genes that either were not expressed during arabidopsis leaf development, or did not have orthology between maize and arabidopsis (table 4.8; p.161).

	Gene ID	4d	6d	8d	Protein type & role
Auxin	*AT1G12110*	26	44	47	**NRT1**: Nitrate transporter
related	*AT5G02840*	19	31	35	**REVEILLE 4**: Circadian clock. Histone acetylation
	AT1G15750	17	22	20	**TOPLESS**: Transcriptional repression. IAA interactor. Auxin repression.
	AT2G45970	17	18	16	**Cytochrome P450**: Metabolism. Cutin biosynthesis
	AT5G42190	15	26	29	**SKP1-like**: Protein ubiquitination and degradation
	AT3G13300	14	21	19	**Enhancer of mRNA-decapping protein**: mRNA degradation. Vein development.
	AT4G05160	9	14	13	**4-coumarate-CoA ligase-like 7**: JA biosynthesis
	AT1G51950	9	11	11	**Auxin-responsive protein IAA18**: ARF negative regulator. Auxin repression.
	AT4G29810	7	13	12	**Mitogen-activated protein kinase kinase 2**: Increases stress tolerance. MAPK cascade.
	AT2G39480	7	8	7	**ABC transporter B family member 6**: Auxin transport
	AT1G30330	5	8	8	**Auxin response factor 6**: JA biosynthesis. Floral development.
	AT1G72320	5	6	5	**Pumilio homolog 23**: RNA binding. Abaxial/adaxial patterning.
Cytokinin	*AT1G79040*	817	904	1121	**Photosystem II 10 kDa polypeptide**: Photosystem II assembly.
related	*AT2G40490*	114	114	105	**Uroporphyrinogen decarboxylase 2**: Chlorophyll biosynthesis.
	AT3G07480	41	65	67	**2Fe-2S ferredoxin-like protein**: Mitochondria/chloroplast electron transport
	AT1G50480	33	52	58	**Formate-tetrahydrofolate ligase**: Metabolism
	AT5G66510	23	27	27	**Gamma carbonic anhydrase 3**: Mitichondrial respiratory chain
	AT2G46540	21	27	25	**Unknown**
	AT1G19050	17	16	13	**Two-component response regulator ARR7**: Negative regulator of cytokinin signalling.

Table 4.7: *AINTEGUMENTA-like* protein target genes expressed during arabidopsis leaf development that encode auxin and cytokinin responsive proteins. Gene expression during arabidopsis leaf development in 4 day old (4d), 6 day old (6d) and 8 day old (8d) leaves (Mi *et al.*, 2016).

No genes clustered to group 1 suggesting that AIL regulation of leaf development differs between arabidopsis and maize. Some genes did cluster to group 2, suggesting that the roles of these genes in leaf development may be conserved between arabidopsis and maize, but the regulatory networks are not. Finally, three arabidopsis genes clustered to group 3. In this case, there is no possibility that the roles of these genes are conserved in maize and arabidopsis leaf development. This analysis indicates that, although there may be some conservation in the role of cytokinin and auxin signalling in maize and arabidopsis leaf development, the interaction of AIL proteins with the hormone signalling networks is likely to be distinctly different.

Overall, the comparison of arabidopsis and maize AIL target genes indicates that the role of AIL proteins may differ substantially during monocot and eudicot leaf development. More specifically, given there was no orthology in the cytokinin or auxin responsive gene target groups, AIL protein interactions with hormone signalling networks may not be conserved between eudicots and monocots. Assuming that AIL gene targets are similar between maize and rice (given the closer phylogenetic relatioship), and given that the auxin and cytokinin hormone signalling networks are key regulators of leaf development, it may be predicted that the effects of *ZmANT1* expression on arabidopsis leaf development will differ markedly from the effect of ZmANT1 in rice. This prediction can be tested by generating transgenic *ZmANT1* arabidopsis lines. This approach would also test whether a candidate monocot Kranz regulator can confer Kranz-like traits on a eudicot C_3 species.

	Gene ID	4d	6d	8d	Shared orthogroup	F & H expr.	Protein type & role
Group 1 (Expressed during arabidopsis leaf development and has ANT consensus region)							
	No orthologs						
Group 2 (Expressed during arabidopsis leaf development but no ANT consensus region)							
Auxin	AT5G13930	43	52	45	GRMZM2G175812	F>H	**Chalcone synthase**: Metabolism.
related	AT2G33860	6	8	8	GRMZM2G030710	F>H	**AtARF3/ZmARF24**: KANADI protein interactor, abaxial/adaxial patterning.
Cytokinin	AT3G07230	79	123	92	GRMZM2G112795	F>H	**Unknown**
related	AT1G51980	49	59	59	GRMZM2G005036	F>H	**Mitochondrial processing peptidase**: Mitochondrial respiratory chain.
	AT3G63140	207	267	297	GRMZM2G111216	H>F	**Chloroplast stem-loop binding protein**: Chloroplast rRNA synthesis.
	AT1G28230	5	5	7	GRMZM2G009344	H>F	**Purine permease**: Cytokinin transport.
Group 3 (Not expressed during arabidopsis leaf development)							
Auxin	AT4G37390	0	0	0	GRMZM2G053338	F>H	**IAA-amido synthetase**: Auxin homeostasis
related	AT5G22220	0	0	0	AC233850.1-FGT005	F≈H	**E2F**: Cell cycle control.
Auxin &	AT5G62380	0	0	0	GRMZM2G167018	H>F	**NAC protein**: Transcription factor, morphogenesis.
cytokinin	AT5G62380	0	0	0	GRMZM2G156977	H>F	**NAC protein**: Transcription factor, morphogenesis.
related							

Table 4.8: AIL regulation of auxin and cytokinin responsive genes in maize leaf development is not likely to be conserved in arabidopsis leaf development. Arabidopsis orthologs ("gene ID") of maize auxin or cytokinin responsive AIL target genes ("shared orthogroup"), grouped by expression and regulation properties of arabidopsis orthologs. Gene expression during arabidopsis leaf development in 4 day old (4d), 6 day old (6d) and 8 day old (8d) leaves (Mi *et al.*, 2016). Gene expression during Kranz development- upregulated (F>H) or downregulated (H>F) (Wang *et al.*, 2013b).

4.2.10 Preliminary study shows ZmANT1 effects on leaf development in arabidopsis are opposite to those in rice

4.2.10.1 *$ZmANT1_{pro}$::ZmANT1* transformation into arabidopsis produced multiple transgenic lines and good seed set

To test whether the effect of *ZmANT1* expression on leaf development differs between arabidopsis and rice, a *$ZmANT1_{pro}$::ZmANT1* construct was transformed into arabidopsis. The *$ZmANT1_{pro}$* promoter consisted of a 2960 bp region upstream of the *ZmANT1* start codon and was used instead of a constitutive promoter to reduce the risk of sterile transgenic plants; AtANT is involved in ovule development, and *$35S_{pro}$::AtANT* lines have been shown to be sterile (Krizek, 1999; Kang *et al.*, 2007). Given the amino acid sequence similarity of ZmANT1 to AtANT, it is possible that the maize gene could perform a similar role if constitutively expressed. A *SHR/shr-6* heterozygous T-DNA insertion arabidopsis line (Yu *et al.*, 2010) was used for the transformation, but this section will only focus on the non-mutant *SHR/SHR* progeny. Good transformation rates were obtained (table 4.9, p.163), and three lines (*$ZmANT1_{pro}$::ZmANT1-3*, *$ZmANT1_{pro}$::ZmANT1-4* and *$ZmANT1_{pro}$::ZmANT1-5*) were taken forward for T_2 genotyping and expression analysis. These lines all showed segregation ratios consistent with a single insertion event, tested using a chi-square test of independence (*$ZmANT1_{pro}$::ZmANT1-3*: $\chi^2(1, N = 30) = 2.5$, $p > 0.05$; *$ZmANT1_{pro}$::ZmANT1-4*: $\chi^2(1, N = 30) = 2.1$, $p > 0.05$; *$ZmANT1_{pro}$::ZmANT1-5*: $\chi^2(1, N = 30) = 0.5$, $p > 0.05$). Expression levels of *ZmANT1* were not consistent across the lines with gene silencing occurring in *$ZmANT1_{pro}$::ZmANT1-5* (fig.4.22; p.167). As such, only the two lines expressing *ZmANT1* were used for phenotypic analysis.

Line	Genetic background	Transgene	T_1 survival	T_1 seed set
ZmANT1$_{pro}$::ZmANT1-1	*SHR/SHR*	Yes	Yes	Good
ZmANT1$_{pro}$::ZmANT1-2	*shr-6/shr-6*	Yes	No	N/A
ZmANT1$_{pro}$::ZmANT1-3	*SHR/shr-6*	Yes	Yes	Good
ZmANT1$_{pro}$::ZmANT1-4	*SHR/shr-6*	Yes	Yes	Good
ZmANT1$_{pro}$::ZmANT1-5	*SHR/shr-6*	Yes	Yes	Good
ZmANT1$_{pro}$::ZmANT1-6	*SHR/SHR*	Yes	Yes	Good
ZmANT1$_{pro}$::ZmANT1-7	*SHR/SHR*	Yes	Yes	Good
ZmANT1$_{pro}$::ZmANT1-8	*SHR/shr-6*	Yes	Yes	Good
ZmANT1$_{pro}$::ZmANT1-9	*SHR/shr-6*	Yes	Yes	Good
ZmANT1$_{pro}$::ZmANT1-10	*SHR/shr-6*	Yes	Yes	Good
ZmANT1$_{pro}$::ZmANT1-11	*shr-6/shr-6*	Yes	No	N/A
ZmANT1$_{pro}$::ZmANT1-12	*SHR/shr-6*	Yes	Yes	Good
ZmANT1$_{pro}$::ZmANT1-13	*SHR/SHR*	Yes	Yes	Good
ZmANT1$_{pro}$::ZmANT1-14	*SHR/SHR*	Yes	Yes	Good
ZmANT1$_{pro}$::ZmANT1-15	*SHR/shr-6*	Yes	Yes	Good
ZmANT1$_{pro}$::ZmANT1-16	*SHR/shr-6*	Yes	Yes	Good

Table 4.9: *ZmANT1$_{pro}$::ZmANT1* arabidopsis transformation summary

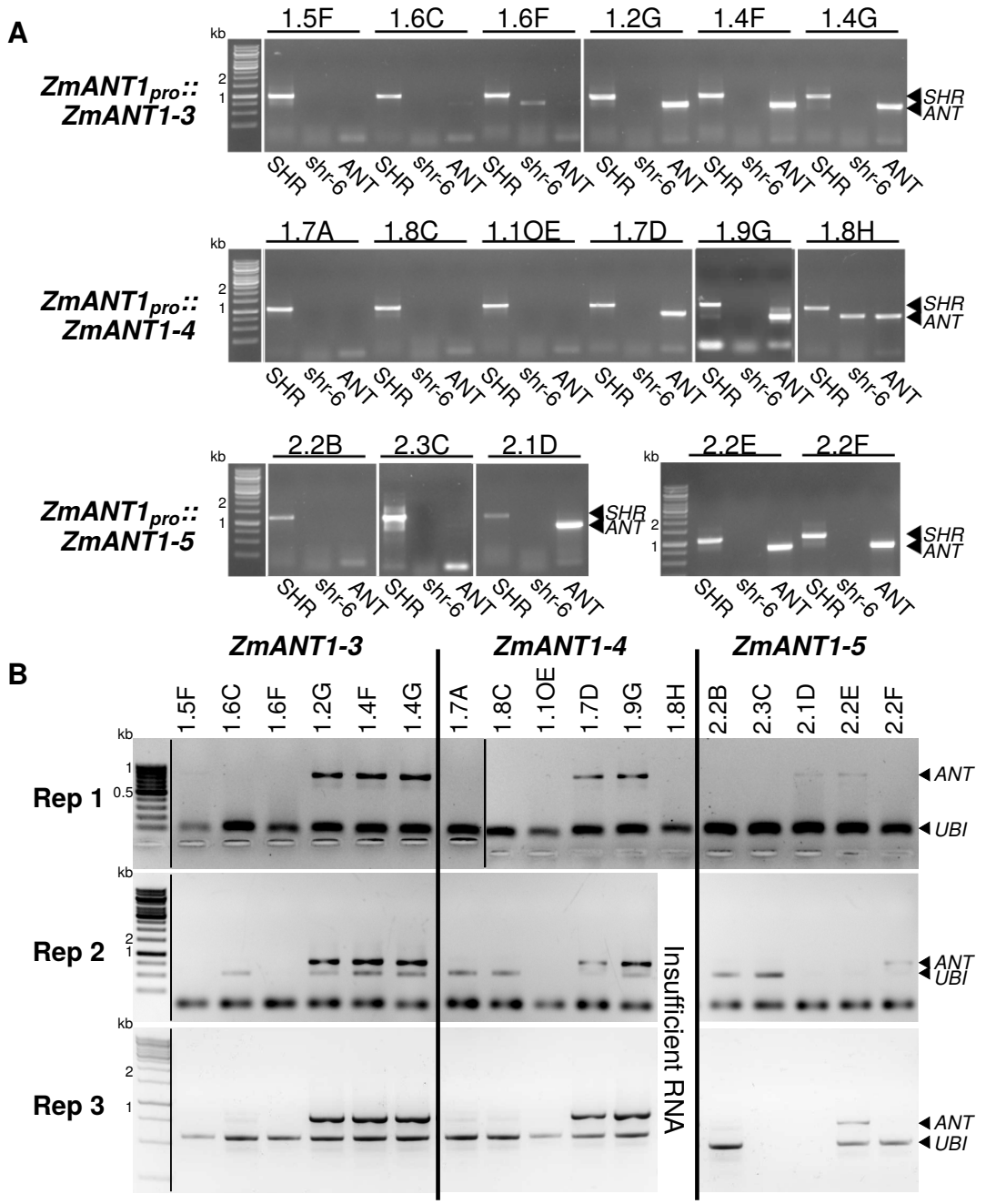


Figure 4.22: Composite of gel images showing *ZmANT1$_{pro}$:ZmANT1* arabidopsis genotyping and expression analysis. T_2 individuals from three *ZmANT1$_{pro}$:ZmANT1* arabidopsis lines were analysed for **A)** transgene insertion using PCR on genomic DNA. *SHR* and T-DNA insertion "*shr-6*" were also amplified to check genetic background. **B)** transgene expression using RT-PCR. Three cDNA synthesis replicates (Rep 1-3) were tested, *AtUBI* amplification was used as a positive control with primer pairs that amplified either 100bp (Rep 1) or 536bp (Rep 2-3). Each ID code is a unique identifier for one individual which corresponds with the RT-PCR ID, but is not related to the transformation event or family.

4.2.10.2 ZmANT1 activity in arabidopsis increases leaf area

To investigate the effect of *ZmANT1*$_{pro}$*::ZmANT1* expression on arabidopsis leaf development, leaf 5 was harvested from transgenic plants and wild-type siblings. The leaf was cleared and photographed using darkfield imaging and the images were analysed using Leaf Image Analysis Interface (LIMANI) software (Dhondt *et al.*, 2012) for leaf area, vein length, enclosed vein and free-ending vein length and number, and vein branching points (fig.4.23; p.166). Palisade cells, bundle sheath cells and vascular bundles were also imaged, using differential interference contrast (DIC) microscopy.

Leaf area was found to be significantly larger in *ZmANT1* expressing arabidopsis plants compared to wild-type siblings (fig.4.24, A; p.167). To investigate whether this increase was due to an increase in cell proliferation or in cell size, palisade cell measurements were taken from points mid-way between the leaf margin and mid-rib. No significant differences in palisade cell number or size were found in the middle of the leaf lamina (fig.4.24, B & C). Given that overall leaf size differed, but no cellular differences were observed in the middle of the leaf, these data show that ZmANT1 induced leaf growth through cellular proliferation and/or expansion either near the mid-rib or at the margin of the leaves.

4.2.10.3 ZmANT1 activity increases vein initiation in arabidopsis

To investigate the effect of ZmANT1 on vein development in arabidopsis, LIMANI calculated vein measurements were analysed. A significant increase in total vein length was observed in the *ZmANT1* expressing siblings, compared to wild type (fig.4.25, A; p.169), indicating that ZmANT1 has an effect on arabidopsis vein development. To test whether this increase in total vein length was solely due to the increased leaf size, the average whole leaf vein density was measured. To give a proportional measure of vein density, total vein length was then divided

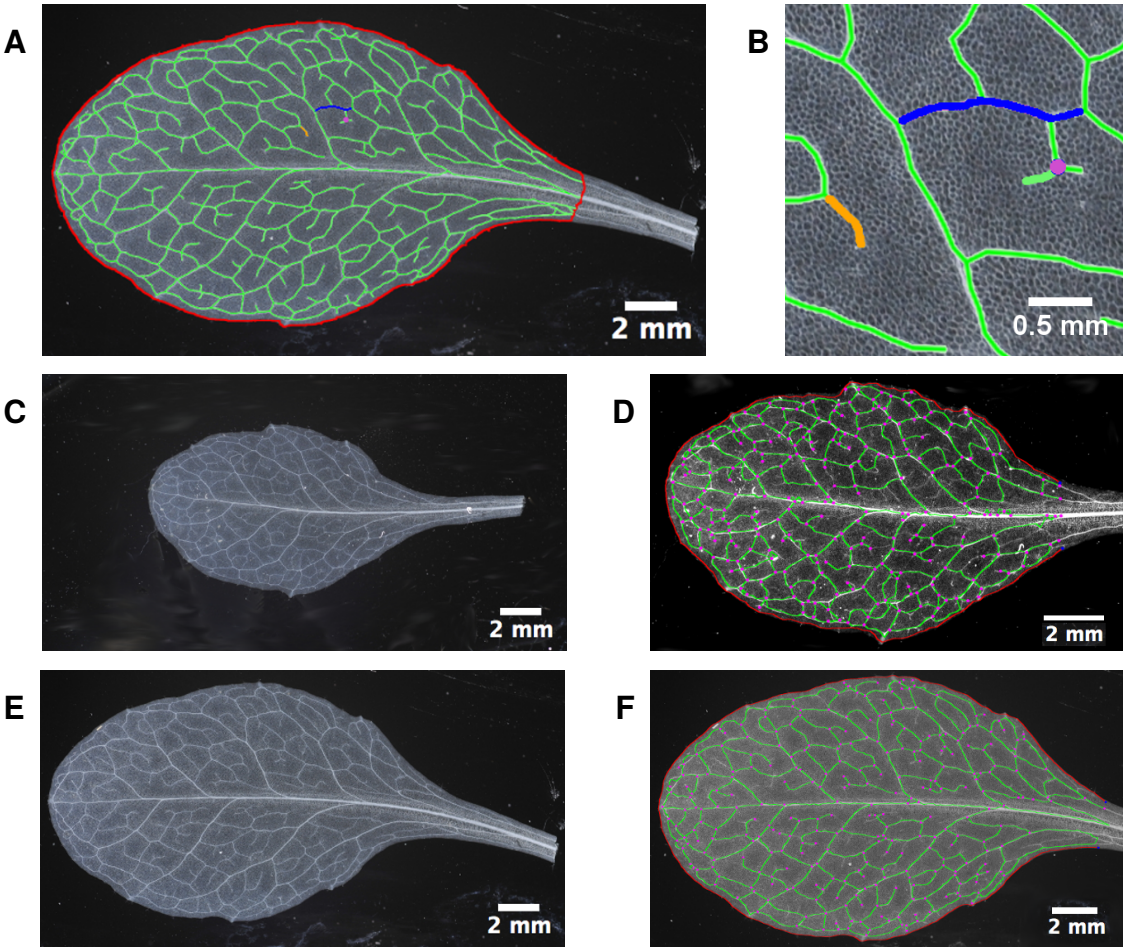


Figure 4.23: Arabidopsis leaf was measured using Leaf Image Interface (LIMANI) software. Darkfield photographs of cleared arabidopsis leaf 5- **A-B)** overlaid with the leaf traits measured: leaf area (red), total vasculature (green), enclosed veins (blue), free-ending veinlets (orange) and vein branching points (purple)- in a whole leaf (A) and a magnified image (B). **C-F)** from a segregating *ZmANT1*$_{pro}$*::ZmANT1* arabidopsis line - wild-type T_2 sibling- darkfield photograph (C) overlaid with LIMANI output (D) and *ZmANT1* T_2 sibling- darkfield photograph (E) overlaid with LIMANI output (F).

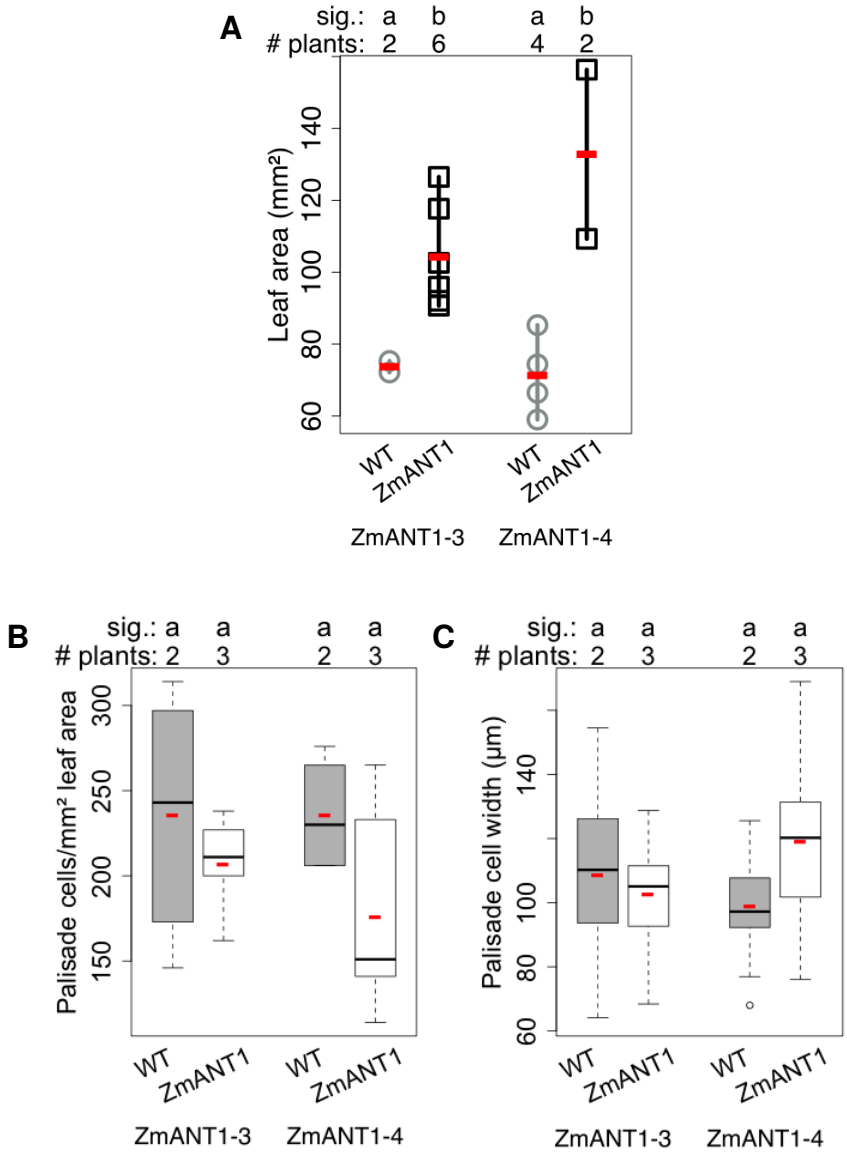


Figure 4.24: *ZmANT1$_{pro}$::ZmANT1* arabidopsis plants have larger leaves compared to wild-type. Leaf measurements from T_2 transgenic and wild-type siblings in two *ZmANT1$_{pro}$::ZmANT1* lines. **A)** Strip chart showing total leaf 5 area (mm^2) **B-C)** Boxplots showing- Palisade cell - number per mm^2 leaf area (B) and width (µm) (C). "Sig."- different letters indicate significant differences in group means ($p<0.05$). # plants- number of individual plants used for analysis. **Strip charts**: Square/circle- individual data point, vertical line- range of data, red dash- mean. **Boxplots**: Box- data within 25^{th}-75^{th} percentiles, with black horizontal line- median, red dash- mean, whiskers- lowest and highest value still within 1.5 IQR, open circle- value outside 1.5 IQR.

by the $\sqrt{}$leaf area. *ZmANT1* expressing siblings showed an increase in vein density compared to wild-type (fig.4.25, B), suggesting that the increase in total vein length was not solely due to increased leaf area. As such, ZmANT1 may induce increased procambium initiation in transgenic arabidopsis leaves. Procambium initiation frequency can be inferred from the number of free-ending veinlets and enclosed veins present in the leaf, and it was found that *ZmANT1* expressing siblings had significantly more free-ending veinlets per mm^2 leaf area than wild-type (fig.4.25, C). There was also a significant increase in enclosed vein length per $\sqrt{}$leaf area (fig.4.25, D) in *ZmANT1* expressing siblings. The significant increases in free-ending veinlets and enclosed veins indicates an increase in vein patterning complexity and provides evidence that ZmANT1 increases procambium initiation in arabidopsis.

4.2.10.4 ZmANT1 activity increases vein and bundle sheath width in arabidopsis

Given that constitutive expression of AtANT in arabidopsis and of ZmANT1 in rice increased the cross-sectional area of vascular bundles (Kang *et al.*, 2007 and section 4.2.6.2; p.135, respectively), the effect of ZmANT1 on vein size in arabidopsis was also investigated. A trend for increased vein width was found in transgenic siblings, compared to wild-type (fig.4.26, A; p.170). Bundle sheath size was also increased by ZmANT1 in rice. To investigate whether this effect was conserved in arabidopsis, bundle sheath width and length was also measured (fig.4.26, B-C; p.170). A trend for increased bundle sheath width in *ZmANT1* expressing siblings compared to wild-type was observed. These data indicate that the effects of ANT on vascular cell proliferation or expansion, and on bundle sheath size may be conserved between eudicots and monocots.

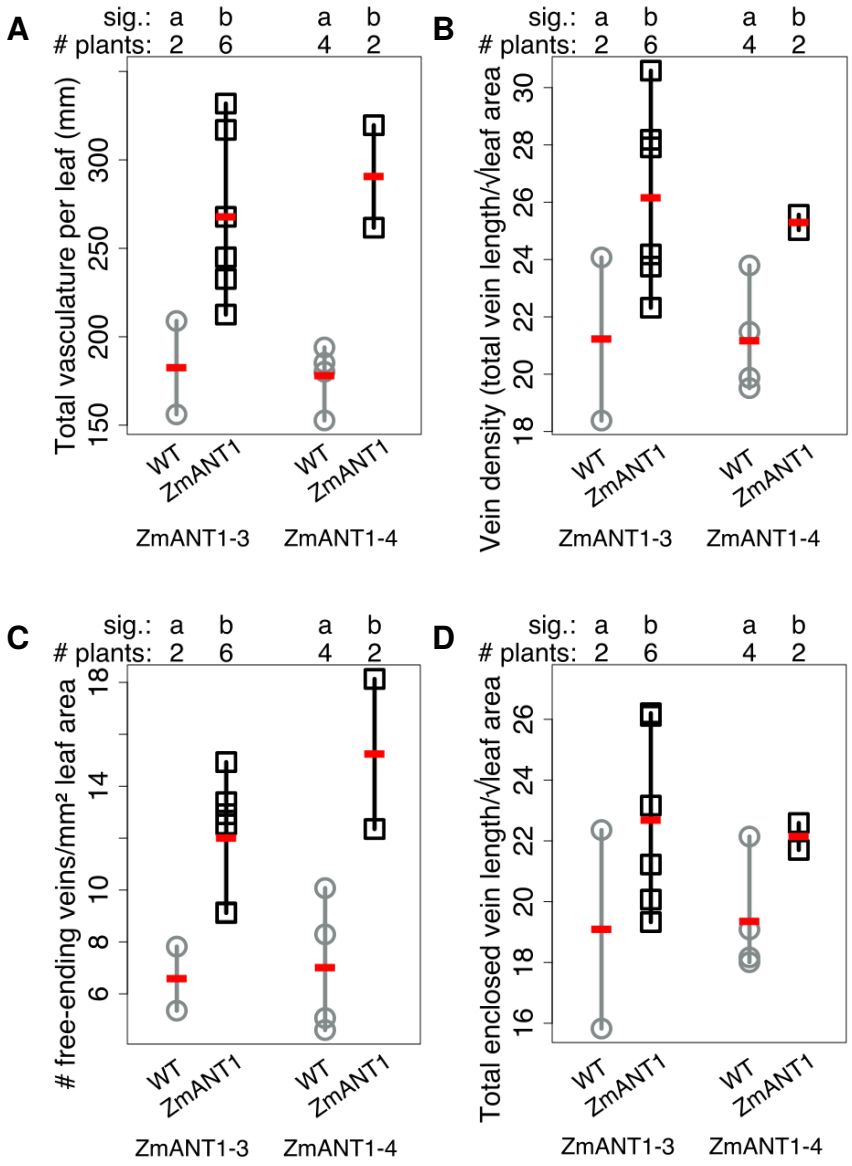


Figure 4.25: *ZmANT1$_{pro}$::ZmANT1* transgenic arabidopsis lines have more vasculature and a higher vein density compared to wild-type. Whole leaf analysis of T_2 segregating siblings from three *ZmANT1$_{pro}$::ZmANT1* arabidopsis lines. **A-D)** Strip charts showing whole leaf- total vasculature (mm) (A), vein density (mm/mm) (total vein length/$\sqrt{}$leaf area) (B), free-ending veinlet number per mm^2 leaf area (C) and enclosed vein density (total enclosed vein length/$\sqrt{}$leaf area) (D). "Sig."- different letters indicate significant differences in group means ($p<0.05$). # plants- number of individual plants used for analysis. **Strip charts**: Square/circle- individual data point, vertical line- range of data, red dash- mean.

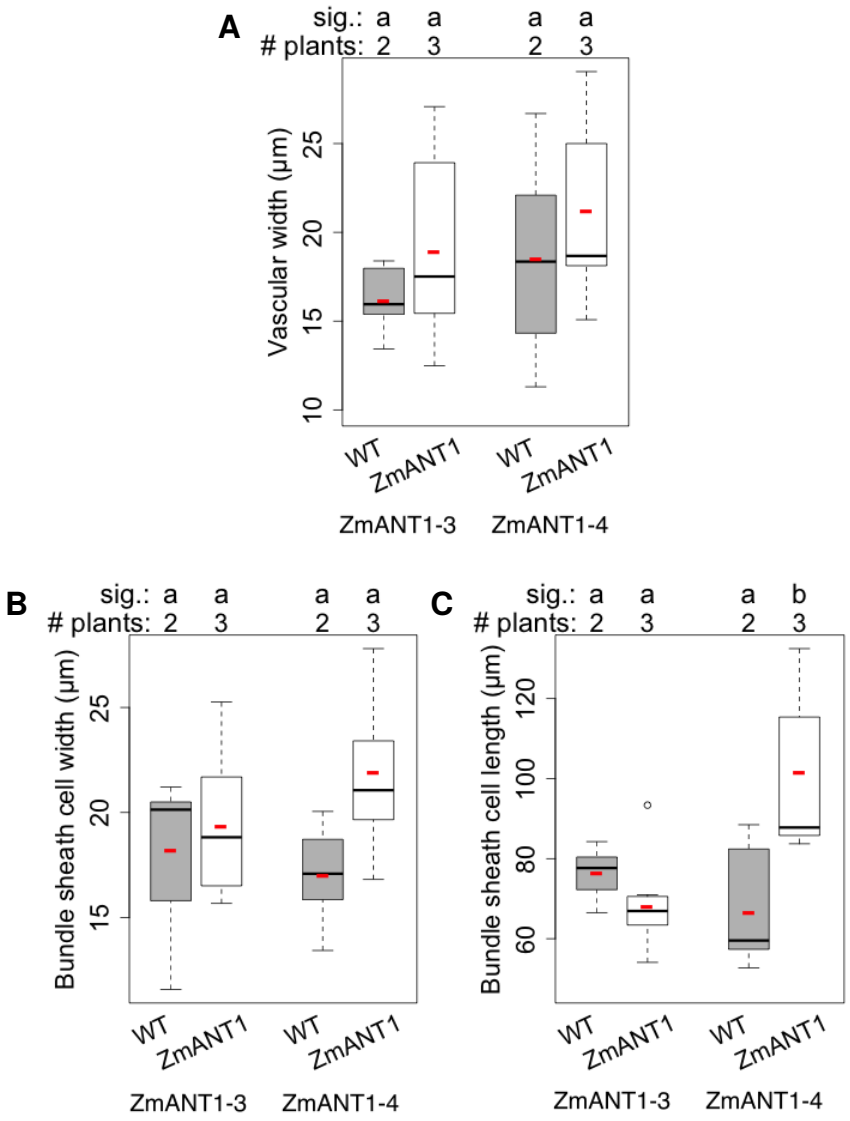


Figure 4.26: *ZmANT1$_{pro}$::ZmANT1* transgenic arabidopsis lines show increased vein width trend. T_2 segregating siblings from two *ZmANT1$_{pro}$::ZmANT1* arabidopsis lines. Boxplots showing **A)** vein width and **B-C)** bundle sheath width (B) and length (C). "Sig."- different letters indicate significant differences in group means ($p<0.05$). # plants- number of individual plants used for analysis. **Boxplots**: Box- data within 25th-75th percentiles, with black horizontal line- median, red dash- mean, whiskers- lowest and highest value still within 1.5 IQR, open circle- value outside 1.5 IQR.

4.3 Discussion

4.3.1 The effect of ZmANT1 in leaf development is not conserved between arabidopsis and rice

4.3.1.1 ANT protein role in leaf lamina growth and vein initiation is not conserved between arabidopsis and rice

ZmANT1 expression in rice and arabidopsis resulted in contrasting leaf development phenotypes; leaf lamina showed a non-significant trend for growth restriction in rice, whereas in arabidopsis leaf area was significantly increased. Cell divisions in the ground tissue are key drivers of lamina growth (Foster, 1952; Dewitte *et al.*, 2003), and M cell counts in transgenic rice plants suggest that ZmANT1 expression restricts these divisions. In contrast, *AtANT* overexpression in arabidopsis increases ground meristem cell proliferation (Mizukami & Fischer, 2000; Kang *et al.*, 2007), and a similar phenotype is seen in *ZmANT1* expressing arabidopsis leaves. Interestingly, the phenotypic similarity between $ZmANT1_{pro}$*::ZmANT1* and $35S_{pro}$*::AtANT* plants may indicate that the effects of ANT proteins on arabidopsis leaf development are non cell-autonomous or that the $ZmANT1_{pro}$ gives a broad expression domain during leaf development. In either case, these phenotypic similarities indicate that the contrasting effects of ZmANT1 in rice and arabidopsis are unlikely to be caused by different promoters used to drive expression in the two species. Instead, the data suggest that the role of ANT proteins in regulating ground tissue development differs between arabidopsis and rice.

The effect of *ZmANT1* expression on procambium initiation was also different between rice and arabidopsis; ZmANT1 did not have a measurable effect on vein initiation in rice but did have a significant positive effect in arabidopsis. Current understanding of leaf development maintains that for procambium initiation to occur in leaves, the ground tissue must maintain meristematic competence.

Published evidence shows that *AtANT* over-expression in arabidopsis acts to prolong meristematic competence in ground tissue (Mizukami & Fischer, 2000; Kang *et al.*, 2007), and this study indicates that ZmANT1 can play a similar role during arabidopsis leaf development. In contrast, constitutive over-expression of *ZmANT1* in rice did not increase procambium initiation, instead, there was a non-significant trend for ground tissue cell proliferation and restricted growth. If the rice experimental results show similar, but significant, differences when a larger sample size is used, these results may indicate that the regulation of cell division and procambium initiation in ground leaf tissue differs between the two species and that the role of ANT proteins in regulating ground tissue meristematic competence is not conserved between rice and arabidopsis.

4.3.1.2 ZmANT1 interactions with auxin and cytokinin regulatory networks are not conserved between arabidopsis and rice

Cytokinin is a key hormone for maintaining meristematic competence in rice shoots (Kurakawa *et al.*, 2007) and this study showed that exogenous cytokinin significantly increased rice leaf width with a non-significant trend for higher mesophyll cell numbers. ZmANT1 over-expression appeared to be antagonistic to these cytokinin-mediated effects on leaf development, however larger sample sizes are required to make any statistically valid conclusions. Histological studies on rice leaf development in chapter 3 indicate that mesophyll cell number is determined by the extent of ground tissue cell division in the developing primordia. Therefore the results in this section may indicate that ZmANT1 inhibits cytokinin-mediated ground tissue proliferation in rice. There were also trends for increased bundle sheath cell size in ZmANT1 rice and cytokinin treated wild-type rice, but these effects were suppressed when the two inducers were combined. Overall, these data indicate that ZmANT1 may act as an antagonist to cytokinin,

however, the mechanism of this antagonism is not clear, and further replication of the experiments is required to show statistically valid data. *In silico* analysis of maize promoter regions showed that putative ANT targets include both auxin and cytokinin related genes. Given that auxin is known to be antagonistic to cytokinin (Ioio *et al.*, 2008), it is possible that constitutive *ZmANT1* expression modulates the relative levels of cytokinin and auxin in rice leaves, promoting auxin-mediated inhibition of cytokinin signalling.

In contrast to the effects of ZmANT1 in rice, the protein did not inhibit ground tissue proliferation in arabidopsis. As in rice, cytokinin maintains meristematic competence in arabidopsis leaves and some molecular mechanisms for this maintenance have been elucidated. For example, cytokinin is involved in controlling cyclin D3 (CYCD3) activity (Dewitte *et al.*, 2007). CYCD3 proteins are only active in non-differentiated cells where they drive cell proliferation by promoting cell cycle progression (Dewitte *et al.*, 2003). In arabidopsis, ANT proteins appear to have an indirect role in regulating CYCD3 in that over-expression of *AtANT* prolonged *CYCD3;1* expression (Mizukami & Fischer, 2000) even though there is evidence that AtANT does not directly regulate *CYCD3* genes (Randall *et al.*, 2015a,b). The lack of a direct interaction is supported by *in silico* analysis which shows that *CYCD3* gene promoters do not have the ANT binding DNA consensus sequence. These data indicate that, in contrast to the effect of ZmANT1 in rice, ANT proteins do not inhibit cytokinin-mediated regulation of ground tissue during arabidopsis leaf development. Given also that the predicted auxin and cytokinin related AIL target genes in arabidopsis were not orthologous to those from maize, it is likely that the role of ANT proteins in the regulation of hormone signalling during leaf development is not conserved between monocots and eudicots.

4.4 Future perspectives

4.4.0.1 ZmANT1 has potential for Kranz anatomy engineering in eudicots

Although *ZmANT1* expression is not sufficient to cause Kranz-like characteristics in a monocot C_3 context, it does cause increased vein density in arabidopsis. This indicates that ZmANT1 is a good candidate for engineering Kranz anatomy into C_3 eudicots, which may be economically valuable for oil crop production. ZmANT1 may be required to engineer C_4 in a monocot context, but further investigation into fundamental leaf development differences between plant species is required.

4.4.0.2 Role of other *AINTEGUMENTA-like* genes in Kranz development

ZmANT1 is one member of an expanded gene clade in monocots from which other closely related genes have been identified as putative Kranz regulators in maize. Given the sequence similarities between AIL proteins, it is probable that the target genes overlap. It is therefore possible that the effect of ANT1 on auxin and cytokinin signalling networks in monocot leaf development is modulated by the activity of other AIL proteins. More specifically, co-regulation of gene targets by different AIL proteins may be important for veination patterning. There is some evidence to support this theory. For example, a mutant maize line with decreased *ZmCRL5* expression showed perturbations in Kranz patterning (Yardley, 2013, BA report), and experiments with an RNAi knock-down line of *OsAIL5* in rice showed decreased vein initiation (see appendix B; p.233). These preliminary findings indicate that other AIL genes may work in combination with ZmANT1 during Kranz development and should not be dismissed from further study.

4.4.0.3 ZmANT1 interaction with cytokinin in maize

A distinctive feature of Kranz development is the maintenance of competency to initiation procambium in ground tissue, which prolongs the developmental window for vein initiation leading to increased vein density. In maize, high levels of *ZmANT1* expression are observed during Kranz development in tissues where ground cells are proliferating and procambium is initiating. This may indicate that, in contrast to the effect of ZmANT1 in rice, the protein does not inhibit cytokinin-mediated control of ground tissue proliferation in maize. Notably, a gene encoding an auxin synthetase protein (an enzyme that reduces free-auxin levels through conjugation of IAA with amino acids) is highly upregulated during maize Kranz development, and this gene was identified as an AIL protein target gene. Therefore, it is possible that during Kranz development, an AIL protein promotes the sequestration of auxin in ground tissues thereby altering the relative cytokinin-auxin levels. Consequently, cytokinin-mediated ground tissue cell division is maintained even though ZmANT1 (which is shown to inhibit cytokinin-signalling in monocot leaf development) is up-regulated. Little is known about the auxin and cytokinin levels in developing Kranz tissue, and alterations in the hormone balance may be key to engineering Kranz anatomy in a C_3 context.

5

The role of *DOF AFFECTING GERMINATION-like* genes in Kranz development

5.1 Introduction

5.1.1 DAG-like proteins are identified as putative Kranz regulators in monocots and eudicots

The *in silico* analysis of ANT protein target genes in the previous chapter identified the maize gene *GRMZM2G171852* as a potential ANT target. This gene encodes a DNA-binding with one finger (Dof) protein which is homeologous to *GRMZM2G114998* (Hughes *et al.*, 2014). Strikingly, the maize transcriptome analysis also identified these genes as putative Kranz regulators. Due to their orthology with the arabidopsis *DOF AFFECTING GERMINATION* (DAG) genes (Wang *et al.*, 2013b; Fouracre *et al.*, 2014), the genes were named *DAG-like 1 HOMEOLOG* (*ZmDLK1H*) and *DAG-like 1* (*ZmDLK1*), respectively (table 5.1; p.178). Interestingly, transcriptome analysis of developing leaves in two Cleomaceae species *Gynandropsis gynandra* (C_4 eudicot) and *Tarenaya hassleriana*

(C_3 dicot) also identified a Dof gene which was upregulated during C_4 leaf development. This gene is orthologous to *AT4G24060* (a gene in the *DAG* clade) (Külahoglu *et al.*, 2014; Huang *et al.*, 2016). These observations indicate that the role of *DAG-like* (*DLK*) genes in C_4 development may be conserved between monocots and eudicots.

		P1&2	P3&4	P5	I	E
GRMZM2G114998	**F**	43	149	237	164	119
(ZmDLK1)	**H**	32	66	124	49	43
GRMZM2G171852	**F**	175	439	507	184	142
(ZmDLK1H)	**H**	122	224	334	172	117

	BS	M
GRMZM2G114998	38	0
GRMZM2G171852	15	0

Table 5.1: *DOF AFFECTING GERMINATION*-like genes identified as putative Kranz regulators in maize. mRNA expression in relative transcript abundance in- foliar (F) and husk (H) plastochron 1 (P1), 3&4 (P3&4) and 5 (P5), immature (I) and expanded (E) leaves (Wang *et al.*, 2013b), and in mature bundle sheath (BS) and mesophyll (M) cells (Li *et al.*, 2010).

5.1.2 DAG-like proteins are mobile transcription factors that can act in protein complexes as transcriptional repressors

Dof protein functional domains have been identified and the family are characterised by a highly conserved zinc finger DNA binding domain in the N' terminal half of the protein (Yanagisawa, 1995). The DNA binding motif for a tobacco DLK protein (NtBFF1) has been identified as ACTTTA (Baumann, 1999). Dof proteins also have a bipartite nuclear localisation domain (Krebs *et al.*, 2010), and DLK proteins have a conserved intercellular trafficking motif (ITM) that confers protein mobility across cells (Chen *et al.*, 2013) (fig.5.1; p.180). The transcrip-

tional activation domain has been identified at the C' terminal end of DOF proteins (Yanagisawa, 2001; Diaz *et al.*, 2002). Given these functional domains, ZmDLK proteins may regulate gene transcription in a non-cell autonomous manner.

Dof proteins have been found to interact with other TFs to regulate transcription of downstream target genes. For example, in seed endosperm, the maize Dof protein ZmPBF directly interacts with ZmO2 (OPAQUE2, a bZIP TF) to regulate expression of *Zein* genes (Vicente-Carbajosa *et al.*, 1997) and the barley DLK protein HvSAD interacts with HvGAMYB (a MYB protein) to repress α-*amylase* gene expression (Diaz *et al.*, 2005). Similarly, the ZmDLK proteins may be able to operate in transcriptional complexes to modulate gene expression during leaf development.

5.1.3 DAG-like proteins are implicated in auxin responses and leaf development

Auxin plays a fundamental role in procambium initiation and vein development, and two *DLK* genes have been shown to be associated with auxin during leaf development. *AtDOF4.6* expression in wild-type arabidopsis overlaps with *ATHB8*, but the domain is broader, especially in undifferentiated leaf tissue. This expression domain is further expanded in leaf primordia treated with an auxin transport inhibitor, suggestive of auxin-related regulation of transcript accumulation (Gardiner *et al.*, 2010). In contrast, mRNA levels of *NtBFF1* were not affected when tobacco mini-explants were treated with exogenous auxin. However, the protein was required for auxin-responsive downstream effects, suggesting post-translational auxin-mediated regulation of NtBFF1 (Baumann, 1999). These two examples imply that DLK genes can be regulated by auxin at both the transcriptional and post-translational levels.

Both *NtBFF1* and *AtDOF4.6* have a broad expression domain in leaf primordia that becomes restricted to vasculature later in development. The barley gene

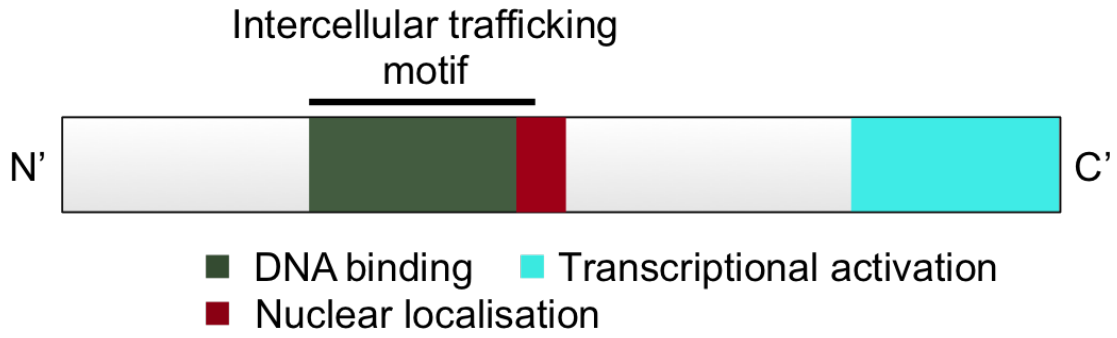


Figure 5.1: Functional domains in DAG-like proteins. DAG-like proteins have a N' terminal DNA-binding with one finger (Dof) domain (green) and a nuclear localisation domain (NLS) (red). There is a conserved intercellular trafficking motif (ITM) (black bar) that confers protein mobility across cells, and a C' terminal transcription activation domain (blue) (Yanagisawa, 1995; Krebs *et al.*, 2010; Chen *et al.*, 2013). Image adapted from Le Hir & Bellini (2013).

HvSAD and arabidopsis gene *AtDOF4.1* (*AtITD1*) also have this expression pattern (Isabel-LaMoneda *et al.*, 2003; Ahmad *et al.*, 2012). Constitutive expression of *AtDOF4.1* in arabidopsis caused growth retardation and plants produced small, curled leaves (Ahmad *et al.*, 2012). Strikingly, this phenotype was also seen in *$35S_{pro}$::AtOBP2* (a non-DLK DOF gene) arabidopsis lines. AtOBP2 activates transcription of two cytochrome P450 genes, thereby regulating auxin biosynthesis, and constitutive expression was shown to elevate free auxin levels in transgenic plants (Skirycz *et al.*, 2006). Given the similarity in phenotypes, *AtDOF4.1* over-expression may also be elevating auxin levels (Ahmad *et al.*, 2012). Because *DLK* genes are implicated in auxin-mediated responses, and the expression patterns of all published *DLK* genes correlate with pre-procambial leaf primordia tissues and developing vasculature, ZmDLK1 proteins may play a role in procambium initiation and vein development in maize leaves.

The aim of this chapter was to determine whether *ZmDLK1* genes function as Kranz regulators. If the proteins have a role in procambium initiation, and this role is conserved between rice and maize, vein patterning in rice should be perturbed in lines constitutively expressing the *ZmDLK1* or *ZmDLK1H* genes. Alternatively, if protein function is more related to auxin metabolism and/or signalling, the transgenic rice plants may exhibit phenotypes that are consistent with perturbations in auxin responses.

5.2 Results

5.2.1 The *DAG-like* clade underwent gene duplications in both monocots and eudicots

To understand the evolutionary relationships between the *ZmDLK* genes and published *DLK* genes, a maximum-likelihood phylogenetic tree was inferred

using amino acid sequences from six eudicot and seven monocot species and RAxML software (Stamatakis, 2014) (fig. 5.2; p.183). *Amborella trichopoda* was included as a species that diverged before the monocot-eudicot divergence, and *Physcomitrella patens* was used as an outgroup. To resolve the tree and increase branch length accuracy, basal monocot and eudicot species were also included; *Acquilegia coerulea* diverged after the monocot-eudicot divergence (at the base of the eudicots), and *Spirodela polyrhiza* diverged at the base of the monocots.

Gene duplication has been linked with the evolution of novel phenotypes and the tree shows that DLK genes underwent several gene duplications after the monocot-eudicot divergence; both in monocots and eudicots. These duplication events mean that the resulting proteins could have undergone sub- or neo-functionalisation, altering protein activity during leaf development and possibly giving rise to Kranz development regulators in both eudicot and monocot species.

5.2.2 Verification and correction of *ZmDLK1* and *ZmDLK1H* gene models using RNAseq data

To verify gene model annotations, the *ZmDLK1* and *ZmDLK1H* gene models in the Phytozome database (Goodstein *et al.*, 2012) were compared with maize RNAseq data (Wang *et al.*, 2013b). Only one transcript was predicted for each of the candidate regulators, and the *ZmDLK1* gene model correlated with the RNAseq data (fig. 5.3; p.184). However, the *ZmDLK1H* gene model was truncated, and the 5′ half of the transcript was misannotated as an untranslated region (fig. 5.4; p.185). In fact, RNAseq data showed that the *ZmDLK1H* gene has a similar structure to *ZmDLK1* and has an intron in the 5′ half of the gene.

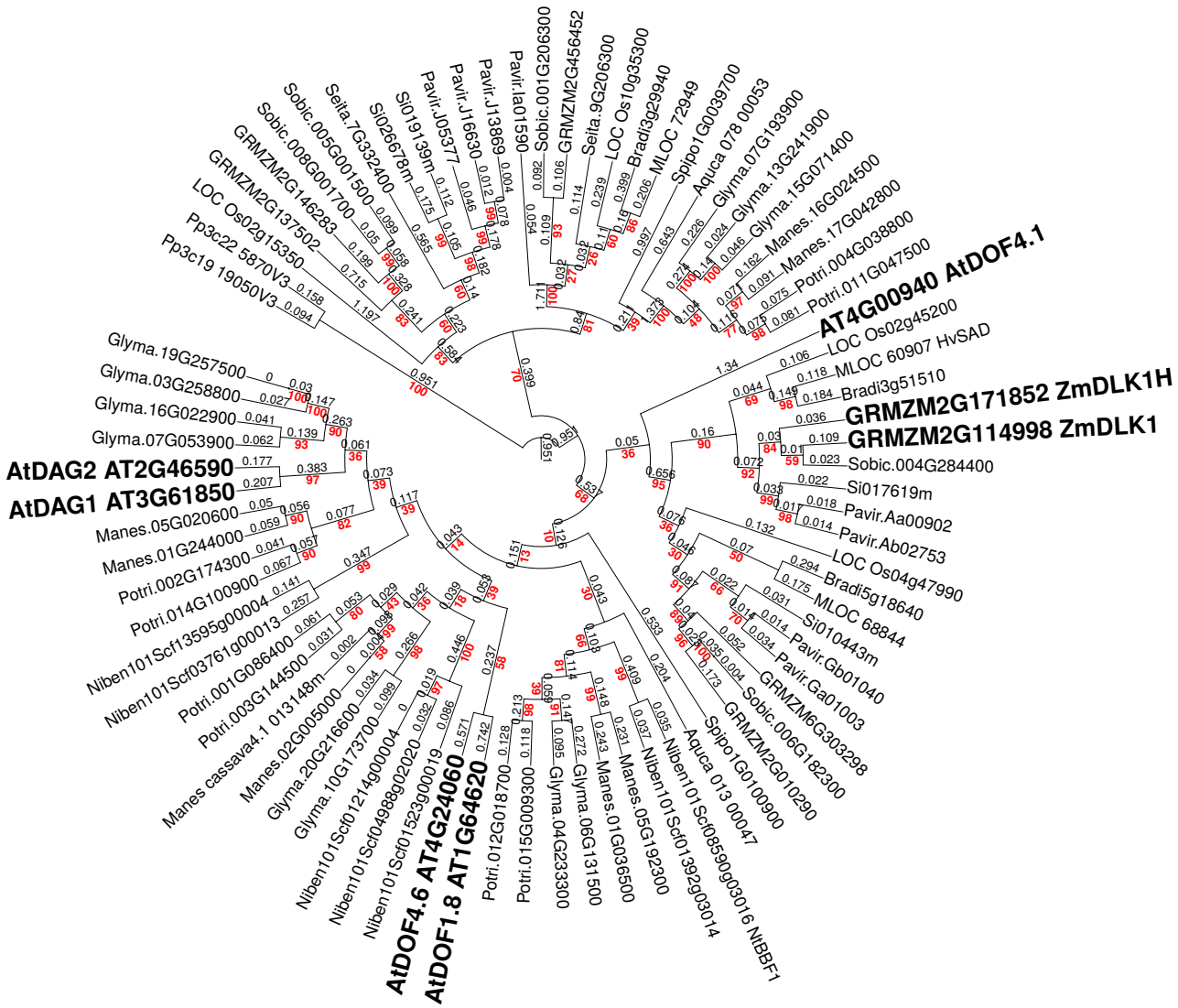


Figure 5.2: Phylogenetic tree of DAG-like proteins. Maximum-likelihood inference of phylogenetic relationships between the DAG-like (DLK) proteins in monocot and eudicot species. Branch length measured in number of amino acid substitutions per site (black branch labels), bootstrap values shown for each node (red branch labels). Arabidopsis DAG proteins ("At" bold labels) and putative Kranz candidates ("GRMZM" bold labels) are highlighted.

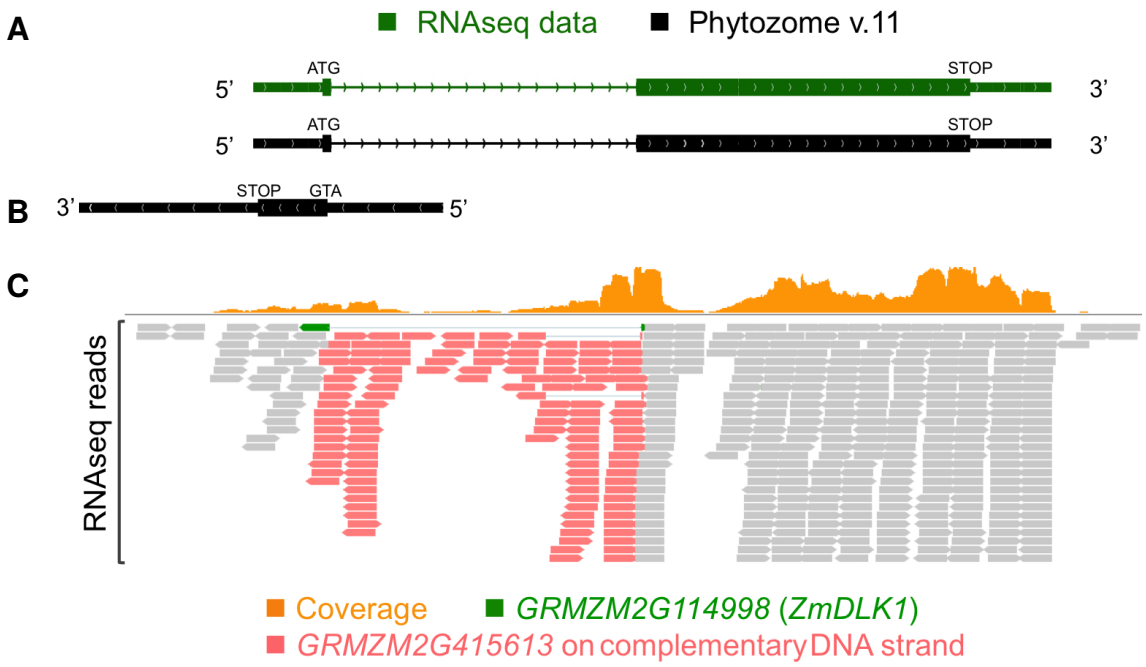


Figure 5.3: *ZmDLK1* gene model is correctly annotated in Phytozome.

A-B) *ZmDLK1* gene model (A) and *GRMZM2G415613* (an anti-sense transcript) (B) from Phytozome v.11 (black), *ZmDLK1* is compared to full length transcript from RNAseq data for *ZmDLK1* (green). Translation start ("ATG") and stop ("STOP") sites are indicated. Thin horizontal line with vertical dashes- intron, medium horizontal bar- untranslated region (UTR), thick horizontal bar- exon. **C)** Wang *et al.* (2013b) RNAseq transcript reads from maize leaf primordia (plastochrons 3 & 4) aligned to genome models in (A & B). RNA reads mapped to genome using TopHat (S. Kelly, personal communication; Kim *et al.*, 2013). Coverage of genome by reads (orange), RNA reads (gray), reads including *ZmDLK1* intron (green) and putative annotation of reads from *GRMZM2G415613* on complementary DNA strand (pink).

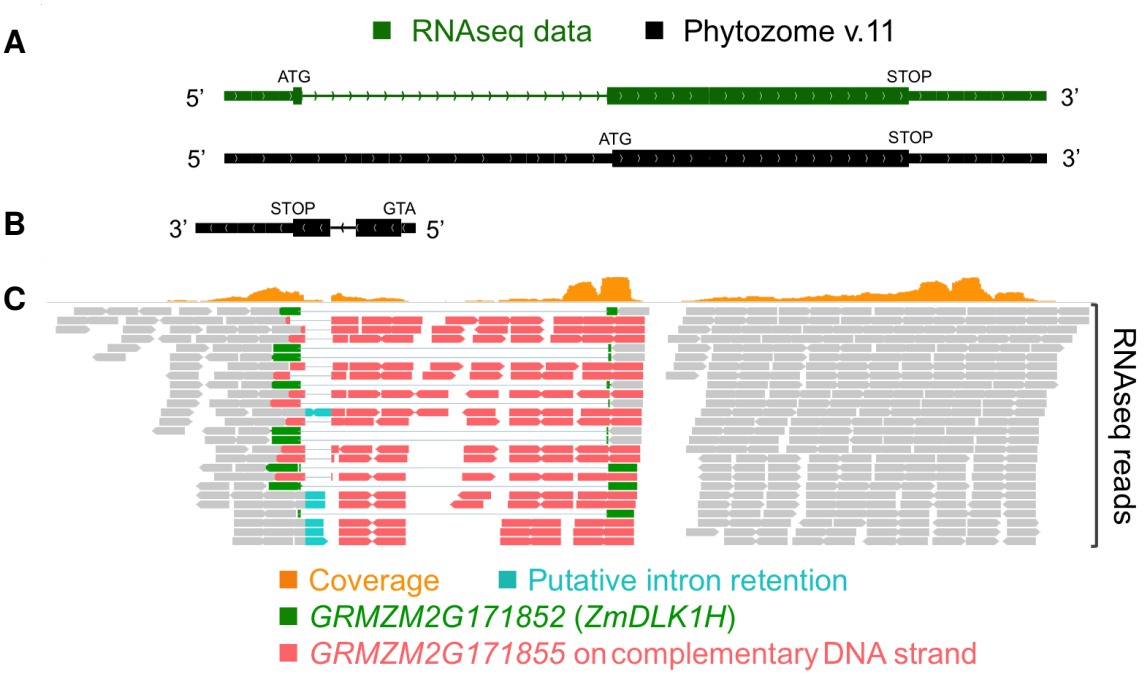


Figure 5.4: Mis-annotation of *ZmDLK1H* gene model in Phytozome.

A-B) *ZmDLK1H* gene model (A) and *GRMZM2G171855* (an anti-sense transcript) (B) from Phytozome v.11 (black), compared to full length transcript from RNAseq data for *ZmDLK1H* (green). Translation start ("ATG") and stop ("STOP") sites are indicated. Thin horizontal line with vertical dashes- intron, medium horizontal bar- untranslated region (UTR), thick horizontal bar- exon. **C)** Wang *et al.* (2013b) RNAseq transcript reads from maize leaf primordia (plastochrons 3 & 4) aligned to genome models in (A & B). RNA reads mapped to genome using TopHat (S. Kelly, personal communication; Kim *et al.*, 2013). Coverage of genome by reads (orange), RNA reads (gray), putative intron retention (blue), reads including *ZmDLK1H* intron (green) and putative annotation of reads from *GRMZM2G171855* on complementary DNA strand (pink).

5.2.3 *ZmDLK1* and *ZmDLK1H* may be regulated by non-coding RNA

Genes can be encoded on either strand of genomic DNA, and two genes can be present in the same region of the genome on complementary strands. In these situations, the transcript from one of the DNA strands (the anti-sense RNA) can regulate the expression of the gene located on the complementary strand (reviewed in Pelechano & Steinmetz, 2013). The loci for both *ZmDLK1* and *ZmDLK1H* contain two open reading frames (ORFs), one on each DNA strand, and RNAseq data shows that all of the ORFs are transcribed in young maize leaf primordia (fig. 5.3 & 5.4). Notably, the anti-sense transcript sequences (*GRMZM2G415613* and *GRMZM2G171855*) overlap with the *ZmDLK* coding regions, so may have a role in regulating *ZmDLK* gene expression in the developing maize leaf.

5.2.4 Non-coding RNA is not expressed from the rice *OsDLK1* locus

To investigate whether the putative anti-sense RNA regulatory pathway is conserved in rice, RNAseq data for the rice *OsDLK1* locus was analysed (van Campen *et al.*, 2016) (fig. 5.5; p.187). In contrast to maize, very few putative anti-sense RNA transcripts were detected from the *OsDLK1* locus during rice leaf development. This suggests that the regulation of DLK1 genes during leaf development differs between rice and maize.

5.2.5 Conserved C_3 versus C_4 differences between monocot DLK1 protein sequences are prevalent in the C' region

To investigate whether DLK protein sequences have diverged in C_3 and C_4 monocots, protein alignments were generated using MergeAlign (Collingridge & Kelly,

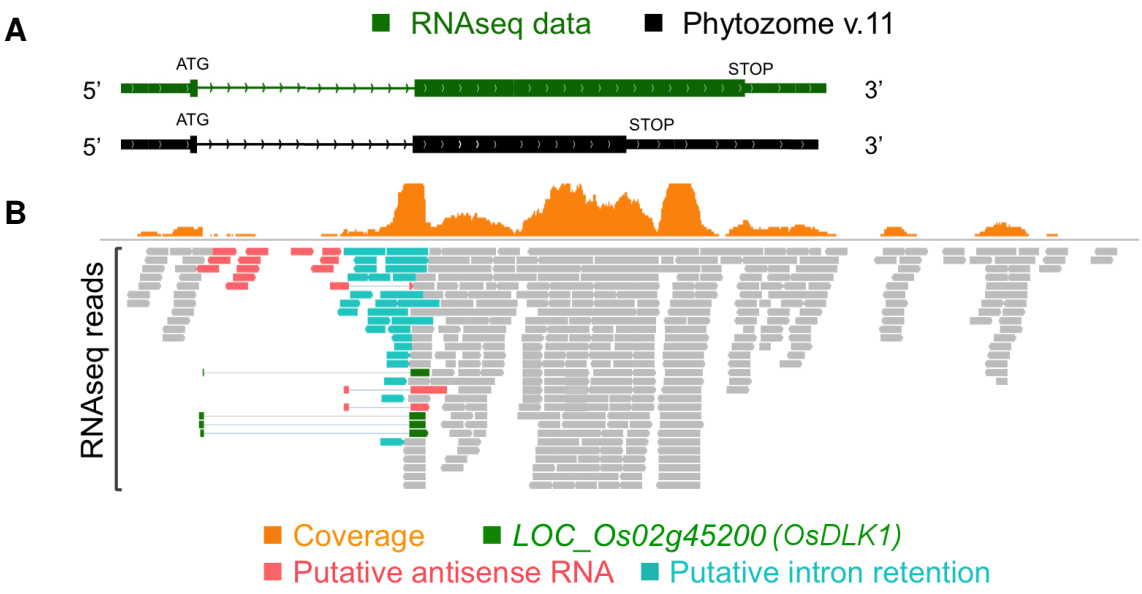


Figure 5.5: Mis-annotation of *OsDLK1* gene model in Phytozome.

A) *OsDLK1* gene model from Phytozome v.11 (black), compared to full length transcript from RNAseq data (green). Translation start ("ATG") and stop ("STOP") sites are indicated. Thin horizontal line with vertical dashes- intron, medium horizontal bar- untranslated region (UTR), thick horizontal bar- exon. **B)** van Campen *et al.* (2016) RNAseq transcript reads from rice leaf primordia (plastochron 3) aligned to genome models in (A). RNA reads mapped to genome using TopHat (S. Kelly, personal communication; Kim *et al.*, 2013). Coverage of genome by reads (orange), RNA reads (gray), reads including *OsDLK1* intron (green) and putative reads from complementary DNA strand (pink).

2012) for four C_4 species (maize, *Sorghum bicolor*, *Setaria italica*, *Panicum virgatum*) and three C_3 species (rice, *Brachypodium distachyon* and *Hordeum vulgare*) (fig. 5.6; p.189). The ITM, DNA binding and NLS domains were all highly conserved between species, however, there were differences in the putative transcriptional activation domain. Given that the C_4 species used for the alignment originate from two independent origins of C_4 (Grass Phylogeny Working Group II, 2012), the conserved residue differences may represent differing abilities of C_3 vs C_4 DLK1 proteins to activate or repress target gene transcription.

5.2.6 Developmental consequences of *ZmDLK1* overexpression in rice

5.2.6.1 Transformation of *ZmDLK1* into rice led to low rates of regeneration and poor seed set

To test the activity of ZmDLK1 in a C_3 developmental context, a *35S$_{pro}$::HYG-ZmUBI$_{pro}$::ZmDLK1* construct was transformed into rice callus using agrobacterium-mediated gene transfer. T_0 plants were regenerated on cytokinin-rich media but regeneration rates were poor and transgenic plants showed retarded growth and were weak. Only two transgenic T_0 lines set seed (table 5.2; p.188). T_1 seed grown from these lines, exhibited high seedling mortality with the surviving plants segregating for the retarded growth phenotype. Only three of these weak plants survived whilst the other siblings were vigorous.

Line	Transgene	T_0 survival	T_0 seed set
ZmDLK1-1	Yes	Yes	Poor
ZmDLK1-2	Yes	Yes	Poor
ZmDLK1-3	Yes	No	NA

Table 5.2: *ZmUBI$_{pro}$::ZmDLK1* transformation regeneration summary.

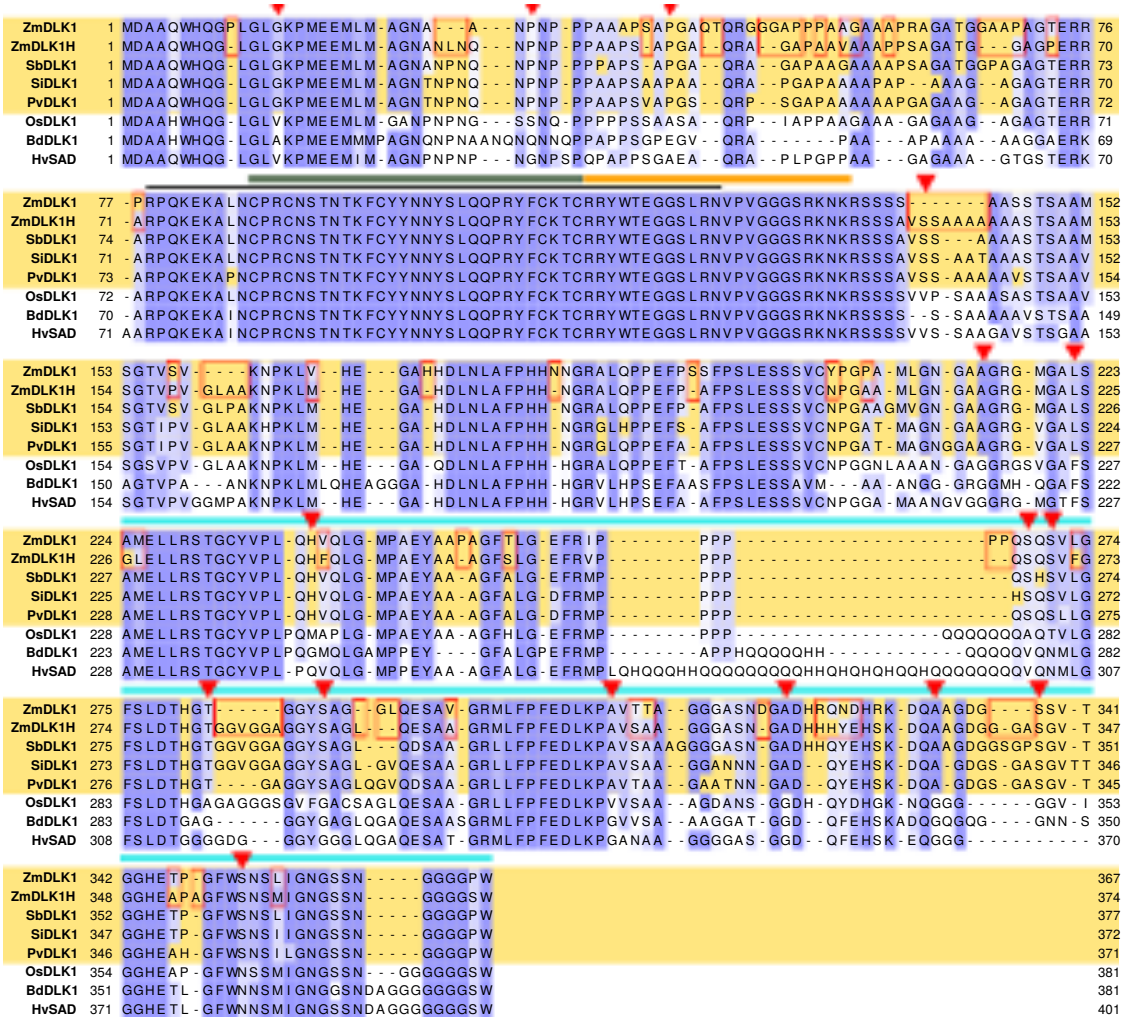


Figure 5.6: DLK1 amino acid sequence differences between C_3 and C_4 monocots are present in the putative transcriptional activation domain. DLK1 protein sequences alignment from monocot species; four C_4- *Zea mays* (Zm), *Sorghum bicolor* (Sb), *Setaria italica* (Si), *Panicum virgatum* (Pv) (yellow background) and three C_3- *Oryza sativa* (Os), *Brachypodium distachyon* (Bd) and *Hordeum vulgare* (Hv) (white background). N' terminal amino acid (a.a.)- $ZmDLK1_1$, C' terminal a.a.- $ZmDLK1_{367}$. Functional domains annotated by horizontal lines- Intercellular trafficking motif (thin black), DOF DNA binding (green), nuclear localisation sequence (orange) and putative transcriptional activation (blue). Conserved amino acid differences between C_3 and C_4 species (red arrows). Amino acid differences between ZmDLK1 and ZmDLK1H are highlighted (red boxes). Degree of a.a. residue conservation shown by colour intensity proportional to Blocks Substitution Matrix (BLOSUM62) values.

All three of the weak plants and two of the more vigorous siblings were tested for transgene insertion using PCR and for transgene expression using RT-PCR (fig.5.7; p.192). The retarded growth phenotype was associated with *ZmDLK1* expression, with no transcript expressed in the vigorous individuals. These data suggest that ubiquitous expression of *ZmDLK1* in rice decreases plant fitness and that the transgene may have an effect on plant regeneration from callus.

5.2.6.2 *ZmDLK1* expression retards plant growth but does not perturb vein patterning in rice

ZmDLK1 expression led to a striking retardation in rice growth (fig. 5.8; p.193). To test whether this phenotype was due to a specific defect in leaf development, the shoot apex was harvested from a 70 days old surviving transgenic plant and transverse sections were compared with a 17 days old wild-type rice apex that had the same number of plastochrons (fig. 5.9; p.194). The *ZmDLK1* expressing plant had slightly fewer procambium initiation sites in plastochron 3, however, the overall vein patterning in developing rice leaves, and the primordium size was similar to wild-type. Leaf ground tissue cell size and number were also similar between the two at each plastochron stage imaged. These results suggest that leaf tissue patterning is not greatly effected by *ZmDLK1* overexpression, however, cell proliferation or expansion at later stages of leaf development are effected, leading to decreased growth rates.

5.2.7 Developmental consequences of *ZmDLK1H* overexpression in rice

5.2.7.1 Transformation of *ZmDLK1H* into rice was not successful

To test the effect of ZmDLK1H activity in a C_3 developmental context, a *ZmUBI*$_{pro}$*:: ZmDLK1H* construct was transformed into rice callus using agrobacterium-mediated gene transfer, and a regeneration step was carried out on cytokinin-rich media. Although this procedure was repeated three times, no T_0 plants regenerated. *ZmDLK1H* over-expression, therefore, appears to inhibit callus regeneration in rice.

5.2.8 Gene silencing of *ZmDLK1H* in the presence of truncated *GRMZM2G171855*

Given that *ZmDLK1H* over-expression is deleterious to rice callus regeneration, it is probable that gene expression levels are tightly regulated during normal development to enable organ growth. To investigate potential post-transcriptional regulation of *ZmDLK1H* transcript levels, the effect of the anti-sense transcript *GRMZM2G171855* on *ZmDLK1H* expression was tested by co-transformation into rice and expression analysis in transgenic plants.

5.2.8.1 A maize genomic DNA fragment was used to co-transform *ZmDLK1H* and △*GRMZM2G171855*

RNAseq data showed that *GRMZM2G171855* is expressed during leaf development, however, the transcription initiation site appears to be mis-annotated in Phytozome. Given that DNA motifs indicative of transcription start sites have been characterised (Deng & Roberts, 2006), manual genomic sequence analysis was used to identify a putative transcription start site on the complementary strand of the *ZmDLK1H* coding DNA sequence (fig.5.10; p.196). These data suggest

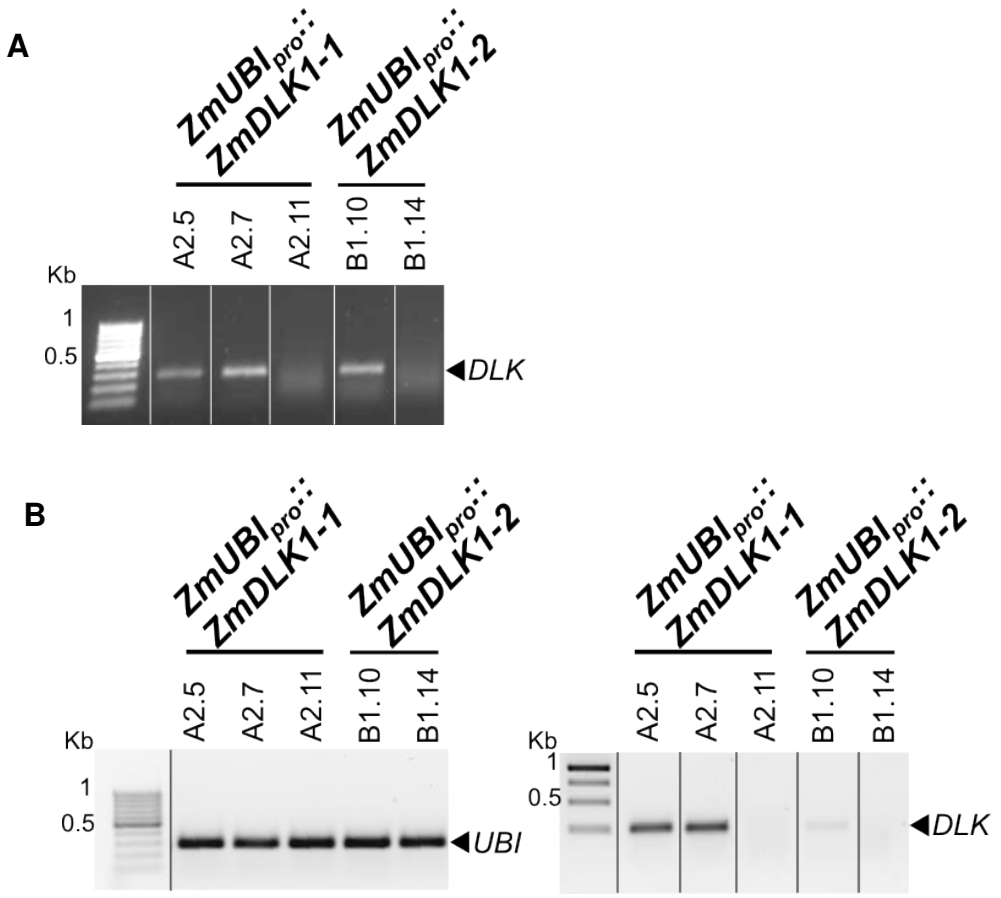


Figure 5.7: Composite of gel images showing *ZmUBI$_{pro}$::ZmDLK1* genotyping and expression analysis. T_1 individuals from three *ZmUBI$_{pro}$::ZmDLK1* lines were analysed for **A)** Transgene insertion by PCR with genomic DNA. **B)** Transgene expression using RT-PCR *OsUBI* amplification was used as a positive control. Each PCR well ID is a unique identifier for one individual which corresponds with the RT-PCR ID. ID code refers to independent transformation events (A2/B1), and the T_1 individual in the family (.#).



Figure 5.8: Constitutive *ZmDLK1* expression retards rice growth. A-B) 70 day old T_1 siblings from the *ZmUBI$_{pro}$::ZmDLK1-1* rice line; transgenic plant with no *ZmDLK1* expression (A) compared with sibling that showed constitutive *ZmDLK1* expression (B). Scale bar- 5cm

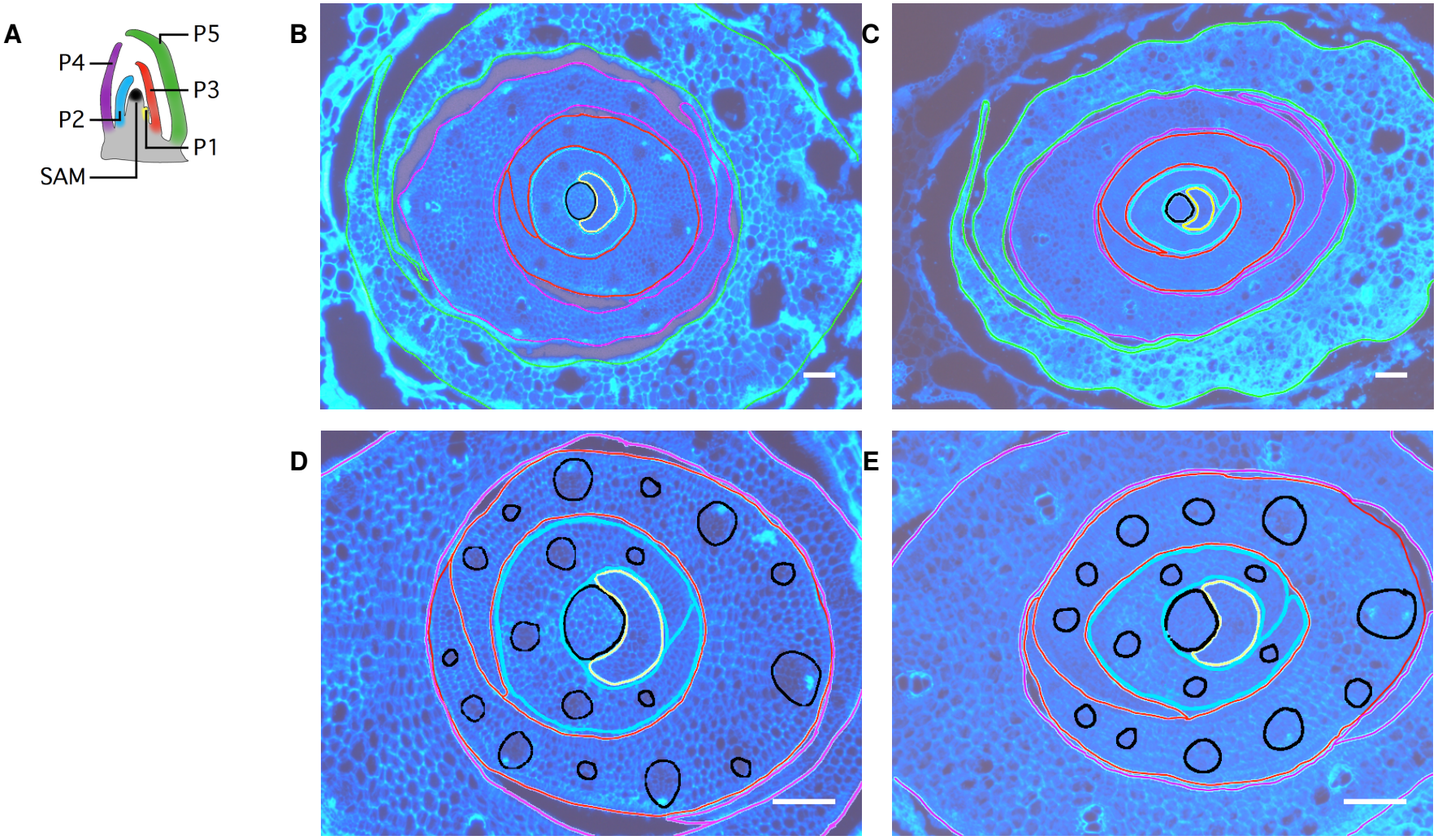


Figure 5.9: *ZmDLK1* transgenic rice plant does not display vein patterning purturbations in leaf primordia. Figure legend is on p.195.

Figure 5.9: *ZmDLK1* transgenic rice plant does not display vein patterning purturbations in leaf primordia. A) Diagram showing longitudinal cross section of monocot shoot apex with colour coded shoot apical meristem (SAM) and plastochrons 1 to 5 (P1-5). **B-E)** UV images of transverse sections of rice shoot apices stained with calcofluor-white comparing vein (black outline) spacing in a 17 day old wild-type and a 70 day old *ZmUBI*$_{pro}$*::ZmDLK1-1* rice plant that expressed *ZmDLK1*. P2 (blue outline)- wild-type (B) and *ZmDLK1* (C), P3 (red outline)- wild-type (D) and *ZmDLK1* (E).

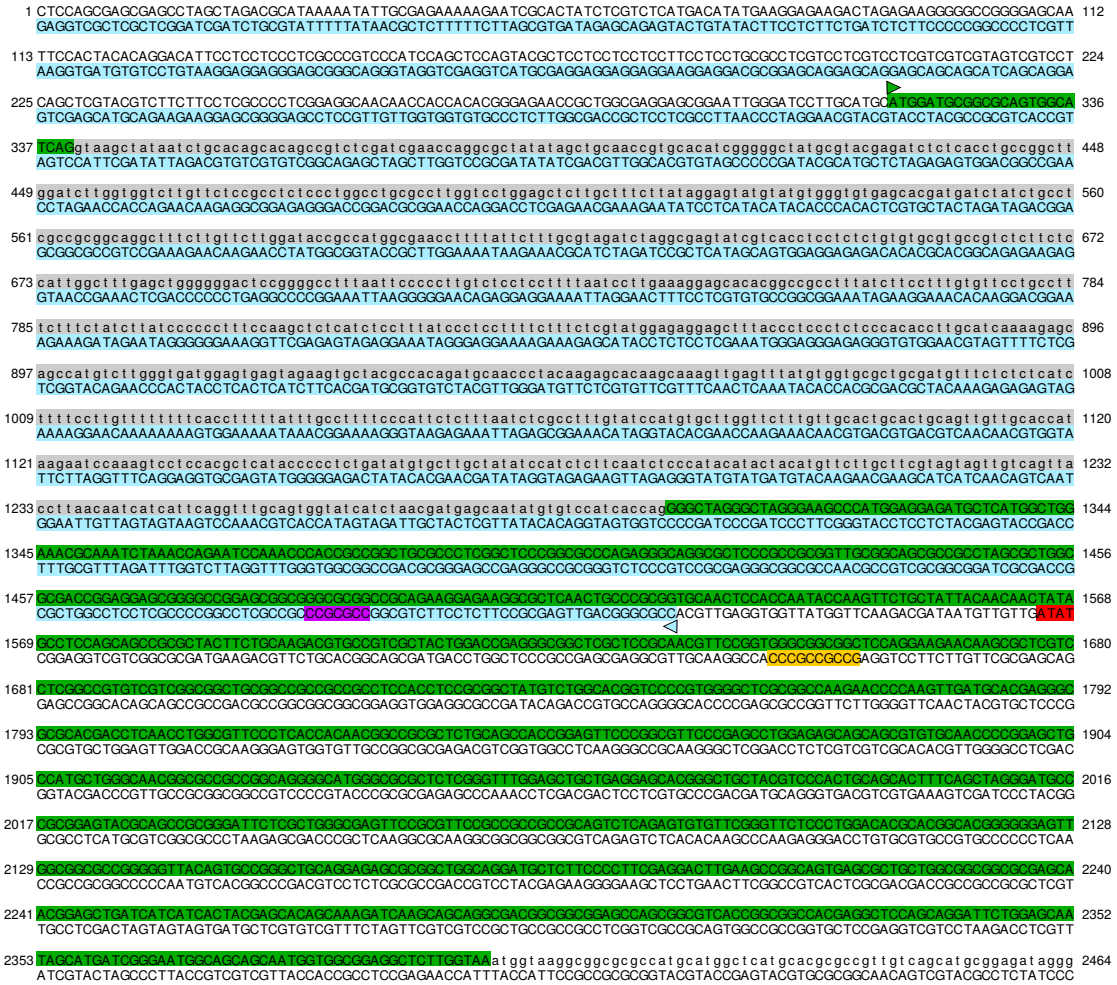


Figure 5.10: Anti-sense RNA transcription start site is predicted in the *ZmDLK1H* exon genomic region. Genomic sequence showing DNA strand coding for *ZmDLK1H* (top strand) and the complementary strand (bottom strand). *ZmDLK1H* coding DNA sequence (CDS) (green) and intron (grey), direction of transcription shown with green arrow. Putative *GRMZM2G171855* transcript (light blue) showing direction of transcription (blue arrow) and annotated putative transcriptional initiation elements- TATA box (red), upstream BRE element (yellow) and downstream BRE element (purple).

that the *GRMZM2G171855* transcript contains approximately 250 bp of complementary sequence to the *ZmDLK1H* mRNA (this region will be termed the anti-sense region) that may regulate *ZmDLK1H* transcript levels.

To test whether the *GRMZM2G171855* anti-sense region can regulate levels of *ZmDLK1H* transcript, a construct was designed that included the anti-sense region plus non-complementary *ZmDLK1H* sequence (the whole second exon was used). The DNA sequence was cloned from genomic DNA into the pVec8 vector in order to utilise the bi-directional nature of the $35S_{pro}$ promoter (Zheng *et al.*, 2007). As such, in transgenic plants, *ZmDLK1H* sequence expression would be driven by the *UBI* promoter and expression of the anti-sense region was activated by the *35S* promoter (fig.5.11; p.198).

5.2.8.2 Co-transformation of *ZmDLK1H* and △*GRMZM2G171855* into rice led to the regeneration of plants through *ZmDLK1H* gene silencing

The *ZmDLK1H*–△*GRMZM2G171855* construct was transformed into rice callus using agrobacterium-mediated gene transfer. T_0 plants regenerated well and transgenic T_0 lines set seed (table 5.3; p.199). Three lines were taken forward for transgene expression analysis, and all three lines showed *ZmDLK1H* gene silencing (fig. 5.12; p.200). These preliminary data indicate that the anti-sense region of *GRMZM2G171855* can prevent *ZmDLK1H* transcript accumulation and thus may be a regulatory mechanism controlling *ZmDLK1H* transcription in plants.

5.2.9 Hormone and light responsive regulatory elements identified in *ZmDLK* putative promoter regions

To further elucidate potential regulators of the two *ZmDLK* genes, 3 kbp nucleotide regions upstream of the translational start sites were analysed for known regulatory elements (REs) using NSITE-PL software (Shahmuradov & Solovyev, 2015) (table

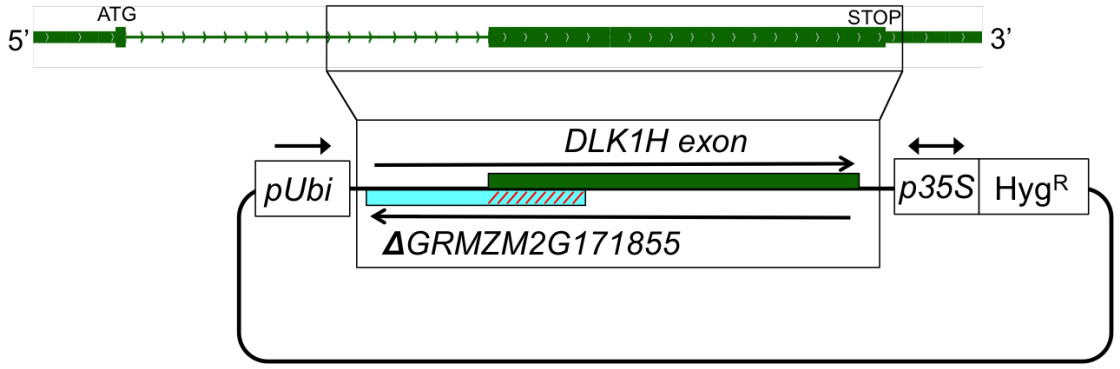


Figure 5.11: Construct to test the effect of *GRMZM2G171855* antisense region on *ZmDLK1H* expression Genomic DNA from the *ZmDLK1H* gene locus was cloned into the pVec8 plasmid backbone. The relationship between the cloned region to the overall *ZmDLK1H* gene model is indicated- the gene model is annotated with the UTRs (thin green box), exons (thick green box) and intron (green line with vertical dashes) and the region cloned for the construct is outlined by a black box. The construct consists of a *DLK1H* exon (thick green box) driven by a ubiquitin gene promoter and a truncated *GRMZM2G171855* region (blue) that includes the anti-sense region (red hashes) transcriptionally activated by the 35S promoter.

5.4; p.202). Strikingly, there was little overlap in REs between the two genes. The majority of *ZmDLK1* motifs were related to light responsiveness and seed development, and binding domains for bZIP and bHLH transcription factors were also identified. In contrast, *ZmDLK1H* REs were related to abscisic acid (ABA)-responsiveness and to the earliest stages of morphogenesis, containing motifs that are also found in homeotic gene promoters. These data indicate that the two proteins may have roles at different stages of plant development.

Given that experimental evidence suggests the anti-sense transcript coded in the *ZmDLK1H* gene locus regulates *ZmDLK1H* expression, the REs in the putative promoter regions of the anti-sense transcripts in both the *ZmDLK1* and *ZmDLK1H* gene loci were also analysed (table 5.5; p.203). Interestingly, both promoter regions contained an RE that is bound by an AP2 protein (a family of proteins that are over-represented in the Kranz dataset) plus a motif that is bound by a bZIP protein and is also found upstream of the *AtARR8* gene (a negative regulator of cytokinin signalling). The promoter region of the anti-sense *ZmDLK1* transcript also contained light and auxin responsive elements.

Line	Transgene	T_0 survival	T_0 seed set
ZmDLK1H-1	Yes	Yes	Good
ZmDLK1H-2	Yes	Yes	Good
ZmDLK1H-3	Yes	Yes	Good
ZmDLK1H-4	Yes	Yes	Good
ZmDLK1H-5	Yes	Yes	Good
ZmDLK1H-6	Yes	Yes	Good
ZmDLK1H-7	Yes	Yes	Good
ZmDLK1H-8	Yes	Yes	Good
ZmDLK1H-9	Yes	Yes	Good

Table 5.3: *ZmUBI$_{pro}$::ZmDLK1H* transformation regeneration summary.

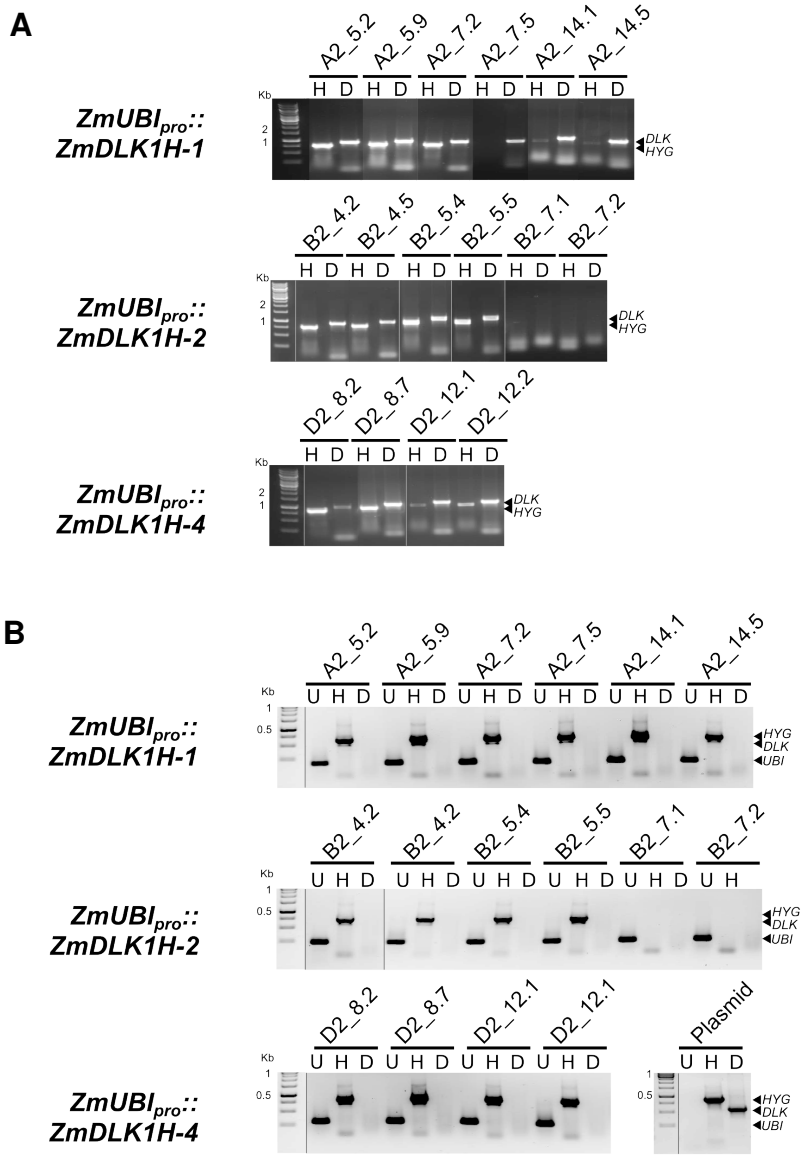


Figure 5.12: Composite of gel images showing *ZmUBI$_{pro}$::ZmDLK1H* genotyping and expression analysis. T_2 individuals from three *ZmUBI$_{pro}$::ZmDLK1H* lines were analysed for- **A)** Transgene insertion by PCR with genomic DNA using amplification of the hygromycin resistance gene (H) and *ZmDLK1H* (D). **B)** Transgene expression using RT-PCR. *OsUBI* amplification was used as a positive control (U), the hygromycin resistance gene (H) and *ZmDLK1H* (D) were amplified to test transgene expression. Each PCR well ID is a unique identifier for one individual which corresponds with the RT-PCR ID. ID code refers to independent transformation events (A2/B2/D2), the T_1 family (_#) and the T_2 individual in the family (.#).

Two of these motifs are bound by Auxin Response Factors (ARFs). Given that many ARFs are repressors, it is possible that high auxin levels stop gene silencing of *ZmDLK1* genes during cellular differentiation in Kranz development. Overall, this analysis indicates that hormone signalling networks may play a large role in regulating *ZmDLK* gene expression.

5.2.10 Sub-functionalisation of *DLK1* and *DLK1H*

To infer whether the gene duplication event in maize led to novel expression patterns for one of the homeologs, the regulatory elements upstream of *LOC_0s02g45200* (*OsDLK1*) were analysed using NSITE-PL software (Shahmuradov & Solovyev, 2015) (table 5.6; p.204). Strikingly, the *OsDLK1* promoter contained a combination of motifs that were only present in either *ZmDLK1* or *ZmDLK1H* promoter regions in maize. These data suggest that the maize gene expression patterns sub-functionalised after the duplication event. However, because no other C_4 monocots display two copies of the *DLK1* gene, it is unlikely that this sub-functionalisation was a key driver during Kranz evolution.

	Motif	Location 5′ of ATG	Regulatory Element	Description
ZmDLK1	GACACGTGTC	1837-1828	ABRE from *AtSUC1*	ABA-responsive
	TCCACGTGGTGGG	2106-2118	OsNAS1-OsIRO2-BS from *OsMTN*	Binding of bHLH protein
	CCACGTGGCC	1774-1765	Box II from *PcCHS*	Binding of bZIP protein
	tTCTTTTTGCAAtTTa	829-844	CCAAT box from *ZmpMS1*	Binding of HMG protein in endosperm
	GCCACGTGGC	1775-1766	PA in tobacco	Binding of TATA box binding protein
	GCCACGTGCC	515-524	G-box 10 from tobacco	Expression enhancer
	AAaCAAtACACAAC	638-651	Box BP-1 from *OsGT3*	Expression in endosperm
	ATGAaTAaATAAgGTAG	667-651	n+e element from *HvC-HORDEIN*	Expression in endopserm
	TCATGCATGT	2341-2332	Z2 from *Zmα22kDaZEIN*	Expression in endosperm
	ATTTTTTTAT	276-285	Fp12/III from *GmCHS8*	Expression in seed
	TGgCACGTGGCA	525-514	G-box 1 from *AtRBCS-1A*	Light responsive, binding of bZIP protein
	AtTcATTTTTATT	662-674	AT-1 from *PsRBCS-3.6*	Light responsive
	AAACTgTAaCATTT	782-795	5′-III from *NtRBCS-3A*	Light responsive
	GCCACGTGGC	1775-1766	G-box 1 from tobacco	Light responsive
	cGGACACGTGTCGg	1826-1839	G-box 1 from *NtZGT*	Light responsive
	GCTGTAGTAAGGTG	2175-2188	UN C1 from *LeRBCS3B*	Light responsive
ZmDLK1H	GCCACGTACA	951-960	A3 from *HvA22*	ABA-responsive
	GCCACGTACc	2733-2742	ABRC1 from *HvA22*	ABA-responsive
	TGCCAaCGG	2764-2772	CE1 from *HvA22*	ABA-responsive
	GCGTaGGGAACCGTGC	565-581	ANT motif	ANT protein binding
	CTACTCTACTg	2010-2020	STY1 BS from *AtYUC4*	Auxin related, expression in apical meristem
	AATCTCCAcCCA	1906-1895	Box-L4 from *DcPAL1*	Binding of MYB protein
	GTACGTGGCG	959-950	Lwt from tobacco	Binding of TATA box binding protein
	cGATcATTTCtTAAC	1286-1272	Regulatory Motif 3 from *GmA2B1a*	Expression in embryo
	CTCTCTCTCTCTCTCT	1434-1449	CT-rich from *CrWRKY1*	Expression in vasculature
	AGAGAGAG	1223-1215	GA-5 from *AtSTK*	Regulation of homeotic genes
	GAGAGAGAGAGAGAGAc	1450-1434	GAGA from *GmGSA1*	Regulation of homeotic genes

Table 5.4: Regulatory elements present in *ZmDLK1* and *ZmDLK1H* putative promoter regions. Plant regulatory elements (REs) present in 3 kbp genomic region upstream (5′) from the translation start site (ATG) for *ZmDLK1* and *ZmDLK1H*. Motif homology to published REs is indicated by capital letters, description of REs- motif name and the gene where it was first identified are annotated using the RegsitePL database (Shahmuradov & Solovyev, 2015).

	Motif	Location 5′ of ATG	Regulatory Element	Description
ZmDLK1	TGTCTC	1546-1541	D5 from *GmGH3*	Auxin responsive
antisense	TGTCTC	1535-1530	D5 from *GmGH3*	Auxin responsive
transcript	tGaCGCCGCCGCCGCCGCC	842-824	GCC box from arabidopsis	Ethylene responsive, binding of AP2 protein
	TaTGACGTGTTT	2198-2187	C/G-box from *AtARR8*	Binding of bZIP protein
	ACTCTcCAGTACTC	2885-2872	TAC box from *NtEAS4*	Binding of a repressor protein in tobacco
	CtCAACGTGGCG	312-301	G-box from *CrCathB3*	Light responsive
	GCACaGTTAAAC	1464-1475	Box 2r from *NtCab-E*	Light responsive
	TgCACGTGGCAT	1576-1587	G box from *PsRBCS-3.6*	Light responsive
	AAgATTTTTATT	2122-2133	AT-1 from *LeRBCS-3A*	Light responsive
	AAAGaGGTGCTG	2716-2705	Alfin1 BS5 from *MsPRP2*	Saline-responsive
	CgAATTAACCAAaaCAT	1727-1743	L-box from *LsRBCS3A*	Unknown
	TaTATAATTTTTT	2346-2334	AT-2a from *PsGS1a*	Unknown
	ATTTATATATaTA	2385-2373	AT-rich ELS from *GmBiP6*	Unknown
	aAAACACGTGATAg	2025-2012	Box 1 from *NtprxC2*	Wounding-response
ZmDLK1H	TGTGACGTGcTT	1306-1317	C/G-box from *AtARR8*	Binding of bZIP protein
antisense	ATCGTGGGAA	2414-2423	BRRE from *AtFLD*	Brassinosteroid responsive
transcript	gGCtGCgGCCGCCGCCGCC	174-192	GCC box from arabidopsis	Ethylene responsive, binding of AP2 protein
	ATTTTTccATgATTAATaAA	1747-1766	ATRE-BS2 from *SrGLB3*	Expression in roots

Table 5.5: Regulatory elements present in *ZmDLK1* and *ZmDLK1H* anti-sense transcript putative promoter regions.

Plant regulatory elements (REs) present in 3 kbp genomic region upstream (5′) from the putative transcription start site for *ZmDLK1* and *ZmDLK1H* anti-sense transcripts. Motif homology to published REs is indicated by capital letters, description of REs- motif name and the gene where it was first identified are annotated using the RegsitePL database (Shahmuradov & Solovyev, 2015).

	Motif	Location 5′ of ATG	Regulatory Element	Description
OsDLK1	ACGTGGCGC	538-530	Em1a from *TaEm*	ABA-responsive
	CaCTGTCACGAT	1632-1621	W-box 4 from *AtABI5*	ABA-responsive
	GCCAAGTGGA	2378-2369	-190 half G-box (core) from *AtADH3*	ABA-responsive
	ACGcCGGGAgGCaTGC	1713-1698	ANT motif	ANT protein binding
	CTCTCcCTtTCTaTCT	169-184	CT-rice from *CrWRKY1*	Expression in vasculature
	TTGATCTAGTtGATtAg	1720-1704	GS3A-F1 from *PsGS3A*	Expression in vasculature
	AATATTTTTATg	228-239	AT-1 from *LeRBCS-3A*	Light responsive
	CAAATATAAG	1067-1076	CArG from *LeACS4*	MADS protein binding
	AGAGAGAGA	192-184	GA-5 from *AtSTK*	Regulation of homeotic genes
	CCCCTCCTC	609-601	C-rich Q from *LeRBCS2*	Unknown
	TaTATAATTTTTT	2606-2618	AT-2a from *PsGS1a*	Unknown
	CCTCTCCTTgaCTC	135-122	CT-element from *PsGS1a*	Unknown

Table 5.6: Regulatory elements present in *OsDLK1* putative promoter region. Plant regulatory elements (REs) present in 3 kbp genomic region upstream (5′) from the translation start site (ATG) for *OsDLK1*. Motif homology to published REs is indicated by capital letters, description of REs- motif name and the gene where it was first identified are annotated using the RegsitePL database (Shahmuradov & Solovyev, 2015).

5.3 Discussion

5.3.1 ZmDLK1 is a negative regulator of plant growth in rice

Transgenic rice plants with constitutive *ZmDLK1* expression showed decreased growth rates and reduced viability. Given that no veination patterning perturbations were identified, it is postulated that ZmDLK1 can effect the rate of cell proliferation and/or expansion but is not sufficient to alter cell differentiation processes. DLK1H has a very similar amino acid sequence to ZmDLK1, especially in the DNA binding domain. It is therefore probable that the gene targets of the two proteins are conserved and that the effect of ZmDLK1H in rice also acts to restrict organ growth during early leaf development.

5.3.2 *ZmDLK1* and *ZmDLK1H* transcript levels regulated by anti-sense RNA

Given that over-expression of *ZmDLK1* was shown to be detrimental to plant growth and survival, there may have been selective pressure for finely tuned control of *DLK* gene expression. In maize, the two *DLK* gene loci investigated contain anti-sense sequences that are co-expressed with the genes during leaf development. Notably, knock-down of *ZmDLK1H* expression was observed when both *ZmDLK1H* and the anti-sense transcript were expressed in rice. This observation suggests that during maize leaf development, *DLK* activity may be regulated through anti-sense RNA based post-transcriptional mechanisms.

5.3.3 Amino acid differences between ZmDLK1 and ZmDLK1H modulate effects on target genes during leaf development

Both *ZmDLK1H* and *ZmDLK1* over-expression decreased rice callus regeneration efficiency but the effect of *ZmDLK1H* was more severe. Interestingly, the two proteins have highly conserved amino acid sequences, however, several motifs are missing from the ZmDLK1 sequence. For example, ZmDLK1H has a motif in the putative transcription activation domain that consists of six hydrophobic residues; this motif is not present in ZmDLK1. The residue differences between the two proteins may modulate transcriptional activation of target genes, and this may explain the differences in regeneration efficiency. These amino acid differences may also modulate target gene transcription during Kranz development. Given that the two *ZmDLK* genes have different promoter sequences, *DLK1* and *DLK1H* may be part of a regulatory system that can fine-tune target gene expression during maize leaf development.

5.3.4 *ZmDLK1* and *ZmDLK1H* are expressed at different times during leaf development

The regulatory elements in the putative promoter region of *ZmDLK1H* indicate that the protein may be involved in regulating plant growth in the early stages of organogenesis in maize, before procambium initiation. Given that no regeneration was observed when *ZmDLK1H* was constitutively expressed in calli grown on cytokinin-rich media, it is possible that the protein acts to repress cell proliferation in ground tissues during shoot organogenesis. Given that the barley ortholog HvSAD represses α-*amylase* expression in the endosperm (Diaz *et al.*, 2005), and that amylases are important regulators of sugar-signalling, a network that

is known to regulate cell division (reviewed in Gibson, 2005), it is possible that ZmDLK1H represses cell division during leaf development by modulating sugar levels. *ZmDLK1* regulatory elements are associated with different pathways. For example, the promoter contains several regulatory elements that are present upstream of RUBISCO small subunit genes, and that are known to be regulated by light-signalling. It is therefore likely that ZmDLK1 regulates gene expression at later stages of leaf development than ZmDLK1H. Interestingly, ZmDLK1 is preferentially expressed in bundle sheath cells and there is some evidence that the closely related poplar DLK protein PtWALLDOF up-regulates cell wall biosynthesis (Arruda & Gerhardt, 2016). As such, ZmDLK1 may have a role in the suberisation of bundle sheath cells during Kranz development.

5.3.5 Hormone signalling regulates *ZmDLK* transcript levels

Both *ZmDLK* genes and the putative anti-sense transcript promoter regions contain motifs that have been implicated in hormone responses. For example, the *ZmDLK1H* promoter region contains a regulatory element that has been shown to bind STYLISH1 (STY1), a protein that regulates auxin biosynthesis in arabidopsis apical meristem tissue. The regulation of ZmDLK1 also appears to be linked to auxin signalling because the anti-sense transcript contains auxin responsive elements. *ZmDLK1* transcript levels may therefore be down-regulated by auxin. Interestingly, ABA and ethylene responsive elements are also seen in the promoter regions, with ABA-related motifs in the *ZmDLK1H* promoter and ethylene-related elements in the complementary anti-sense transcript promoter. ABA and ethylene have been shown to regulate shoot growth during water-stress (reviewed in Sharp & LeNoble, 2002), and it is therefore possible that ZmDLK1H may play a role in regulating shoot growth under drought conditions.

5.3.6 Mis-calculation of transcript levels for *ZmDLK* genes during leaf development due to anti-sense transcripts

The manual analysis of RNAseq data from Wang *et al.* (2013b) showed that a high proportion of reads that were mapped to the *ZmDLK1* and *ZmDLK1H* gene loci in the P1-P5 samples were from the intronic regions, and likely to be anti-sense transcripts from the complementary strand. This means that transcript levels quantified for the *ZmDLK* genes are likely to be inaccurate during early stages of leaf development. Given that ZmDLK1 inhibits rice shoot growth, it is possible that the *DLK* genes are actually downregulated during the early stages of Kranz development through upregulation of the anti-sense transcripts. This could not be detected in the transcriptomic analysis as it was not possible to analyse from which DNA strand the read originated. To confirm transcript levels, RT-PCR analysis of RNA from the different leaf primordia could be conducted, using primers designed to amplify across the intronic region.

5.4 Future perspectives

5.4.1 The effect of ZmDLK1 and ZmDLK1H on rice bundle sheath development

Given that constitutive expression of *ZmDLK1* is deleterious to rice plant growth, the role of ZmDLK1 in bundle sheath development could not be tested in this study. RNAseq data from maize bundle sheath and mesophyll cells (Li *et al.*, 2010) shows that there is very little *ZmDLK* anti-sense mRNA in bundle sheath cells (data not shown). As such, differential expression levels between the cells types are likely to be an accurate representation of 'functional' transcript levels. To test the effect of *ZmDLK1* expression in C_3 BS cells, the promoter sequence from the *Zoysia*

japonica Phosphoenolpyruvate carboxykinase (*ZjPCK*$_{pro}$) gene can be used to drive BS specific expression of *ZmDLK1*. In this regard, preliminary experiments show that rice plants transformed with *ZjPCK*$_{pro}$*::ZmDLK1* or *ZjPCK*$_{pro}$*::ZmDLK1H* regenerate well and show good growth rates. Bundle sheath morphology and cell wall characteristics need to be analysed in the T_1 plants to investigate whether DLK1 proteins can condition Kranz-traits in rice leaves.

5.4.2 Conserved DLK protein role in eudicots and monocots during Kranz development

A *DLK* gene ortholog was up-regulated in the later stages of the eudicot *Gynandropsis gynandra* Kranz development (Külahoglu *et al.*, 2014; Huang *et al.*, 2016) and phylogenetic analysis of the *DLK* clade inferred that gene expansion occurred in both monocot and eudicot species. It is therefore possible that the role of *DLK* genes in Kranz development is conserved between the two plant groups. Given that *AtDLK* gene expression has been shown to localise to vasculature and bundle sheath cells, it is possible that there is a conserved role in, for example, the suberisation of bundle sheath cells in C_4 monocots and eudicots. This can be tested by bundle-sheath specific expression of *ZmDLK* genes in arabidopsis.

6

General Discussion

6.1 Overview

Convergent evolution of leaf morphology gave rise to Kranz anatomy on over 60 independent occasions during angiosperm evolution (Sage *et al.*, 2011a). Kranz is a key feature in the majority of C_4 photosynthetic plants and these plants show increased radiation-use efficiency and drought resistance, features that are highly valued in domesticated crops due to the resulting higher yields (reviewed in Sage & Zhu, 2011). This thesis aimed to further our understanding of the developmental mechanisms that underpin Kranz anatomy by characterising candidate Kranz regulators using *in-situ* hybridisation and constitutive gene expression experiments in rice.

In-situ hybridisation experiments led to the refinement of the Kranz development model proposed by Langdale & Nelson (1991) and Fouracre *et al.* (2014). The results showed that the spatio-temporal expression of putative Kranz regulators correlated with key developmental stages; *ZmMIL2* transcripts localised to the cell layer from which all procambium originates, *ZmSHR2*, *ZmMIL1* and *ZmANT1* transcripts accumulated at the very earliest stage of procambium initiation and *ZmDOT1* expression patterns suggest that it may be an inhibitor of vein

initiation. Later in Kranz development, bundle sheath (BS) and mesophyll (M) cell differentiation is correlated with the expression patterns of *SHORTROOT* pathway genes. *ZmSHR* and *ZmSCR* transcript accumulation profiles differ depending on the developmental stage of the tissue and are potentially activated by *IDD* gene function; BS differentiation is correlated with the presence of *ZmRVN1* transcripts and M cell differentiation with *ZmNKD1* transcripts. Expression patterns of gene orthologs in rice were found to differ from maize, suggesting that gene regulatory changes were a factor in the evolution of Kranz anatomy.

Constitutive expression experiments tested the role of ZmANT1 in rice procambium initiation. Surprisingly, there were no effects on vein initiation patterns, but *ZmANT1* over-expression did restrict cell divisions in the leaf ground tissue through a pathway that was antagonistic to cytokinin signalling. These data suggest that changes in auxin and cytokinin homeostasis during leaf development may have occurred during the evolution of Kranz anatomy.

The gene targets of *ZmANT1* were predicted *in silico* and a putative Kranz regulator was identified as an ANT target gene. The role of this gene (*ZmDLK1H*) and the homeolog (*ZmDLK1*) in leaf development were investigated using constitutive expression in rice. Over-expression was detrimental to callus regeneration and plant growth. Although there were no perturbations to vein initiation patterns, ZmDLK1 did restrict the cell divisions and/or expansion. Analysis of regulatory elements and transcriptome data indicated that the expression patterns of the two homeologs were sub-functionalised for activity in apical meristems and in photosynthetic cells, respectively, and that *ZmDLK1* may have a role in bundle sheath differentiation during Kranz development. An ortholog of this gene was also identified as a putative eudicot Kranz regulator, and thus the recruitment of this gene to Kranz development may have occurred during the evolution of C_4 photosynthesis in both eudicots and monocots.

This discussion will address the implications of the experimental results on our

current understanding of monocot leaf development, and highlight experimental approaches that can be used to further characterise the regulatory networks involved in Kranz anatomy. The section will conclude by discussing the implications of this work for engineering Kranz anatomy in the C_3 developmental context.

6.2 *In situ* hybridisation and constitutive expression results have progressed understanding of monocot leaf development

6.2.1 Regulatory mechanisms underpinning leaf development in eudicots are not conserved in C_3 monocots

The majority of our understanding of leaf development and procambium initiation arises from studies in eudicots, especially arabidopsis. However, the results obtained in this thesis suggest that the regulation of leaf development in monocots, particularly in C_3 plants, may differ substantially from that in eudicots. This was particularly apparent in terms of *ZmANT1* function during leaf development in transgenic arabidopsis and rice plants in that the effects of ZmANT1 on ground tissue proliferation were opposite. Different expression patterns of gene orthologs known to be involved in eudicot procambium initiation were seen in monocots; in arabidopsis, procambium markers are first expressed in cells adjacent to existing procambium, however, in monocots, for the the markers tested, there are always cells between procambium initiation sites. These results suggest that although the genes involved in vein development may be conserved between eudicots and monocots, the regulation of these genes is not necessarily the same. As such, characterisation of gene function in monocots is essential to understand the development of vein patterning in C_3 and C_4 contexts.

6.2.2 Differences exist between C_3 and C_4 ground tissue proliferation

Analysis of sequential shoot apex sections showed that cell proliferation is restricted to a region between mature vasculature and initiating procambium in rice, however, proliferation occurs across the whole leaf blade in maize. There are differences in gene expression patterns between maize and rice that may contribute to these cell division differences. For example, *ZmMIL2*, *ZmMIL1* and *ZmSHR1* are expressed in maize procambium whereas the expression of orthologs in rice is only seen in mature vasculature. Therefore, these gene products may play a role in promoting cell division in adjacent plant tissue. *DLK* genes are likely to play a direct role in regulating cell proliferation or expansion as *ZmDLK1* over-expression inhibited leaf growth in rice. Analysis of transcriptome data suggests that *ZmDLK1* and *ZmDLK1H* could be down-regulated during Kranz development through expression of anti-sense RNA. Significantly, in rice leaf development very few putative *OsDLK* anti-sense transcripts were identified, and therefore ground tissue proliferation may be restricted in rice leaves through OsDLK activity. *ZmANT1* also inhibited cell division in rice and was antagonistic to cytokinin-signalling. Given that cytokinin is known to enhance cell division in ground tissue (Giulini *et al.*, 2004; Kurakawa *et al.*, 2007), it is likely that the regulation of auxin and cytokinin homeostasis differs between C_3 and C_4 tissues. Interestingly, expression of a predicted ZmANT1 target *IAA-amido synthetase* is upregulated during Kranz development and this may play a role in promoting ground tissue cell division in the C_4 context by sequestering free-auxin. Overall, the regulation of auxin-cytokinin homoeostasis and the subsequent effects on gene expression could explain the different patterns of ground tissue proliferation in rice and maize.

6.2.3 DOT1 could be a key regulator of vein density in monocots

DOT1 transcript levels are negatively correlated with procambium initiation in both rice and maize. *DOT1* transcripts are downregulated during early vein development in maize, but persist to a much later developmental stage in rice veins. Given that procambium initiation in adjacent ground tissues is only seen once *DOT1* expression is downregulated, the data suggest that this gene could be a key regulator of vein initiation. There has been no functional characterisation of *DOT1* genes in monocots or eudicots, and regulatory pathways are not known. As no perturbations in vein patterning were observed with *ZmANT1* or *ZmDLK1* constitutive expression in rice, and *ZmDOT1* was not identified as a putative ANT target gene, the *ANT* developmental pathway is probably not directly involved with the regulation of *ZmDOT1* expression during Kranz development and further work is required to elucidate the DOT1 regulatory network.

6.2.4 SHR and SCR are not sufficient for bundle sheath and mesophyll cell differentiation in C_4 tissues

The *SHORTROOT* developmental pathway has been predicted to be a key regulator of cell differentiation during Kranz development (Slewinski *et al.*, 2012; Fouracre *et al.*, 2014), and *in-situ* hybridisation data presented here support this theory. However, given that the same GRAS proteins (encoded by *ZmSHR* and *ZmSCR* genes) are localised to both developing BS and M cells, the genes can not be sufficient to pattern the two distinct cell types. Interestingly, *ZmDLK1* expression is localised to bundle sheath cells in mature Kranz anatomy, and the promoter region contains an AATTT motif which is a regulatory element that is important for GRAS protein binding (Hirsch *et al.*, 2009). However, there are

also motifs for bZIP and bHLH transcription factor (TF) binding and this could indicate that other tissue-specific TFs have a role in promoting bundle sheath differentiation. These data show that different experimental approaches need to be considered to understand the role of transcription factors in cell differentiation, and more generally in the initiation of veins during Kranz development.

6.3 Further experimental approaches to investigate Kranz development

6.3.1 Transgenic approaches when ectopic expression is detrimental to plant fitness

Constitutive expression of *ZmANT1* and the *ZmDLK* genes in rice led to low callus regeneration efficiency and poor plant vigour. Given this observation, it is likely that there was selective pressure for transformation events in which the transgene was expressed at low levels, thus not all the developmental effects of the transgenes were seen. Further, the low rate of plant survival meant that small sample sizes had to be used for data collection, limiting the strength of statistical analysis. To resolve this issue, a different approach to generating transgenic plants must be taken. Preliminary results using the bundle sheath specific promoter $ZjPCK_{pro}$ have shown that BS-specific expression gives better rates of callus regeneration for *ZmDLK* transgenic plants, but that $ZjPCK_{pro}$*::ZmANT1* is also detrimental to plant survival (results not shown). Unfortunately, there is a limitation in this approach, even when plant regeneration is efficient, as only a narrow developmental window (a single cell-type in this case) can be assessed. To characterise all the effects of the candidate regulators on rice leaf development, a chemical-inducible system could be used to limit the timing of transgene expression (reviewed in

Moore *et al.*, 2006). The glucocorticoid dexamethasone (DEX)-inducible system works well in rice (Ouwerkerk *et al.*, 2001), and preliminary data show that plant regeneration and survival rates are high for DEX-*ZmANT1* and DEX-*ZmDLK* transgenic lines (data not shown). In summary, chemical-inducible expression could allow characterisation of transgene effects at different developmental stages, even though ectopic expression of the genes are detrimental to plant fitness.

6.3.2 The effect of transcriptional complexes on Kranz development in C_4 species

Constitutive over-expression of several candidate Kranz regulators did not perturb leaf development in rice. For example, the *ZmMIL1*, *ZmMIL2* and *ZmSCR1* transgenic rice lines did not show any differences compared to wild-type (P. Wang and T. Hughes, personal communication). There is evidence that the orthologs of these genes act in transcriptional factor complexes (Royo *et al.*, 2009; Cui *et al.*, 2007; Cui & Benfey, 2009), and therefore ectopic expression in rice may not give Kranz-like traits due to the absence of interacting proteins. The developmental effects of these candidate Kranz regulators may be more clearly seen if the gene is mutagenised within a C_4 context; in this case, the transcriptional complexes will be less effective at activating downstream gene targets, thereby perturbing Kranz development. Insertional mutant libraries are available for maize (Cowperthwaite *et al.*, 2002; McCarty *et al.*, 2005; Vollbrecht *et al.*, 2010; Williams-Carrier *et al.*, 2010) and sorghum (Antre *et al.*, 2013; Jiao *et al.*, 2016), and these mutaganised C_4 plant populations could be used to identify and characterise the effect of candidate Kranz regulator knock-down.

RNAi silencing may also give valuable information about the role of candidate regulators in C_4 species. This approach would provide transgenic lines displaying varying levels of gene expression, so could be effective if complete suppression of

gene activity is lethal. RNAi has been successful in maize (Cigan *et al.*, 2005) and *Setaria viridis* (a potential model C_4 species) (T. Brutnell, personal communication), so could be used to study the effect of different candidate regulator expression levels on Kranz development.

6.3.3 Identifying Kranz regulator developmental networks

If ZmMIL and ZmSCR proteins regulate Kranz development in transcriptional complexes, the interacting partners could be identified using biochemical approaches (reviewed in Fukao, 2012). *In vitro* bait-prey assays using purified candidate regulator proteins as bait, with protein extracts from developing maize leaves followed by mass-spectrometry (MS) may reveal the transcriptional complexes present during Kranz development. Yeast-two hybrid experiments with co-expressed genes such as *ZmANT1*, *ZmSHR2* and *ZmMIL1* could also be used to identify protein-protein interactions that may operate as a transcriptional complex.

Mapping the regulatory network for the candidate genes could identify key upstream regulators of Kranz anatomy. DNA promoter regions from candidate genes can be used as bait during *in vitro* bait-prey assays, and transcriptional regulators identified using protein extracts from developing maize leaves. This approach may reveal the earliest regulators of Kranz development.

6.3.4 Auxin and cytokinin homeostasis during C_3 and C_4 leaf development

Given the key roles of auxin and cytokinin in controlling vascular patterning and development in plants (reviewed in Chandler & Werr, 2015), it is likely that changes in hormone homeostasis during leaf development were involved in the evolution of Kranz anatomy. Identifying the differences in auxin and cytokinin distribution

during C_3 and C_4 leaf development may provide information about the earliest stages of Kranz development. Hormone distribution can be visualised through the use of reporter genes, where hormone responsive promoter elements are linked to genes encoding reporter proteins such as GFP (green fluorescent protein) or GUS (β-glucuronidase). Auxin distribution has been extensively studied using the DR5 system, a synthetic promoter based on auxin response elements (Ulmasov *et al.*, 1997), and a cytokinin-responsive system based on type B-response regulator (Type B-ARR) promoter motifs has also been developed (Zürcher *et al.*, 2013). The DR5 system has been implemented in rice and maize (Scarpella, 2003; Gallavotti *et al.*, 2008) and was successfully used to investigate auxin maxima during maize leaf sheath development (Johnston *et al.*, 2015). In contrast, the cytokinin system has only been tested in arabidopsis plants and maize protoplasts. If cytokinin and auxin distribution differences are identified between C_3 and C_4 leaf development, this knowledge could be used to modulate hormone signalling in the C_3 context to give C_4-like patterns, which may enable the engineering of Kranz anatomy in C_3 plants.

6.4 Research implications for engineering Kranz anatomy in C_3 species

Engineering C_4 photosynthesis into C_3 plants could give crop varieties that are higher yielding and more drought resistant. Existing C_3 leaf anatomy could be used to underpin C_4-like photosynthesis, however this is unlikely to give yield improvements over conventional varieties because Kranz anatomy is required for the highest photosynthetic efficiency (Sage & Zhu, 2011). The work presented in this thesis has direct implications for engineering Kranz into C_3 species in that it has identified and refined the potential roles of key candidate regulators in

Kranz development and has elucidated differences in early leaf development in C_3 and C_4 plants.

6.4.1 Modulating procambium initiation competency

A subset of candidate Kranz regulators have been implicated in Kranz procambium initiation through *in situ* hybridisation. Modulating the spatio-temporal expression patterns of these genes in C_3 leaves to give C_4-like patterns may be key to the engineering of Kranz anatomy in crops such as rice. Further, given that auxin-cytokinin homoeostasis is influenced by the vascular development regulator ANT1, it is possible that altering hormone distribution patterns in C_3 plants may increase competency to initiate procambium in developing leaf tissues. Overall, the evidence presented in this thesis gives new targets for altering procambium initiation in C_3 leaves.

6.4.2 Gene stacking for transcriptional complexes

There is a discrepancy between the significance of candidate genes in Kranz development predicted using *in situ* hybridisation and the effect of those genes in rice constitutive expression studies. This is likely due to the role of transcriptional complexes during Kranz development. The results presented in this thesis can be used to guide protein-protein interaction studies and gene stacking strategies to enable functional transcriptional complexes to be engineered into the C_3 contex.

In summary, this body of work aimed to address the role of candidate transcription factors in Kranz development. The results have identified key differences between C_3 and C_4 leaf development that may have been fundamental to the evolution of Kranz anatomy. The findings can be used to guide further research into Kranz anatomy, and are important for engineering C_4 photosynthesis into C_3 species to increase crop yields.

Appendices

A

Phylogenetic inference of rice orthologs to candidate Kranz regulators

A.1 ANT1 phylogeny

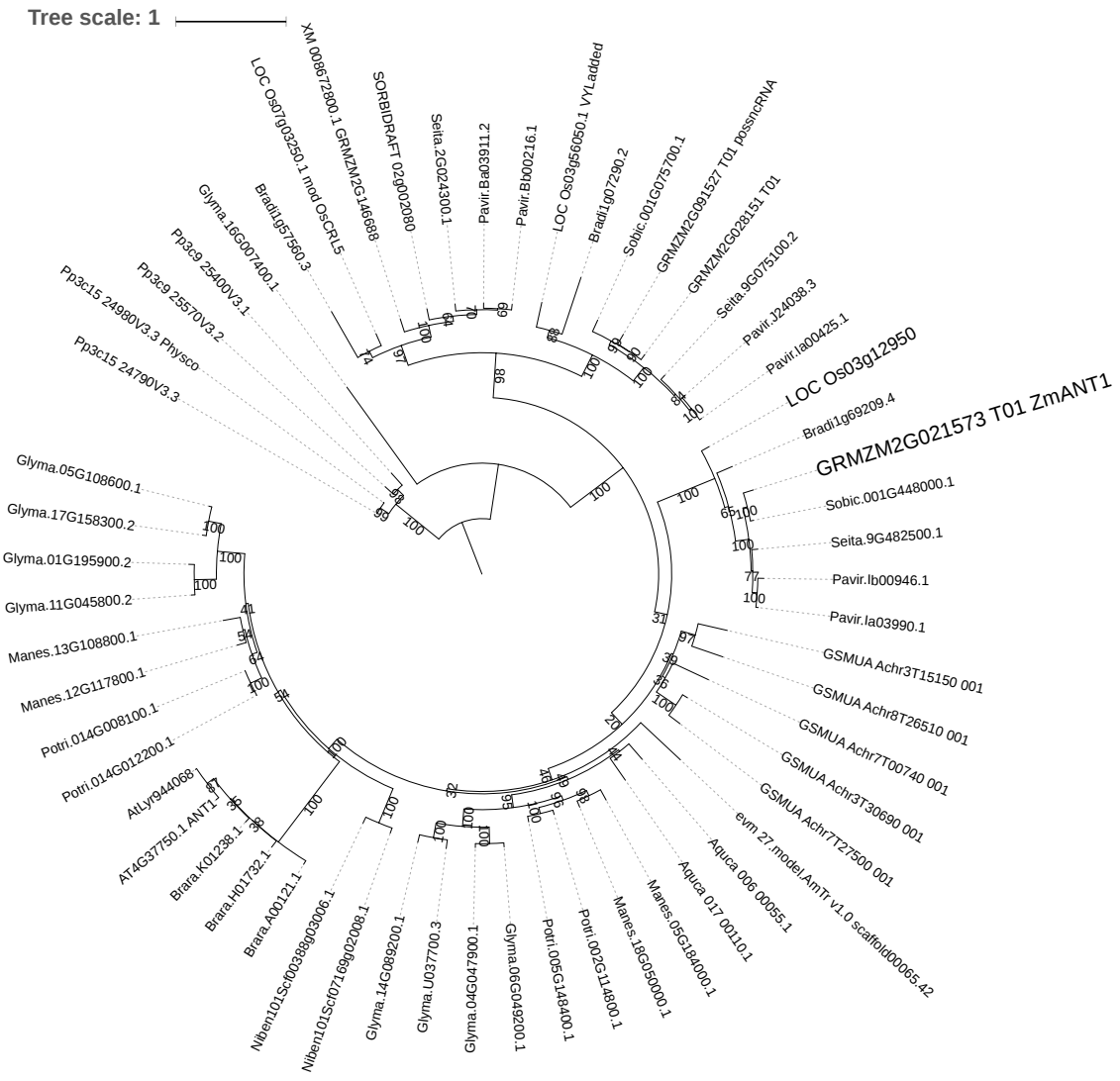


Figure A.1: Phylogenetic tree of ANT1 proteins Maximum-likelihood inference of phylogenetic relationships between the AINTEGUMENTA (ANT1) proteins in monocot and eudicot species. Branch length measured in number of amino acid substitutions per site, bootstrap values shown for each node. Putative Kranz candidates (enlarged "GRMZM" labels) and rice (enlarged "LOC" bold labels) are highlighted.

A.2 CRF1 phylogeny

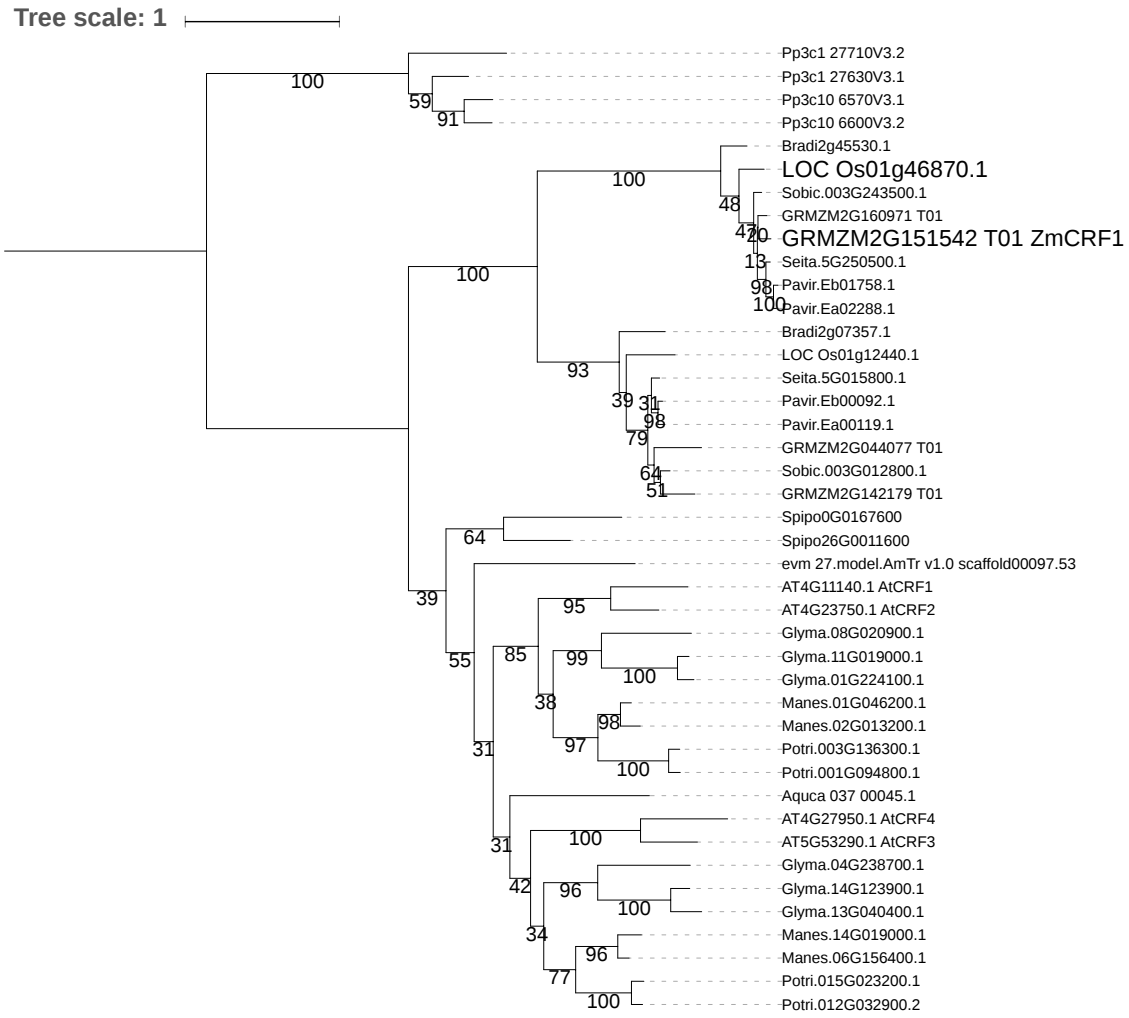


Figure A.2: Phylogenetic tree of CRF1 proteins Maximum-likelihood inference of phylogenetic relationships between the CYTOKININ RESPONSE FACTOR 1 (CRF1) proteins in monocot and eudicot species. Branch length measured in number of amino acid substitutions per site, bootstrap values shown for each node. Putative Kranz candidates (enlarged "GRMZM" labels) and rice (enlarged "LOC" bold labels) are highlighted.

A.3 DOT1 phylogeny

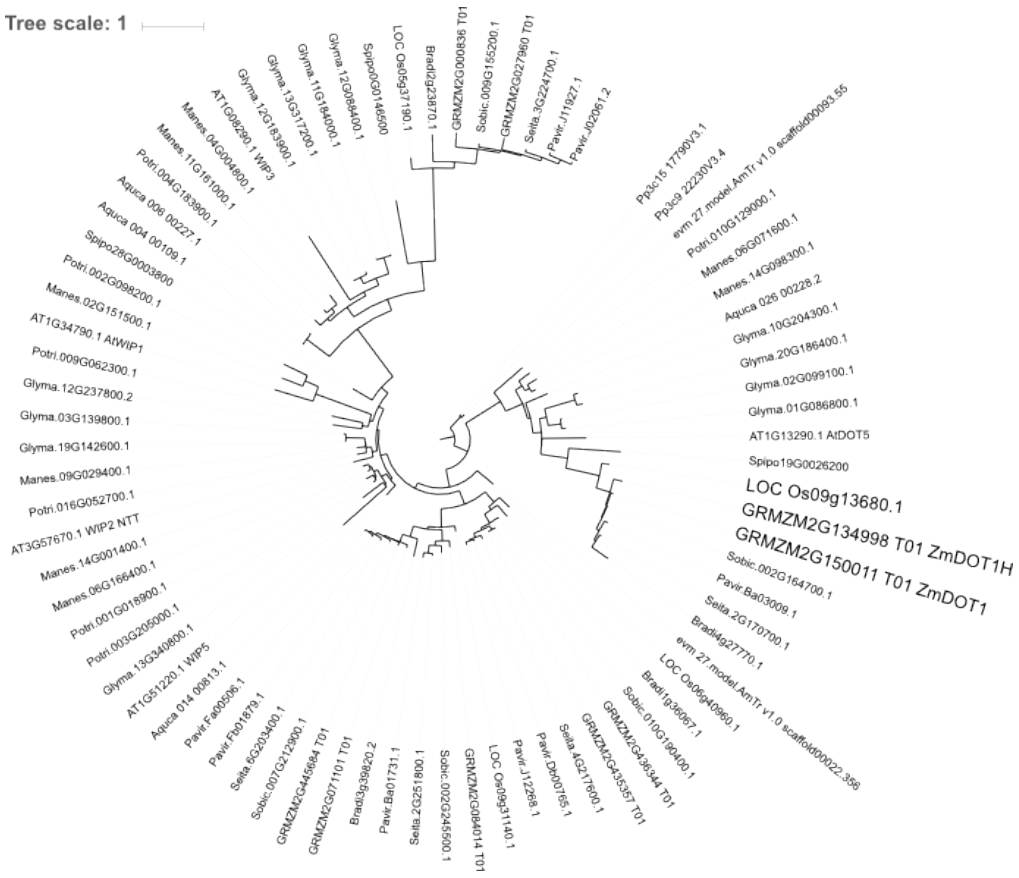


Figure A.3: Phylogenetic tree of DOT1 proteins Maximum-likelihood inference of phylogenetic relationships between the DEFECTIVELY ORGANISED TRIBUTARIES 1 (ZmDOT1/AtDOT5) proteins in monocot and eudicot species. Branch length measured in number of amino acid substitutions per site, bootstrap values shown for each node. Putative Kranz candidates (enlarged "GRMZM" labels) and rice (enlarged "LOC" bold labels) are highlighted.

A.4 NKD1 phylogeny

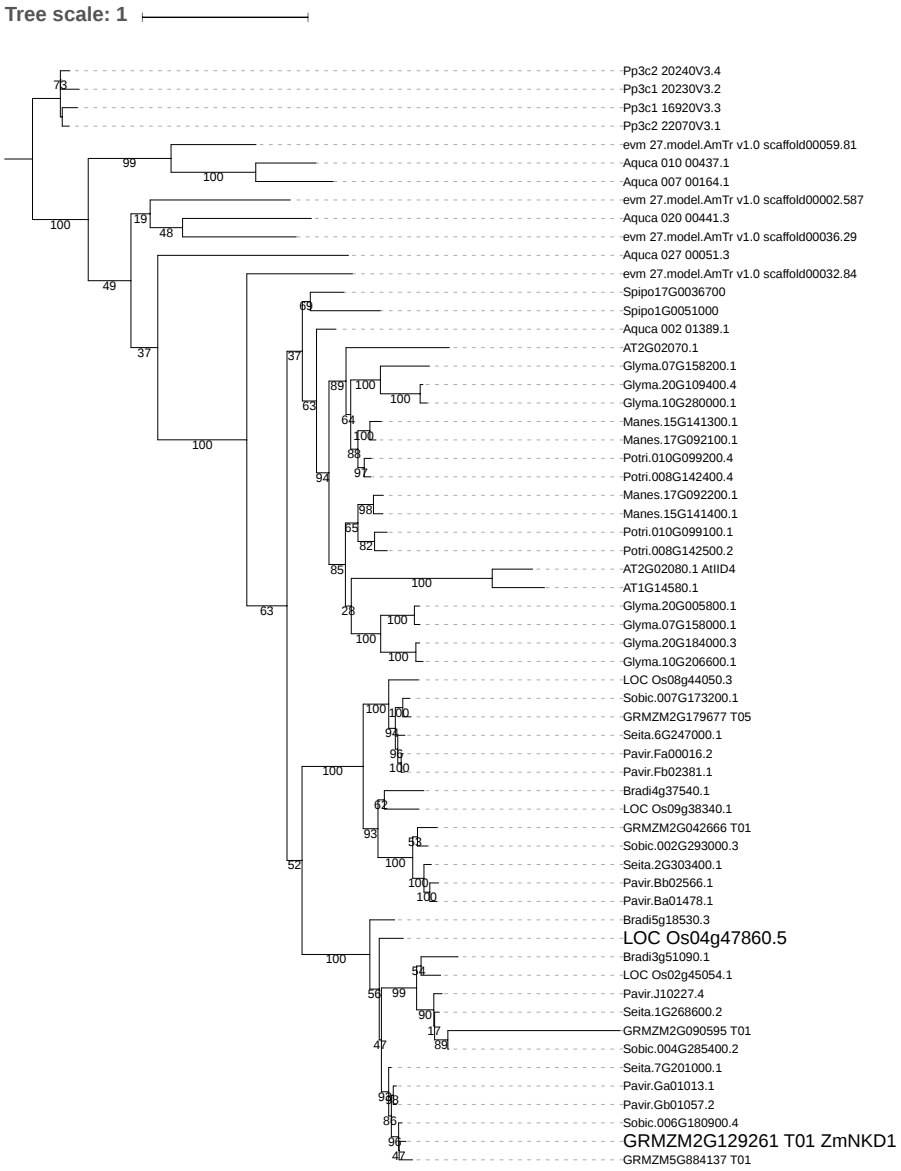


Figure A.4: Phylogenetic tree of NKD1 proteins Maximum-likelihood inference of phylogenetic relationships between the NAKED (NKD) proteins in monocot and eudicot species. Branch length measured in number of amino acid substitutions per site, bootstrap values shown for each node. Putative Kranz candidates (enlarged "GRMZM" labels) and rice (enlarged "LOC" bold labels) are highlighted.

A.5 MIL phylogeny

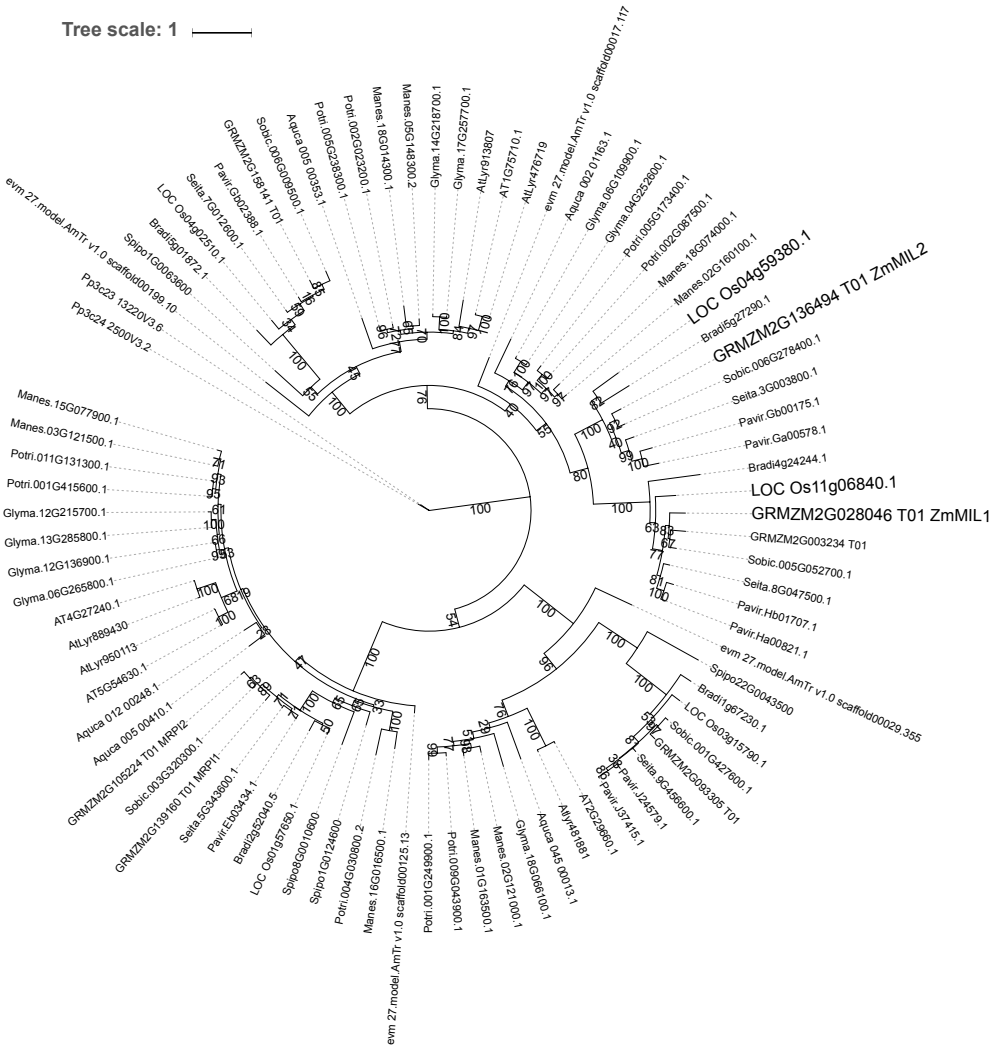


Figure A.5: Phylogenetic tree of MIL proteins Maximum-likelihood inference of phylogenetic relationships between the MRPI-like (MIL) proteins in monocot and eudicot species. Branch length measured in number of amino acid substitutions per site, bootstrap values shown for each node. Putative Kranz candidates (enlarged "GRMZM" labels) and rice (enlarged "LOC" bold labels) are highlighted.

A.6 RVN1 phylogeny

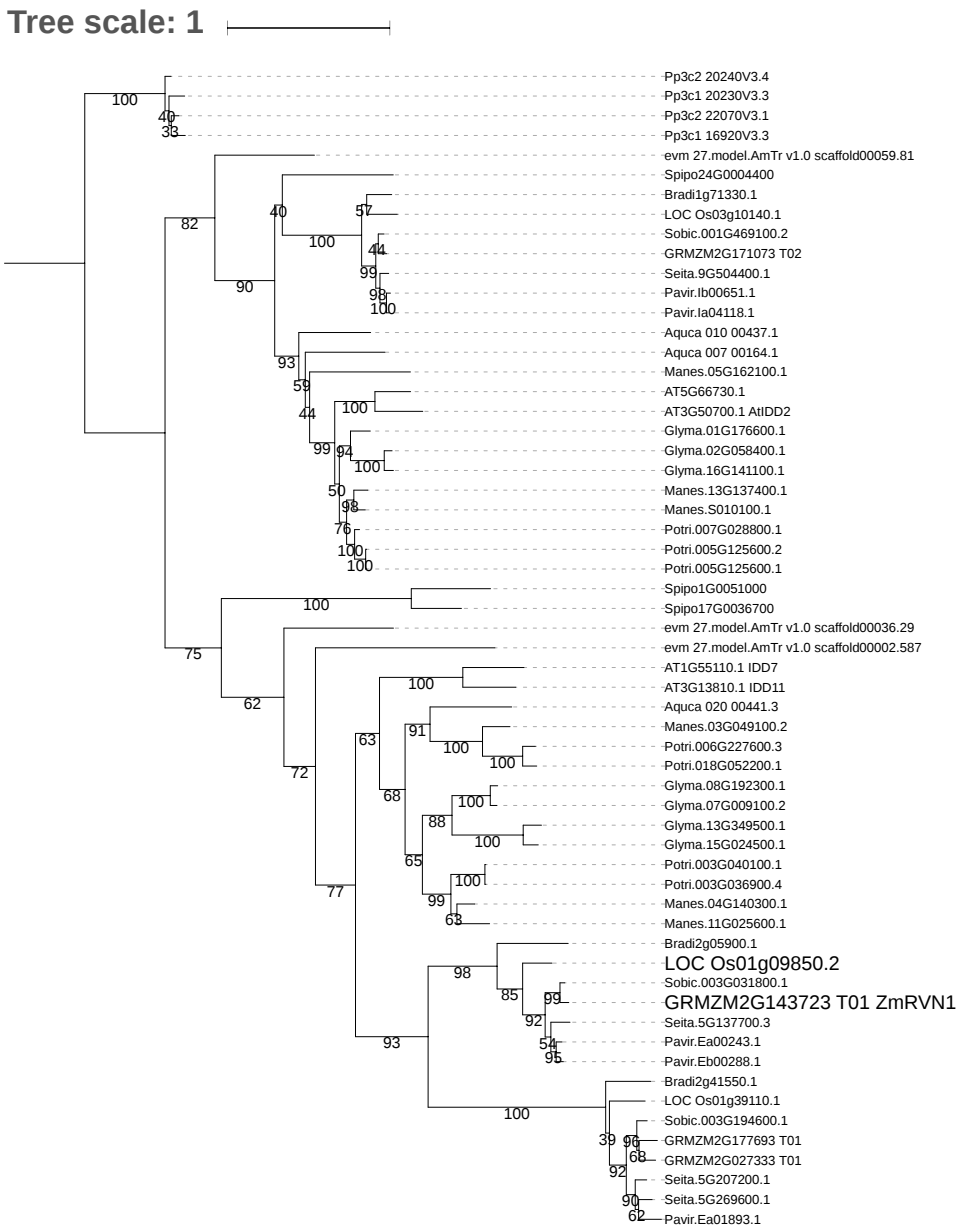


Figure A.6: Phylogenetic tree of RVN1 proteins Maximum-likelihood inference of phylogenetic relationships between the RAVEN1 (RVN1) proteins in monocot and eudicot species. Branch length measured in number of amino acid substitutions per site, bootstrap values shown for each node. Putative Kranz candidates (enlarged "GRMZM" labels) and rice (enlarged "LOC" bold labels) are highlighted.

A.7 SCR1 phylogeny

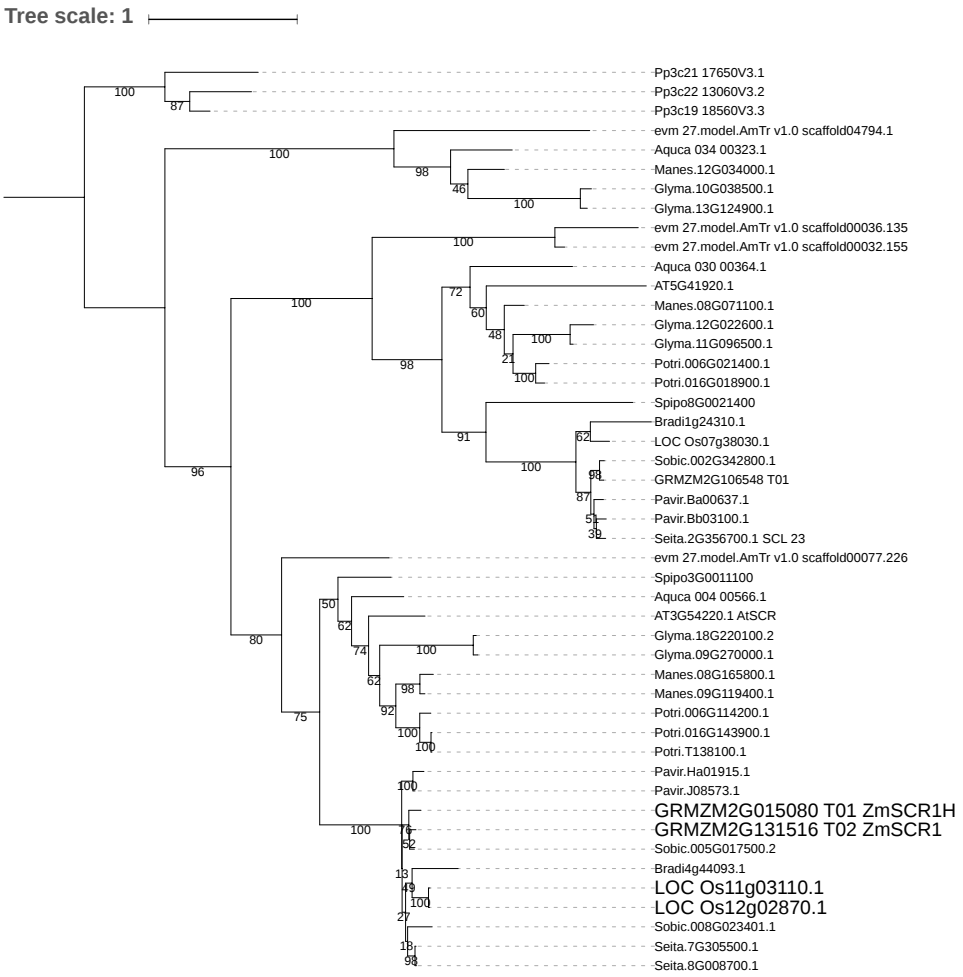


Figure A.7: Phylogenetic tree of SCR proteins Maximum-likelihood inference of phylogenetic relationships between the SCARECROW (SCR) proteins in monocot and eudicot species. Branch length measured in number of amino acid substitutions per site, bootstrap values shown for each node. Putative Kranz candidates (enlarged "GRMZM" labels) and rice (enlarged "LOC" bold labels) are highlighted.

A.8 SHR phylogeny

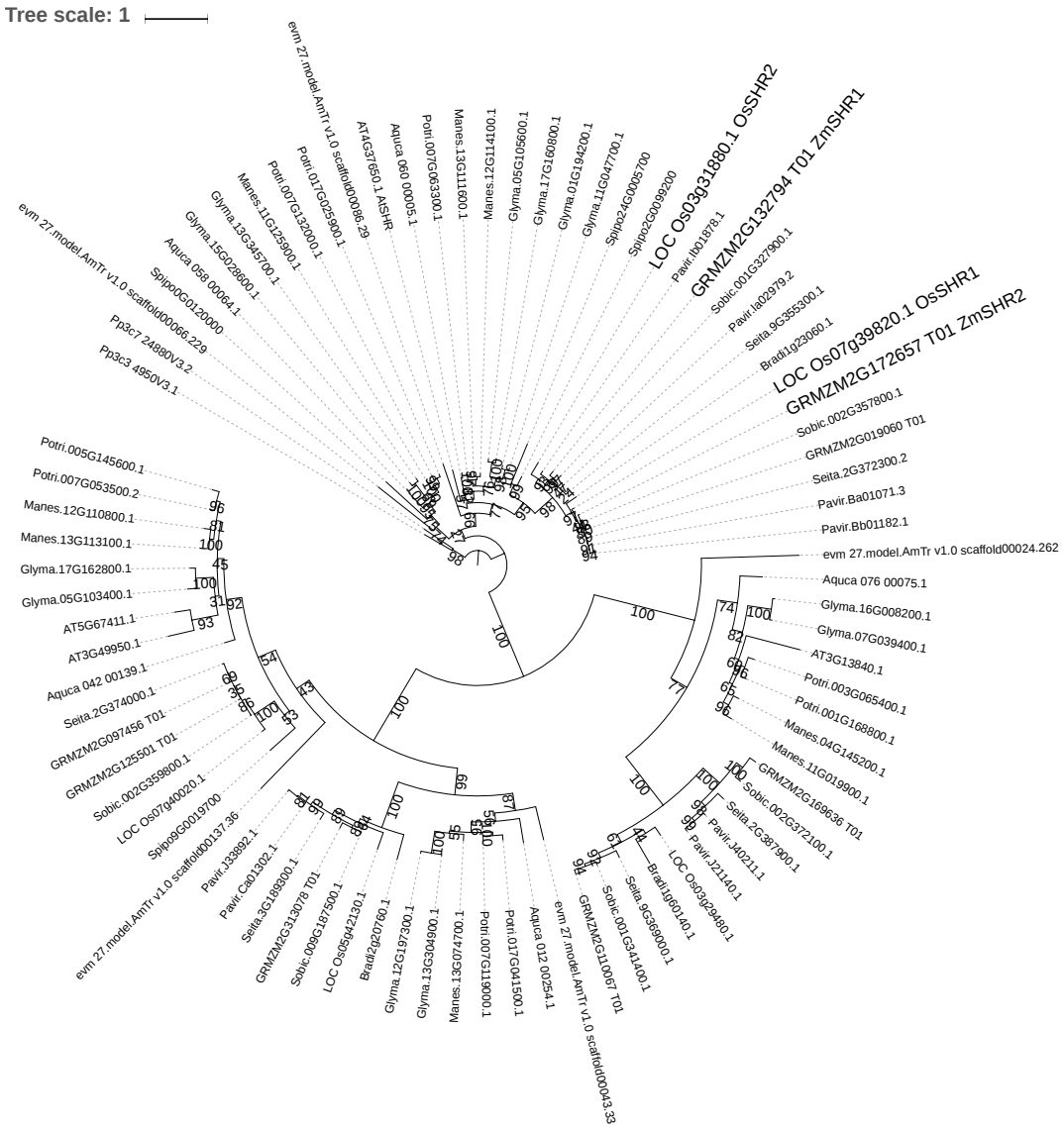


Figure A.8: Phylogenetic tree of SHR proteins Maximum-likelihood inference of phylogenetic relationships between the SHORTROOT (SHR) proteins in monocot and eudicot species. Branch length measured in number of amino acid substitutions per site, bootstrap values shown for each node. Putative Kranz candidates (enlarged "GRMZM" labels) and rice (enlarged "LOC" bold labels) are highlighted.

B

AINTEGUMENTA-like protein targets and supplementary experiments

B.1 AIL protein target genes in maize and arabidopsis

To investigate AIL target genes in maize and arabidopsis, 5 kbp (maize) or 1 kbp (arabidopsis) DNA regions upstream of predicted open reading frames were searched for the ANT DNA binding consensus region (5′-gCAC(A/G)N(A/T)TcCC-(a/g)ANG(c/t)-3′); 2 base mis-matches were allowed. The genes were filtered for transcription >5 transcripts per million during leaf development using datasets from Wang *et al.* (2013b) (maize) or Woo *et al.* (2016) (arabidopsis). The genes identified are listed in table B.1 (p.235-242) (maize) and table B.2 (p.243-250) (arabidopsis).

B.2 GO term analysis of Arabidopsis AIL protein target genes

To investigate arabidopsis AIL protein targets, 1 kbp DNA regions upstream of predicted arabidopsis ORFs were searched, and the genes were filtered for mRNA with >5 transcripts per million during arabidopsis leaf development using data from Woo *et al.* (2016). GO term analysis was conducted on these different gene lists using PANTHER (Mi *et al.*, 2016) for biological process and molecular function GO terms. Many more AIL targets were identified in arabidopsis leaf development compared to maize (453 genes in arabidopsis compared to 262 genes in maize). This may be because the arabidopsis transcriptome data included slightly older leaf developmental stages. Comparing the AIL target gene GO profiles between the maize (section 4.2.7; p.143) and arabidopsis datasets shows that similar classes of biological processes (fig. B.1; p.253) and molecular functions (fig. B.2; p.254) are predicted to be regulated by AIL proteins in both developmental contexts. However, arabidopsis AIL targets had more metabolism-related processes, and did not have the carbohydrate transporter GO terms that were found in maize. [1]

[1] Continued on p.251.

Table B.1: Maize AIL target genes mRNA expression in relative transcript abundance in- foliar (F) and husk (H) plastochron 1 (P1), 3&4 (P3&4) and 5 (P5), immature (I) and expanded (E) leaves (Wang *et al.*, 2013b), and in mature bundle sheath (BS) and mesophyll (M) cells (Li *et al.*, 2010). Annotation based on orthology to arabidopsis genes.

Gene ID	FP 1	FP 3&4	FP 5	FI	FE	HP 1	HP 3&4	HP 5	HI	HE	BS	M	Annotation
AC149810.2_FGT004	0	9	24	73	10	0	20	12	2	0	2	9	Protein PAM68
AC185217.3_FGT007	16	15	8	3	3	5	11	11	5	6	0	0	RNA helicase
AC189879.3_FGT003	2182	2935	3144	351	71	1323	2008	2174	135	53	2	0	Protein argonaute 10; AGO10
AC190788.2_FGT004	4	3	5	2	4	7	8	8	1	2	0	0	Unknown
AC193787.3_FGT008	10	13	12	6	2	18	13	20	16	7	4	0	Unknown
AC196475.3_FGT004	13	122	241	4281	582	68	214	418	11329	5480	31	4	Flavone 3′-O-methyltransferase 1; OMT1
AC198366.3_FGT004	1	2	1	0	0	1	2	3	13	5	0	0	Scarecrow-like protein 3; SCL3
AC199843.4_FGT007	2	3	1	2	1	4	3	7	2	2	0	0	RNA helicase
AC200099.4_FGT006	0	0	0	1	54	5	14	4	23	27	0	0	Cytochrome P450
AC212040.2_FGT006	10	11	8	1	0	7	34	19	5	1	1	0	Unknown
AC214771.3_FGT003	2	1	2	5	1	2	13	6	5	2	0	0	Core-2/I-branching beta-1
AC232238.2_FGT004	18	16	63	19	43	41	19	11	5	11	4	0	BZIP transcription factor
AC233850.1_FGT005	109	96	84	10	12	81	90	87	14	15	1	0	Transcription factor E2FB
AC233901.1_FGT004	38	16	4	5	4	27	26	9	24	17	26	4	La-related protein 1C; LARP1C
AC234164.1_FGT004	306	399	391	11	11	123	47	86	15	8	0	0	Scarecrow-like protein 28; SCL28
GRMZM2G002531	60	55	39	61	74	78	68	46	69	62	1	2	ACT-like protein tyrosine kinase family protein
GRMZM2G004057	3592	3491	3093	1810	761	2863	1953	2069	1400	1249	10	2	RIN13
GRMZM2G005036	1567	1319	1393	748	639	1402	1101	845	1220	1195	33	21	Probable mitochondrial-processing peptidase subunit alpha-1
GRMZM2G005256	1106	968	807	179	214	753	680	643	362	361	1	0	Nucleolar protein-like; OLI2
GRMZM2G007466	89	71	95	579	1274	112	138	192	1384	740	13	5	Unknown
GRMZM2G007659	69	78	76	24	3	41	44	46	2	3	0	2	Cell division control protein 6 homolog; CDC6
GRMZM2G008216	111	89	116	321	84	112	254	154	73	119	8	33	Cytidine/deoxycytidylate deaminase family protein
GRMZM2G009344	1	6	4	53	446	139	32	14	54	165	58	4	Purine permease 1; PUP1
GRMZM2G010011	1143	1363	1233	92	4	1236	1146	1425	177	12	0	0	Protein S-acyltransferase 21; PAT21
GRMZM2G010714	227	222	210	67	34	180	190	223	152	64	3	1	Protein trichome birefringence-like 13; TBL13
GRMZM2G011483	12732	12817	10230	1325	400	8839	6874	7591	1200	644	3	3	Uncharacterized protein
GRMZM2G014089	1	2	4	26	35	11	27	17	109	308	5	1	ABC transporter B family member 7; ABCB7
GRMZM2G015872	0	0	2	1	0	0	0	1	26	7	0	0	Uncharacterized protein

Continued on next page

Table B.1 –continued from previous page

Gene ID	FP 1	FP 3&4	FP 5	FI	FE	HP 1	HP 3&4	HP 5	HI	HE	BS	M	Annotation
GRMZM2G016254	43	32	16	77	4	405	532	459	190	509	1	7	Alpha-humulene/(-)-(E)-beta-caryophyllene synthase; TPS21
GRMZM2G016480	1091	1776	2020	502	63	1830	2534	2974	403	574	2	0	Probable inactive receptor kinase
GRMZM2G016521	2	0	1	3	20	9	24	19	20	23	1	22	Uncharacterized protein
GRMZM2G018631	1	1	0	74	25	0	1	2	96	44	0	4	RING-H2 finger proteinL13
GRMZM2G019358	972	716	399	4604	10134	1851	1720	1078	1195	4114	318	266	NAD(P)-binding Rossmann-fold superfamily protein
GRMZM2G019386	209	248	286	386	259	182	257	292	405	320	16	29	Putative uncharacterized protein; F17A8.180
GRMZM2G020196	95	77	41	155	180	58	100	82	197	164	5	1	U-box domain-containing protein 17; PUB17
GRMZM2G021301	8	10	10	7	0	2	3	2	1	0	0	0	Unknown
GRMZM2G022787	463	372	321	153	116	521	410	329	135	110	1	0	Pentatricopeptide repeat-containing protein
GRMZM2G024310	1299	1113	1349	1219	790	967	659	856	1378	1194	82	76	Probablease 24 kDa subunit
GRMZM2G027955	3185	1838	2258	548	3462	3150	1632	1853	3197	4530	11	100	Glucose-1-phosphate adenylyltransferase large subunit 3
GRMZM2G028726	247	355	390	18	2	105	209	401	97	9	0	1	Guanine nucleotide-binding protein subunit gamma 3; GG3
GRMZM2G029135	82	68	64	171	107	128	125	134	399	330	2	1	Unknown
GRMZM2G029810	149	154	131	55	76	212	142	178	166	102	0	0	Unknown
GRMZM2G030710	1400	1266	981	711	118	2057	1895	1361	270	192	5	3	Auxin response factor 3; ARF3
GRMZM2G032905	0	1	1	0	0	0	6	8	7	5	3	0	Agamous-like MADS-box protein AGL16
GRMZM2G033971	62	41	21	82	135	52	33	28	27	58	41	99	Acid phosphatase/vanadium-dependent haloperoxidase-related
GRMZM2G035118	1061	874	918	345	215	2211	2304	1303	742	506	2	1	Pumilio homolog 6
GRMZM2G036368	66	8	6	0	0	41	10	7	0	0	0	0	Heavy-metal-associated domain-containing protein; T15F16.6
GRMZM2G036418	163	137	112	122	149	205	202	156	300	249	8	4	RINT1-like protein MAG2; MAG2
GRMZM2G037469	57	39	47	371	71	44	72	84	611	140	2	2	Probable beta-1
GRMZM2G038152	241	213	174	27	14	215	113	244	24	25	0	0	Kelch motif-containing protein
GRMZM2G041163	2	3	3	21	13	5	10	10	55	85	1	0	Unknown
GRMZM2G041175	5	3	5	331	963	25	74	49	869	790	53	57	Tetraspanin-8; TET8
GRMZM2G042789	4	23	1	26	84	15	136	5	38	1265	12	18	PR-6 proteinase inhibitor family protein
GRMZM2G043331	772	905	932	255	318	620	583	642	304	206	7	4	Unknown
GRMZM2G046587	1194	660	369	508	1609	561	525	453	343	403	470	117	Soluble inorganic pyrophosphatase 4; PPA4
GRMZM2G047097	516	463	444	288	323	546	421	429	465	457	37	36	Mitochondrial fission 1 protein A; FIS1A
GRMZM2G048912	369	363	312	270	1052	400	365	353	189	210	3	6	Putative uncharacterized protein
GRMZM2G049269	2508	2652	3309	3570	386	2668	2485	3488	2602	993	4	2	Probable methyltransferase PMT26
GRMZM2G050137	1525	300	13	1	0	161	5	1	0	0	1	0	Probable N-acetyltransferase HLS1-like
GRMZM2G053199	107	142	104	38	24	165	188	116	125	54	9	9	Serine/threonine-protein phosphatase PP1 isozyme 4; TOPP4
GRMZM2G053338	1698	1123	246	3	1	200	174	88	8	5	0	0	Indole-3-acetic acid-amido synthetase GH3.2; GH3.2
GRMZM2G055279	10	7	11	11	1	10	18	12	5	3	0	0	Copper transporter 6; COPT6
GRMZM2G055527	341	262	206	133	121	241	179	186	198	199	9	16	Unknown
GRMZM2G057618	319	457	535	137	98	202	140	231	318	111	5	1	Protein kinase family protein
GRMZM2G058366	356	398	744	803	579	450	963	1334	648	675	1	1	U11/U12 small nuclear ribonucleoprotein 59 kDa protein

Continued on next page

Table B.1 –continued from previous page

Gene ID	FP 1	FP 3&4	FP 5	FI	FE	HP 1	HP 3&4	HP 5	HI	HE	BS	M	Annotation
GRMZM2G058491	27	10	37	8	40	82	1328	657	1670	761	0	0	P-loop containing nucleoside triphosphate hydrolase
GRMZM2G060070	62	44	43	93	11	62	32	36	23	30	0	1	Unknown
GRMZM2G062452	74	74	81	263	215	111	123	136	357	342	17	28	Unknown
GRMZM2G062577	2248	3373	4708	100	1	2393	6397	8807	401	43	0	0	GDSL esterase/lipase
GRMZM2G063657	370	178	200	267	667	991	708	356	265	162	6	1	PAP/OAS1 substrate-binding domain superfamily
GRMZM2G064695	1933	1547	1628	1306	1060	1371	879	1066	1829	1695	78	32	Succinyl-CoA ligase [ADP-forming] subunit beta
GRMZM2G065088	2	2	1	11	0	1	3	0	166	2	0	0	Germin-like protein subfamily 2 member 1; GLP4
GRMZM2G065471	1	0	0	0	2	1	6	1	15	29	0	0	Probable carboxylesterase 15; CXE15
GRMZM2G066496	1071	1024	690	640	1562	904	853	858	2256	2034	246	33	ACT domain-containing protein ACR12; ACR12
GRMZM2G067575	1	1	1	1001	4911	2	3	7	2449	1644	1336	5	Unknown
GRMZM2G069726	188	190	211	417	150	219	267	212	111	166	3	23	GTP-binding protein; EMB2001
GRMZM2G070442	329	111	29	2	0	886	147	52	0	0	0	0	Protein UPSTREAM OF FLC; UFC
GRMZM2G070503	13	5	7	2	6	21	19	11	9	8	1	0	Putative disease resistance RPP13-like protein 1 RPPL1
GRMZM2G070563	8	8	6	17	32	4	19	11	5	30	23	4	Auxin efflux carrier family protein
GRMZM2G070649	479	436	445	310	201	366	302	317	265	266	4	5	Expressed protein
GRMZM2G071154	264	386	357	1847	1337	243	1199	1090	7087	6827	97	76	DCD (Development and Cell Death) domain protein
GRMZM2G075683	754	673	653	427	187	784	576	718	462	417	17	13	Unknown CKL13
GRMZM2G076943	1	11	40	11	513	54	233	220	1617	1711	0	0	Cysteine-rich receptor-like protein kinase 11; CRK11
GRMZM2G076950	42	54	73	1759	383	75	58	71	55	111	7	17	Pentatricopeptide repeat-containing protein
GRMZM2G077307	1099	1074	1005	1887	5021	2591	5251	3088	9984	8922	107	9	CHY and CTCHY and RING-type zinc finger protein
GRMZM2G078324	9	18	10	455	701	10	8	14	14	216	94	253	YCF37
GRMZM2G078806	1	1	0	13	95	26	119	61	34	59	2	1	Unknown
GRMZM2G079219	1	2	2	2	160	102	272	95	210	150	16	0	Uncharacterized protein
GRMZM2G079470	581	381	266	257	572	526	329	247	504	400	49	0	Scarecrow-like protein 27; SCL27
GRMZM2G080231	8	45	73	340	59	9	66	124	434	134	4	0	Probable beta-1
GRMZM2G080243	119	152	124	46	28	90	82	74	28	34	10	1	Mitochondrial carrier protein
GRMZM2G081803	1	2	1	329	383	4	3	2	6	69	95	162	Chaperonin-like RbcX protein
GRMZM2G082529	608	684	800	904	545	682	985	1085	1044	683	9	4	Endosomal targeting BRO1-like domain-containing protein
GRMZM2G083346	458	426	444	140	137	515	461	371	344	289	3	3	Cleavage and polyadenylation specificity factor subunit 3-I; CPSF73-I
GRMZM2G084037	102	78	38	33	119	96	46	43	41	23	1	0	Unknown
GRMZM2G084496	43	32	16	77	4	405	532	459	190	509	1	7	Alpha-humulene/(-)-(E)-beta-caryophyllene synthase; TPS21
GRMZM2G087692	154	106	97	82	77	148	124	80	108	122	4	3	Putative methyltransferase
GRMZM2G088312	1	2	17	3	88	18	22	7	29	222	3	0	Putative uncharacterized protein T2H7.2; T2H7.2
GRMZM2G090627	3	3	2	7	5	5	6	7	7	7	1	0	Mechanosensitive ion channel protein 6; MSL6
GRMZM2G090865	124	209	236	153	335	865	512	309	301	198	112	33	Unknown
GRMZM2G093490	2	3	5	3	7	3	7	6	3	2	0	0	Leucine-rich repeat protein kinase-like protein; F18L15.90
GRMZM2G096458	1504	1363	1463	37	0	588	262	739	27	25	0	0	Chromatin assembly factor 1 subunit FAS1; FAS1

Continued on next page

Table B.1 –continued from previous page

Gene ID	FP 1	FP 3&4	FP 5	FI	FE	HP 1	HP 3&4	HP 5	HI	HE	BS	M	Annotation
GRMZM2G096695	457	508	466	1661	5988	522	511	332	913	7340	1583	1912	Unknown F24G24.100
GRMZM2G098187	2	22	50	77	91	7	11	23	70	71	2	0	Protein kinase family protein
GRMZM2G099666	38	39	46	473	341	410	262	260	1858	1781	14	21	Acyl-coenzyme A oxidase 2
GRMZM2G099696	856	862	785	1189	533	644	514	588	623	698	11	25	3-oxoacyl-[acyl-carrier-protein] reductase
GRMZM2G100229	1810	1834	2311	822	128	1735	1891	2576	1531	626	2	2	Unknown IQD28
GRMZM2G101513	36	28	19	7	28	65	47	21	31	19	0	0	Putative disease resistance RPP13-like protein 1; RPPL1
GRMZM2G101527	42	44	25	15	39	78	66	43	46	45	0	0	Putative disease resistance protein
GRMZM2G103657	0	0	1	47	2	0	0	1	50	11	1	0	Pollen Ole e 1 allergen and extensin family protein
GRMZM2G103972	0	0	1	1	15	0	2	0	41	565	0	0	Cellulose synthase-like protein D3; CSLD3
GRMZM2G104332	1	0	0	3	1	52	61	14	6	5	0	1	MUD21_2
GRMZM2G105253	761	1031	896	447	5	689	536	651	59	5	0	0	Rop guanine nucleotide exchange factor 1; ROPGEF1
GRMZM2G105387	3	8	53	21	6	14	15	17	109	45	6	3	Agamous-like MADS-box protein AGL12; AGL12
GRMZM2G105391	157	128	97	118	124	154	148	99	130	139	6	3	Glycosyltransferase
GRMZM2G107196	155	182	221	579	473	199	160	200	254	263	16	18	Pyruvate dehydrogenase (acetyl-transferring) kinase
GRMZM2G107444	800	953	1052	603	721	676	678	898	1686	1721	35	24	Membrane steroid-binding protein 2; MSBP2
GRMZM2G107711	251	19	136	888	0	61	39	142	44	2	0	0	Unknown
GRMZM2G108981	241	293	220	8	20	138	157	165	52	22	0	0	Pentatricopeptide repeat-containing protein
GRMZM2G108991	261	403	134	46	0	44	21	23	60	4	0	0	Serine/threonine protein phosphatase 2A subunit B
GRMZM2G111216	0	2	3	503	316	0	3	16	120	221	48	20	Chloroplast stem-loop binding protein of 41 kDa a
GRMZM2G111713	545	504	435	131	75	419	376	357	261	165	0	0	Mitochondrial inner membrane translocase complex
GRMZM2G111928	5	8	5	4	2	7	8	17	6	2	4	1	Probable disease resistance protein RDL5; RDL5
GRMZM2G112074	1193	1750	1703	47	0	446	372	1082	93	87	0	0	DNA replication licensing factor MCM7; MCM7
GRMZM2G112795	3656	2964	2624	509	499	1667	862	1206	854	1309	65	117	Unknown T1B9.10
GRMZM2G113250	1001	754	509	159	269	1333	1237	914	958	1000	8	5	Unknown
GRMZM2G120463	0	5	6	2	1	5	6	9	1	2	0	0	Unknown
GRMZM2G124254	181	247	240	128	171	167	190	219	194	221	10	11	Unknown
GRMZM2G125259	278	215	190	19	29	408	283	173	148	97	0	0	Bromodomain transcription factor; T20K14_180
GRMZM2G126865	158	157	230	1070	344	199	513	590	1548	775	6	1	Transmembrane 9 superfamily member 12; TMN12
GRMZM2G127031	2	1	2	81	97	2	3	4	100	220	13	9	Putative uncharacterized protein
GRMZM2G127162	0	1	0	0	0	13	13	8	0	0	0	0	Unknown
GRMZM2G127609	1162	1232	1587	5511	1118	1279	1138	1400	1847	3940	26	79	20 kDa chaperonin
GRMZM2G130034	456	451	390	322	1916	182	224	208	709	1317	114	10	Ureidoglycolate hydrolase; UAH
GRMZM2G130356	131	95	43	43	400	149	183	111	173	206	11	30	Alpha/beta-Hydrolases superfamily protein
GRMZM2G130375	929	824	796	381	104	694	532	726	189	121	6	1	Beta-galactosidase 1; BGAL1
GRMZM2G131961	3	8	19	111	54	162	163	64	28	4	19	0	Transcription factor TGA3
GRMZM2G132862	51	36	61	10	0	16	6	18	4	3	1	0	Unknown
GRMZM2G134770	579	619	581	265	283	634	692	599	423	319	9	12	RanBP2-type zinc finger protein

Continued on next page

Table B.1 –continued from previous page

Gene ID	FP 1	FP 3&4	FP 5	FI	FE	HP 1	HP 3&4	HP 5	HI	HE	BS	M	Annotation
GRMZM2G134878	33	74	75	32	110	41	88	115	50	89	9	23	Putative uncharacterized protein; T5K6_60
GRMZM2G135400	1	0	3	99	13	24	13	46	258	206	0	0	BRI1 kinase inhibitor 1; BKI1
GRMZM2G135727	4236	3919	3874	1504	473	3251	2427	2568	1634	1156	6	22	60S ribosomal protein L3-1; ARP1
GRMZM2G135960	50	8	7	6	1	13	174	8	35	91	1	0	23.5 kDa heat shock protein
GRMZM2G136300	18	17	21	37	131	137	236	109	393	105	1	3	Amino acid permease 2; AAP2
GRMZM2G136635	462	399	222	38	349	504	237	202	547	1591	52	16	Unknown
GRMZM2G138012	32	34	44	5	7	23	26	38	21	9	0	0	Pentatricopeptide repeat-containing protein; PCMP-E27
GRMZM2G140051	5457	5364	4975	982	522	4772	4521	4543	2932	2200	4	5	Nucleosome assembly protein 1; NAP1
GRMZM2G141185	179	132	95	224	299	245	217	129	189	314	18	6	Uncharacterized protein
GRMZM2G141818	10063	6466	4079	438	152	6589	3473	3103	521	360	0	0	Protein argonaute 4; AGO4
GRMZM2G145482	296	332	303	106	83	237	246	311	161	123	2	1	U-box domain-containing protein 62; PUB62
GRMZM2G145950	47	43	66	0	0	22	9	22	0	0	0	0	Unknown
GRMZM2G147420	65	132	226	33	3	86	175	307	27	14	1	0	Unknown
GRMZM2G147791	5534	4613	4573	3377	1007	2808	1655	2358	1558	1530	20	131	60S ribosomal protein L29-1; RPL29A
GRMZM2G147811	1404	1142	903	664	925	2641	2387	1640	1240	698	23	3	Protein kinase family protein
GRMZM2G149556	855	953	811	1211	417	1272	1669	548	1192	811	7	6	Unknown
GRMZM2G149649	1810	1987	1874	802	398	1368	1018	1209	776	673	3	10	60S ribosomal protein L12-1; RPL12A
GRMZM2G149905	1	2	18	93	6	7	11	48	165	39	0	0	Uncharacterized protein
GRMZM2G150408	548	763	990	17	2	290	218	359	4	4	3	1	Unknown
GRMZM2G150950	141	120	105	71	354	192	185	141	129	213	35	7	Unknown
GRMZM2G152891	1106	1621	1688	1102	820	736	1089	1299	577	1108	60	27	Calmodulin-2; CAM2
GRMZM2G153138	269	358	456	632	170	338	613	627	1104	660	19	4	Unknown
GRMZM2G155397	22	41	33	15	34	30	15	11	0	0	0	0	Unknown
GRMZM2G156977	0	0	0	2	1	2	13	3	20	22	0	0	NAC domain-containing protein 101; NAC101
GRMZM2G159041	106	92	81	197	159	203	275	218	231	287	29	42	Expressed protein
GRMZM2G160842	129	140	63	43	49	207	176	99	65	71	7	8	Adenylyltransferase and sulfurtransferase MOCS3; MOCS3
GRMZM2G161472	115	134	149	21	89	120	100	123	45	36	1	0	Ent-kaurene oxidase
GRMZM2G162266	166	214	274	298	157	314	415	342	227	201	5	5	Protein AE7; AE7
GRMZM2G163247	91	70	63	39	59	142	129	85	87	104	10	11	RNA recognition motif-containing protein
GRMZM2G163776	21	39	11	25	43	4	33	50	335	93	2	1	Genomic DNA
GRMZM2G165290	464	575	502	128	104	372	407	371	281	221	1	0	Pentatricopeptide repeat-containing protein
GRMZM2G167018	13	8	9	23	451	55	39	25	31	30	75	26	NAC domain-containing protein 101; NAC101
GRMZM2G168474	0	0	0	4	14	37	15	2	12	62	1	1	Glycosyltransferase
GRMZM2G168762	2069	1193	1276	1308	1032	1573	852	840	1855	2006	116	89	Probablease 24 kDa subunit
GRMZM2G168791	325	281	276	291	613	371	369	216	325	439	76	3	Putative anthranilate phosphoribosyltransferase
GRMZM2G170941	68	87	76	31	33	60	65	47	42	32	1	1	Pentatricopeptide repeat-containing protein
GRMZM2G171852	175	439	507	184	142	122	224	334	172	117	15	0	Cyclic dof factor 1; CDF1

Continued on next page

Table B.1 –continued from previous page

Gene ID	FP 1	FP 3&4	FP 5	FI	FE	HP 1	HP 3&4	HP 5	HI	HE	BS	M	Annotation
GRMZM2G172529	25	21	27	40	22	24	42	31	194	93	1	1	Putative uncharacterized protein; F13M23.50
GRMZM2G172945	195	215	248	31	1	81	31	83	4	0	0	0	Unknown
GRMZM2G175812	0	1	1	0	0	43	130	18	3	4	0	0	Chalcone synthase; CHS
GRMZM2G176282	226	188	140	206	52	205	216	205	193	70	1	1	EXS (ERD1/XPR1/SYG1) domain-containing protein
GRMZM2G176737	81	122	203	276	325	51	69	161	151	144	11	2	Unknown
GRMZM2G177050	90	73	166	228	341	78	277	263	47	13	20	28	CBL-interacting serine/threonine-protein kinase 11; CIPK11
GRMZM2G177356	0	1	5	58	43	3	9	13	17	12	1	4	Unknown
GRMZM2G177631	1160	1116	1192	3376	25	1260	2200	2748	3707	799	3	2	Cellulose synthase A catalytic subunit 6 UDP-forming; CESA6
GRMZM2G179885	0	0	0	1	4	0	0	0	47	213	0	0	NAC domain-containing protein 18; NAC018
GRMZM2G180879	17	14	6	8	11	22	24	19	3	8	4	0	Unknown
GRMZM2G305409	123	122	118	79	81	150	194	166	136	94	9	6	Unknown
GRMZM2G306062	141	322	438	80	0	147	106	194	6	2	1	0	Type IV inositol polyphosphate 5-phosphatase 7; IP5P7
GRMZM2G309897	214	375	358	64	22	172	267	382	231	51	0	0	Probable leucine-rich repeat receptor-like protein kinase
GRMZM2G311665	931	1091	993	674	604	961	2203	1869	1025	470	4	1	Zinc finger CCCH domain-containing protein 46
GRMZM2G315199	151	96	66	199	267	586	306	180	82	96	89	8	Unknown
GRMZM2G317450	5	9	43	91	13	1	7	25	13	7	0	0	Transcription factor BIM1
GRMZM2G318633	76	98	29	33	56	98	96	137	140	60	1	1	CASP-like protein 2D1
GRMZM2G318671	53	74	86	23	13	33	49	35	26	24	0	0	GCK domain-containing protein
GRMZM2G319109	146	115	135	1055	446	199	150	120	148	211	13	23	Probable transmembrane GTPase FZO-like
GRMZM2G324285	48	48	41	47	56	58	61	40	40	38	6	4	Putative anthranilate phosphoribosyltransferase
GRMZM2G324767	10	8	4	44	1	3	5	7	61	28	0	0	Genomic DNA
GRMZM2G324781	17	1	2	2	5	9	3	5	26	11	0	0	RING finger-related
GRMZM2G326707	1	1	4	140	1537	27	144	78	503	261	4	14	Major facilitator protein
GRMZM2G329073	73	36	22	0	0	13	10	10	0	0	0	0	Expressed protein
GRMZM2G336789	86	83	98	54	24	108	130	103	106	98	3	2	AP-1 complex subunit sigma-1; AAP19-1
GRMZM2G336908	12	5	2	481	1225	31	15	28	495	922	149	9	Monofunctional riboflavin biosynthesis protein RIBA 3
GRMZM2G343080	182	143	151	115	18	160	160	117	161	188	5	2	Unknown
GRMZM2G351160	514	529	411	195	226	567	600	437	388	400	11	3	Galactose oxidase/kelch repeat-containing protein
GRMZM2G351832	60	34	32	26	49	56	54	51	64	43	1	5	Putative uncharacterized protein; T13O15.7
GRMZM2G355021	11	10	7	8	5	17	8	8	11	10	0	0	Lysine-specific histone demethylase 1 homolog 1; LDL1
GRMZM2G362306	462	411	359	97	111	564	327	236	161	114	2	3	Cytoplasmic tRNA 2-thiolation protein 2; CTU2
GRMZM2G369939	1175	1039	1267	986	917	1947	1681	1472	1791	1422	4	6	Probable eukaryotic translation initiation factor 5-2
GRMZM2G377250	593	715	560	283	392	667	473	431	339	268	4	0	Unknown
GRMZM2G377686	7	9	7	12	16	22	15	19	24	10	2	0	Uncharacterized protein
GRMZM2G379540	0	1	4	138	99	1	11	4	428	206	6	7	RHOMBOID-like protein 6
GRMZM2G380665	553	599	583	156	120	413	330	404	215	129	1	3	Pentatricopeptide repeat-containing protein
GRMZM2G381025	16	14	14	4	26	13	22	36	69	110	0	0	UDP-glycosyltransferase 88A1

Continued on next page

Table B.1 –continued from previous page

Gene ID	FP 1	FP 3&4	FP 5	FI	FE	HP 1	HP 3&4	HP 5	HI	HE	BS	M	Annotation
GRMZM2G382077	785	778	646	275	217	721	444	437	265	189	10	7	Serine/threonine-protein phosphatase PP2A-3
GRMZM2G396482	20	59	50	10	16	32	82	50	16	8	8	3	Unknown
GRMZM2G400718	348	591	583	13	0	129	93	241	11	27	0	0	Unknown
GRMZM2G403719	9	6	5	7	7	16	5	4	4	6	1	0	Cysteine-rich receptor-like protein kinase 10; CRK10
GRMZM2G403915	119	45	59	37	42	157	59	55	56	37	1	1	CMP-sialic acid transporter 5
GRMZM2G407790	2	6	1	1	0	5	5	6	0	0	0	0	Unknown
GRMZM2G409722	3	3	3	55	110	41	91	72	190	158	7	27	Uncharacterized protein
GRMZM2G412986	0	0	1	7	188	1	23	9	138	141	1	0	Unknown
GRMZM2G414043	1220	974	990	544	114	1079	923	1290	1317	422	1	0	COP1-interacting protein-related protein
GRMZM2G419969	457	499	423	138	174	454	525	369	348	248	5	9	Multifunctional methyltransferase subunit TRM112-like protein
GRMZM2G421680	80	97	45	23	64	114	64	72	35	30	0	0	Unknown
GRMZM2G423129	8	41	42	47	55	40	148	289	172	129	2	2	Calcium-binding protein KIC; KIC
GRMZM2G429128	353	458	389	137	69	323	459	508	397	175	13	2	TRAM
GRMZM2G435445	309	262	235	293	301	568	561	358	705	810	17	11	Tubby-like F-box protein 7; TULP7
GRMZM2G446220	75	120	46	9	10	139	24	5	2	4	1	0	DOG1 alpha splice variant; DOG1
GRMZM2G449123	1992	1780	1646	952	870	2797	2320	1721	1494	1101	0	0	Pumilio homolog 6
GRMZM2G459590	9	10	3	1	1	17	17	10	6	1	0	0	Replication protein A 70 kDa DNA-binding subunit A; RPA1A
GRMZM2G466908	3	5	2	2	1	5	3	4	17	11	0	0	Auxin induced like-protein
GRMZM2G469477	14	13	13	20	65	8	10	17	27	26	24	7	Putative uncharacterized protein
GRMZM2G476762	0	0	0	174	732	0	0	0	1	2	51	0	Unknown
GRMZM2G479000	13	20	30	227	81	24	42	27	24	47	0	6	Protein PAM68
GRMZM2G512113	338	597	211	1	0	154	433	358	5	2	0	0	Unknown
GRMZM2G568478	5	11	3	1	1	12	10	21	4	4	0	0	Unknown
GRMZM2G589424	0	0	3	3	0	0	2	21	13	3	0	0	Unknown
GRMZM2G702514	1	1	0	0	1	1	2	0	29	15	0	0	NARINGENIN,2-OXOGLUTARATE 3-DIOXYGENASE
GRMZM5G805609	34	4	0	36	18	108	193	42	917	451	1	1	Glucan endo-1
GRMZM5G805685	62	4	35	55	1	122	295	569	202	35	1	0	B3 domain-containing protein
GRMZM5G813055	190	184	177	119	76	183	158	185	123	98	25	11	Lipase class 3-like protein
GRMZM5G813470	674	593	572	375	234	994	734	682	914	670	7	2	Putative zinc transporter
GRMZM5G813532	9	10	24	9	0	4	16	43	49	21	0	0	Zinc finger protein CONSTANS-LIKE 13; COL13
GRMZM5G826174	11	4	2	2	16	37	36	7	47	128	0	1	CLAVATA3/ESR (CLE)-related protein 2; CLE14
GRMZM5G826321	57	67	105	86	22	37	114	168	434	219	2	2	Probable carboxylesterase 12; CXE12
GRMZM5G835704	869	1377	1311	391	560	910	1303	1062	806	849	25	10	Unknown
GRMZM5G836567	138	98	67	19	11	114	83	79	45	18	1	0	Spc97 / Spc98 family of spindle pole body (SBP) component
GRMZM5G849473	0	0	3	3	0	0	2	21	13	3	0	1	Unknown
GRMZM5G854301	301	238	215	255	170	269	213	175	186	283	17	31	Unknown
GRMZM5G856002	13	17	6	5	2	9	12	10	3	1	4	7	Unknown

Continued on next page

Table B.1 –continued from previous page

Gene ID	FP 1	FP 3&4	FP 5	FI	FE	HP 1	HP 3&4	HP 5	HI	HE	BS	M	Annotation
GRMZM5G857090	518	463	185	23	1	257	116	136	18	15	0	0	Unknown
GRMZM5G864414	1048	1156	1206	1275	2361	1126	1214	1403	1876	1892	40	10	Putative serine/threonine protein kinase
GRMZM5G871910	167	134	117	102	83	121	108	111	79	74	3	6	Unknown
GRMZM5G875502	138	113	124	337	740	197	209	213	547	715	25	6	Histone-lysine N-methyltransferase CLF
GRMZM5G882080	6	5	11	1	9	18	18	3	9	2	0	0	Unknown
GRMZM5G882189	26	30	23	9	0	25	29	30	14	3	6	3	Probable galacturonosyltransferase 6; GAUT6
GRMZM5G891783	39	44	73	350	113	38	44	33	45	64	3	9	Putative oxidoreductase/electron carrier
GRMZM5G894156	537	500	480	313	375	854	904	631	364	410	15	13	Vesicle transport v-SNARE 11; VTI11
GRMZM5G894486	53	50	36	11	8	53	35	29	21	15	0	0	Pentatricopeptide repeat-containing protein
GRMZM5G896369	12	4	5	2	1	14	10	7	2	2	15	3	Unknown

Table B.2: Arabidopsis AIL target genes mRNA expression in transcripts per million in- 4, 6 and 8 day old arabidopsis leaves (4d, 6d, 8d respectively) (Woo *et al.*, 2016).

Gene ID	4d	6d	8d	Annotation
AT3G54890	1547	1373	1717	Chlorophyll a-b binding protein 6, chloroplastic; LHCA1
ATCG00470	1254	1343	1407	ATP synthase epsilon chain, chloroplastic; atpE
AT1G06680	1096	1259	1566	Oxygen-evolving enhancer protein 2-1, chloroplastic; PSBP1
AT1G79040	817	904	1121	Photosystem II 10 kDa polypeptide, chloroplastic; PSBR
AT2G10940	744	743	433	Unknown
AT2G31141	353	271	336	Uncharacterized protein
AT5G66190	295	440	500	Ferredoxin–NADP reductase, leaf isozyme 1, chloroplastic; LFNR1
AT1G02780	284	231	204	60S ribosomal protein L19-1; RPL19A
ATMG01390	242	35	8	Unknown
AT3G14210	183	354	412	GDSL esterase/lipase ESM1; ESM1
AT2G07732	173	58	54	Ribulose bisphosphate carboxylase large chain, catalytic domain
AT3G12120	162	220	243	Omega-6 fatty acid desaturase, endoplasmic reticulum; FAD2
AT2G20890	143	205	218	Protein THYLAKOID FORMATION 1, chloroplastic; THF1
AT3G52590	135	137	133	Ubiquitin-60S ribosomal protein L40-2; RPL40B
AT4G34190	130	128	169	Stress enhanced protein 1, chloroplastic; SEP1
AT1G71500	123	171	190	Putative uncharacterized protein ; F26A9.12
AT5G08670	121	156	146	ATP synthase subunit beta-1, mitochondrial
AT3G62410	119	170	214	Calvin cycle protein CP12-2, chloroplastic; CP12-2
AT4G35450	115	132	151	Ankyrin repeat domain-containing protein 2; AKR2
AT4G34110	115	117	114	Polyadenylate-binding protein 2; PAB2
AT2G40490	114	114	105	Uroporphyrinogen decarboxylase 2, chloroplastic; HEME2
AT3G60245	97	123	92	60S ribosomal protein L37a-2; RPL37AC
AT1G16080	96	100	112	Putative uncharacterized protein ; T24D18.18
AT4G38740	95	116	148	Peptidyl-prolyl cis-trans isomerase CYP18-3; CYP18-3
AT1G34000	86	158	168	Unknown
AT1G14345	80	148	138	NAD(P)-linked oxidoreductase-like protein
AT2G04390	79	71	56	40S ribosomal protein S17-1; RPS17A
AT4G00360	78	101	101	Cytochrome P450 86A2; CYP86A2
ATMG00280	77	49	44	Putative uncharacterized mitochondrial protein
AT3G16000	76	75	65	MAR-binding filament-like protein 1; MFP1
AT2G46220	74	118	115	Unknown
AT1G74060	70	57	56	60S ribosomal protein L6-2; RPL6B
AT4G33865	69	86	95	40S ribosomal protein S29; RPS29A
AT1G53210	67	73	80	Putative uncharacterized protein
AT1G56045	66	97	93	60S ribosomal protein L41; RPL41A
AT5G62700	64	55	42	Tubulin beta-3 chain; TUBB3
AT3G48420	61	81	104	Haloacid dehalogenase-like hydrolase domain-containing protein
AT1G78820	61	89	74	D-mannose binding lectin protein with Apple-like carbohydrate-binding domain
AT4G24190	60	57	51	Endoplasmin homolog; SHD
AT2G38170	59	91	143	Vacuolar cation/proton exchanger 1; CAX1
AT1G70490	59	72	84	ADP-ribosylation factor 2-B; ARF2-B
AT1G72930	58	112	119	Toll/interleukin-1 receptor-like protein; TIR
AT1G26850	58	73	72	Probable methyltransferase PMT2
AT1G48450	55	72	76	Unknown
AT2G29650	55	103	119	Sodium-dependent phosphate transport protein 1, chloroplastic; ANTR1
AT3G08030	53	68	71	Unknown
ATMG00160	52	46	40	Cytochrome c oxidase subunit 2; COX2
AT4G27600	51	55	50	NARA5
AT1G66970	51	82	80	Glycerophosphodiester phosphodiesterase GDPDL1; GDPDL1
AT2G39470	51	79	97	Photosynthetic NDH subunit of lumenal location 1, chloroplastic; PNSL1
AT1G15500	51	61	57	ADP,ATP carrier protein 2, chloroplastic; AATP2
AT1G51680	51	80	85	4-coumarate–CoA ligase 1; 4CL1
AT2G32060	50	47	39	40S ribosomal protein S12-2; RPS12C
AT2G31400	49	63	61	Pentatricopeptide repeat-containing protein , chloroplastic

Continued on next page

Table B.2 –continued from previous page

Gene ID	4d	6d	8d	Annotation
AT3G51920	49	64	65	Calmodulin-like protein 9; CML9
AT4G29410	49	51	37	60S ribosomal protein L28-2; RPL28C
AT5G16400	48	68	94	Thioredoxin F2, chloroplastic
AT3G20390	47	78	69	Reactive Intermediate Deaminase A, chloroplastic; RIDA
AT1G21065	47	96	118	Putative uncharacterized protein ; T22I11.11
AT3G24160	43	62	70	AT3g24160/MUJ8_16; PMP
AT2G07786	42	32	26	Unknown
AT4G02510	42	51	33	Translocase of chloroplast 159, chloroplastic; TOC159
AT3G06435	41	1	1	Unknown
AT3G07480	41	65	67	2Fe-2S ferredoxin-like protein; F21O3.19
AT1G49760	41	36	29	Polyadenylate-binding protein 8; PAB8
AT1G60550	41	61	60	1,4-dihydroxy-2-naphthoyl-CoA synthase, peroxisomal; MENB
AT1G29880	40	47	46	Glycine–tRNA ligase, mitochondrial 1
AT3G07470	39	51	52	Unknown
AT2G28900	39	57	68	Outer envelope pore protein 16-1, chloroplastic; OEP161
AT2G16660	39	44	36	Unknown
AT1G06190	39	41	32	Rho-N domain-containing protein 1, chloroplastic; RHON1
AT5G03040	38	67	63	AT5g03040/F15A17_70; iqd2
AT4G24930	38	66	59	Thylakoid lumenal 17.9 kDa protein, chloroplastic
AT2G27020	38	35	35	Proteasome subunit alpha type-3; PAG1
AT5G19290	38	51	49	Esterase/lipase/thioesterase family protein
AT4G37870	36	25	24	Phosphoenolpyruvate carboxykinase [ATP]; PCKA
AT2G33150	36	49	58	3-ketoacyl-CoA thiolase 2, peroxisomal; PED1
AT4G32210	36	42	39	Succinate dehydrogenase subunit 3-2, mitochondrial; SDH3-2
AT4G20440	36	27	20	Small nuclear ribonucleoprotein-associated protein; F9F13.90
AT4G24750	35	39	42	Rhodanese-like domain-containing protein 11, chloroplastic; STR11
AT1G14270	35	43	40	CAAX amino terminal protease family protein
AT5G52780	34	33	39	Uncharacterized protein PAM68-like
AT3G50950	34	39	30	Disease resistance RPP13-like protein 4; RPP13L4
AT1G08510	34	47	55	Palmitoyl-acyl carrier protein thioesterase, chloroplastic; FATB
AT5G47230	34	33	21	Ethylene-responsive transcription factor 5; ERF5
AT4G28025	33	60	64	Putative uncharacterized protein
AT1G50480	33	52	58	Formate–tetrahydrofolate ligase; THFS
AT5G19540	33	46	57	Putative uncharacterized protein
AT3G53580	32	35	34	Diaminopimelate epimerase, chloroplastic; DAPF
AT3G11930	32	48	58	Unknown
AT2G27030	32	45	42	Calmodulin-5; CAM5
AT1G72020	32	57	64	Putative uncharacterized protein ; F28P5.9
AT4G16370	32	44	44	Oligopeptide transporter 3; OPT3
AT5G51720	31	47	54	CDGSH iron-sulfur domain-containing protein NEET; NEET
AT4G25650	31	49	58	Protochlorophyllide-dependent translocon component 52, chloroplastic; PTC52
AT3G12290	31	47	55	Bifunctional protein FolD 2; FOLD2
AT1G22840	30	35	32	Cytochrome c-1; CYTC-1
AT5G40890	29	47	52	Chloride channel protein CLC-a; CLC-A
AT5G57030	29	37	35	Lycopene epsilon cyclase, chloroplastic; LUT2
AT3G46040	29	29	23	40S ribosomal protein S15a-4; RPS15AD
AT3G63460	29	39	33	Protein transport protein SEC31 homolog B; SEC31B
AT2G44050	29	43	40	6,7-dimethyl-8-ribityllumazine synthase, chloroplastic
AT5G21020	28	59	76	Unknown
AT4G14030	28	35	47	Selenium-binding protein 1; SBP1
AT1G35420	28	44	51	Dienelactone hydrolase-like protein
AT5G45670	28	28	9	GDSL esterase/lipase
AT5G53560	27	35	35	Cytochrome b5 isoform E; CYTB5-E
AT1G17200	27	24	28	CASP-like protein 2A1
AT4G37450	27	33	18	Lysine-rich arabinogalactan protein 18; AGP18
AT1G19020	27	13	7	Putative uncharacterized protein
AT1G66670	26	29	26	ATP-dependent Clp protease proteolytic subunit 3, chloroplastic; CLPP3
AT4G27310	26	39	40	B-box domain protein 28
AT3G55360	26	35	36	Very-long-chain enoyl-CoA reductase; ECR
AT1G12110	26	44	47	Protein NRT1/ PTR FAMILY 6.3; NPF6.3
AT4G01100	26	38	38	Mitochondrial adenine nucleotide transporter ADNT1; ADNT1
AT2G37110	25	18	21	Unknown
AT5G51545	25	28	26	Protein LOW PSII ACCUMULATION 2, chloroplastic; LPA2

Continued on next page

Table B.2 –continued from previous page

Gene ID	4d	6d	8d	Annotation
AT4G24820	25	32	30	26S proteasome non-ATPase regulatory subunit 6 homolog; RPN7
AT2G20610	25	47	47	S-alkyl-thiohydroximate lyase SUR1; SUR1
AT1G73530	24	32	32	Unknown T9L24.48
AT4G34980	24	36	46	Subtilisin-like protease SBT1.6; SBT1.6
AT1G72030	24	28	41	Acyl-CoA N-acyltransferases-like protein; F28P5.8
AT3G10850	24	35	34	Hydroxyacylglutathione hydrolase cytoplasmic; GLX2-2
AT2G46505	24	46	35	Succinate dehydrogenase subunit 4, mitochondrial; SDH4
AT5G27560	23	25	28	Putative uncharacterized protein
AT5G11500	23	34	33	Putative uncharacterized protein ; F15N18_90
AT5G66510	23	27	27	Gamma carbonic anhydrase 3, mitochondrial; GAMMACA3
AT1G32640	23	24	13	Transcription factor MYC2; MYC2
AT2G36000	22	34	28	Unknown
AT5G26030	22	23	17	Ferrochelatase-1, chloroplastic/mitochondrial; HEM15
AT3G54470	22	25	17	Uridine 5′-monophosphate synthase; PYRE-F
AT3G26900	22	37	42	Probable inactive shikimate kinase like 1, chloroplastic; SKL1
AT5G48385	21	26	24	FRIGIDA-like protein 3; FRL3
AT2G25520	21	32	29	Probable sugar phosphate/phosphate translocator
AT2G46540	21	27	25	Unknown
AT5G10960	21	21	26	Probable CCR4-associated factor 1 homolog 10; CAF1-10
AT3G17930	21	33	41	Unknown
AT1G03160	21	24	23	Probable transmembrane GTPase FZO-like, chloroplastic; FZL
AT5G58950	20	30	38	Unknown
AT4G30310	20	32	31	FGGY family of carbohydrate kinase
AT4G27230	20	21	22	Probable histone H2A.3
AT4G16143	20	23	19	Importin subunit alpha-2; IMPA2
AT3G47800	20	18	11	Aldose 1-epimerase; T23J7.130
AT1G32160	20	32	41	Unknown
AT1G21640	20	31	36	NAD kinase 2, chloroplastic; NADK2
AT5G45490	20	28	34	Probable disease resistance protein
AT3G55400	19	22	17	Methionine–tRNA ligase, chloroplastic/mitochondrial; OVA1
AT1G09795	19	29	28	ATP phosphoribosyltransferase 2, chloroplastic; HISN1B
AT5G10480	19	26	28	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase PASTICCINO 2; PAS2
AT1G30520	19	21	20	2-succinylbenzoate–CoA ligase, chloroplastic/peroxisomal; AAE14
AT1G51650	19	30	32	ATP synthase subunit epsilon, mitochondrial
AT2G27402	19	11	10	Unknown
AT3G11560	19	28	30	Unknown
AT5G02840	19	31	35	Protein REVEILLE 4; RVE4
AT5G13100	19	18	18	Unknown
AT2G44640	18	19	14	Expressed protein
AT2G46910	18	29	33	Probable plastid-lipid-associated protein 10, chloroplastic; PAP10
AT3G46000	18	32	36	Actin-depolymerizing factor 2; ADF2
AT5G08420	18	22	18	KRR1 small subunit processome component
AT3G59840	18	14	15	Allyl alcohol dehydrogenase-like protein
AT5G63840	18	22	22	Probable glucan 1,3-alpha-glucosidase; PSL5
AT4G34090	18	23	23	Unknown
AT1G73920	18	23	27	Alpha/beta-hydrolase domain-containing protein
AT1G02840	17	20	20	Serine/arginine-rich-splicing factor SR34; SR34
AT1G19050	17	16	13	Two-component response regulator ARR7; ARR7
AT4G16920	17	13	11	Disease resistance RPP5 like protein
AT3G09050	17	22	22	Unknown
AT1G15750	17	22	20	Protein TOPLESS; TPL
AT4G21105	17	33	48	Unknown
AT1G64790	17	21	13	Protein ILITYHIA; ILA
AT2G45970	17	18	16	Cytochrome P450 86A8; CYP86A8
AT3G57800	17	26	25	Transcription factor bHLH60; BHLH60
AT2G29360	17	24	33	Tropinone reductase homolog ; SDR
AT4G26555	16	25	25	Peptidyl-prolyl cis-trans isomerase FKBP16-1, chloroplastic; FKBP16-1
AT5G48380	16	13	8	Probably inactive leucine-rich repeat receptor-like protein kinase
AT1G04640	16	29	25	Octanoyltransferase; LIP2
AT2G38870	16	12	11	PR-6 proteinase inhibitor family protein
AT1G29370	16	17	14	Unknown
AT5G52750	16	12	5	Unknown
AT4G29160	16	25	28	Vacuolar protein sorting-associated protein 32 homolog 2; VPS32.2

Continued on next page

Table B.2 –continued from previous page

Gene ID	4d	6d	8d	Annotation
AT5G03740	16	15	11	Histone deacetylase HDT3; HDT3
AT3G54600	16	25	34	DJ-1 protein homolog F; DJ1F
AT5G06130	16	31	31	Chaperone protein dnaJ-like protein
AT5G07240	15	29	27	IQ-domain 24 protein; T28J14_180
AT3G53500	15	21	19	Serine/arginine-rich splicing factor RS2Z32; RS2Z32
AT4G24520	15	17	17	NADPH–cytochrome P450 reductase 1; ATR1
AT2G22230	15	17	14	Unknown
AT1G11870	15	22	16	Serine–tRNA ligase, chloroplastic/mitochondrial; OVA7
AT4G39940	15	29	33	Adenylyl-sulfate kinase 2, chloroplastic; APK2
AT5G42190	15	26	29	SKP1-like protein 1B; SKP1B
AT1G72940	15	17	9	Unknown
AT3G55220	15	10	8	Spliceosomal associated protein 130A; T26I12.100
AT1G20110	14	22	21	Unknown
AT4G20960	14	12	13	Riboflavin biosynthesis protein PYRD, chloroplastic; PYRD
AT5G01810	14	30	25	CBL-interacting serine/threonine-protein kinase 15; CIPK15
AT1G61790	14	14	10	Probable dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit 3B; OST3B
AT4G11600	14	17	28	Probable phospholipid hydroperoxide glutathione peroxidase 6, mitochondrial; GPX6
AT4G21660	14	14	14	Proline-rich spliceosome-associated (PSP) family protein
AT1G54350	14	24	25	ABC transporter D family member 2, chloroplastic; ABCC2
ATMG01020	14	19	19	Uncharacterized mitochondrial protein
AT1G04630	14	15	16	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13-A; MEE4
AT5G18630	14	23	24	Unknown
AT1G12860	14	15	14	Transcription factor SCREAM2; SCRM2
AT5G11580	14	15	16	Unknown
AT3G13300	14	21	19	Enhancer of mRNA-decapping protein 4; VCS
AT1G79560	13	16	11	ATP-dependent zinc metalloprotease FTSH 12, chloroplastic; FTSH12
AT3G52470	13	18	17	Late embryogenesis abundant hydroxyproline-rich glycoprotein; F22O6_150
AT1G63880	13	20	20	Disease resistance protein RML1B; RLM1B
AT5G26667	13	19	16	UMP-CMP kinase 3; UMK3
AT5G05100	13	20	16	Putative uncharacterized protein ; MUG13.4
AT3G10080	13	24	19	Germin-like protein subfamily 3 member 2
AT5G53370	13	24	27	Probable pectinesterase/pectinesterase inhibitor 61; PME61
AT5G65660	13	19	9	Uncharacterized protein
AT4G16860	13	16	16	Disease resistance protein RPP4; RPP4
AT1G01050	13	15	19	Soluble inorganic pyrophosphatase 1; PPA1
AT1G25145	13	12	13	Probable UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase 4; LPXC4
AT1G15130	13	19	18	Endosomal targeting BRO1-like domain-containing protein
AT1G20980	13	16	16	Squamosa promoter-binding-like protein 14; SPL14
AT4G20850	13	20	13	Tripeptidyl-peptidase 2; TPP2
AT5G19550	13	26	31	Aspartate aminotransferase, cytoplasmic isozyme 1; ASP2
AT2G25610	13	25	26	V-type proton ATPase subunit c”2; VHA-c”2
AT2G31070	13	22	21	Transcription factor TCP10; TCP10
AT5G64030	12	17	15	Probable methyltransferase PMT26
AT5G27670	12	12	15	Probable histone H2A.5
AT5G39210	12	16	14	Protein CHLORORESPIRATORY REDUCTION 7, chloroplastic; CRR7
AT1G68060	12	14	12	Microtubule-associated protein 70-1; MAP70.1
AT1G73990	12	18	18	Serine protease SPPA, chloroplastic; SPPA
AT3G57030	12	15	17	Protein STRICTOSIDINE SYNTHASE-LIKE 10; SSL10
AT5G54110	12	19	20	Vesicle-associated protein 4-1; PVA41
AT2G24120	12	11	6	DNA-directed RNA polymerase 3, chloroplastic; RPOT3
AT5G41210	12	15	18	Glutathione S-transferase T1; GSTT1
AT4G14070	12	16	18	Long-chain-fatty-acid–[acyl-carrier-protein] ligase AEE15, chloroplastic; AAE15
AT4G10710	12	14	14	FACT complex subunit SPT16; SPT16
AT5G08415	12	11	12	Lipoyl synthase, chloroplastic; LIP1P
AT1G75210	11	14	13	Cytosolic IMP-GMP specific 5-nucleotidase, putative
AT4G26670	11	10	9	Mitochondrial import inner membrane translocase subunit TIM22-2; TIM22-2
AT3G28460	11	13	11	Genomic DNA, chromosome 3, P1 clone: MFJ20
AT1G48110	11	14	14	Uncharacterized protein; ECT7
AT1G50740	11	11	7	Protein FATTY ACID EXPORT 5; FAX5

Continued on next page

Table B.2 –continued from previous page

Gene ID	4d	6d	8d	Annotation
AT1G14700	11	18	17	Purple acid phosphatase 3; PAP3
AT5G11270	11	11	10	Protein OVEREXPRESSOR OF CATIONIC PEROXIDASE 3; OCP3
AT5G14210	11	15	13	Unknown
AT2G01720	11	17	14	Dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit 1B; OST1B
AT1G53520	11	10	11	Fatty-acid-binding protein 3; FAP3
AT1G11480	11	12	11	Unknown
AT5G04130	11	11	10	DNA gyrase subunit B, mitochondrial; GYRBM
AT5G64960	11	12	11	Cyclin-dependent kinase C-2; CDKC-2
AT4G24805	11	15	16	Unknown
AT2G24570	11	15	15	Probable WRKY transcription factor 17; WRKY17
AT2G21270	11	15	15	UFD1
AT3G11450	11	13	11	DnaJ and Myb-like DNA-binding domain-containing protein
AT5G23920	11	24	29	Unknown
AT2G43770	11	13	11	Unknown
AT2G37920	11	16	17	Copper ion transmembrane transporter; emb1513
AT1G01510	11	15	13	C-terminal binding protein AN; AN
AT1G64150	11	15	14	GDT1-like protein 1, chloroplastic
AT5G62740	10	11	12	Hypersensitive-induced response protein 1; HIR1
AT1G22360	10	14	16	UDP-glycosyltransferase 85A2; UGT85A2
AT2G46230	10	10	9	Unknown
AT3G59040	10	9	6	Pentatricopeptide repeat-containing protein
AT3G61060	10	21	29	F-box protein PP2-A13; PP2A13
AT1G47640	10	14	16	Unknown
AT4G33380	10	18	19	AT4g33380/F17M5_140
AT1G06200	10	17	21	Peptidase-S24/S26 domain-containing protein; F9P14.6
AT4G23430	10	13	12	Short-chain dehydrogenase TIC 32, chloroplastic; TIC32
AT1G27460	10	14	11	NPGR1
AT1G52730	10	15	14	Unknown
AT3G06860	10	15	17	Peroxisomal fatty acid beta-oxidation multifunctional protein MFP2; MFP2
AT1G15400	10	12	16	Uncharacterized protein
AT2G32070	10	13	12	Probable CCR4-associated factor 1 homolog 7; CAF1-7
AT5G10750	10	15	13	Unknown
AT1G07440	10	20	29	Tropinone reductase homolog
AT3G08640	10	13	13	Protein RETICULATA-RELATED 3; RER3
AT3G06610	10	12	11	DNA-binding enhancer-like protein; F5E6.6
AT2G28740	10	7	7	Histone H4
AT4G05160	9	14	13	4-coumarate–CoA ligase-like 7; 4CLL7
AT1G31300	9	13	13	Unknown
AT5G66590	9	21	30	CAP (Cysteine-rich secretory proteins, Pathogenesis-related 1 protein)
AT3G51050	9	14	15	FG-GAP repeat-containing protein
AT1G28395	9	16	9	Uncharacterized protein
AT5G66850	9	11	10	Mitogen-activated protein kinase kinase kinase 5; MAPKKK5
AT1G36390	9	9	8	GrpE protein homolog; F7F23.11
AT5G42890	9	12	14	Putative uncharacterized protein ; SCP2
AT5G15870	9	6	4	Glycosyl hydrolase family 81 protein; F1N13_10
AT5G20520	9	10	10	Bem46-family protein WAV2; WAV2
AT5G45040	9	6	6	Cytochrome c6, chloroplastic; petJ
AT5G14590	9	16	14	Isocitrate dehydrogenase [NADP], chloroplastic/mitochondrial
AT3G60370	9	9	9	Peptidyl-prolyl cis-trans isomerase FKBP20-2, chloroplastic; FKBP20-2
AT4G13460	9	12	13	Histone-lysine N-methyltransferase family member SUVH9; SUVH9
AT3G12360	9	15	13	Ankyrin repeat-containing protein
AT1G12640	9	14	17	Lysophospholipid acyltransferase 1; LPLAT1
AT2G37050	9	10	10	Unknown
AT2G33255	9	13	14	Haloacid dehalogenase-like hydrolase domain-containing protein
AT1G02170	9	14	14	Metacaspase-1; AMC1
AT3G57785	9	10	13	Expressed protein
AT3G60380	9	15	16	Putative uncharacterized protein T8B10_40; T8B10_40
AT4G23460	9	11	9	Beta-adaptin-like protein C; BETAC-AD
AT4G30810	9	13	14	Serine carboxypeptidase-like 29; SCPL29
AT1G80490	9	12	11	Topless-related protein 1; TPR1
AT4G32320	9	11	11	Putative L-ascorbate peroxidase 6; APX6
AT1G78700	9	13	11	BES1/BZR1 homolog protein 4; BEH4
AT3G10360	9	11	13	Pumilio homolog 4; APUM4

Continued on next page

Table B.2 –continued from previous page

Gene ID	4d	6d	8d	Annotation
AT5G55790	9	11	13	Unknown
AT1G21920	9	15	15	Histone H3 K4-specific methyltransferase SET7/9 family protein
AT1G09150	9	15	15	Pseudouridine synthase and archaeosine transglycosylase domain-containing
AT1G51950	9	11	11	Auxin-responsive protein IAA18; IAA18
AT4G26940	8	12	14	Probable beta-1,3-galactosyltransferase 4; B3GALT4
AT3G08780	8	12	11	Unknown
AT5G06870	8	14	7	Polygalacturonase inhibitor 2; PGIP2
AT2G31660	8	10	7	Importin beta-like SAD2; SAD2
AT2G23820	8	10	8	Unknown
AT5G19180	8	11	11	NEDD8-activating enzyme E1 catalytic subunit; ECR1
AT4G10040	8	10	10	Cytochrome c-2; CYTC-2
AT1G18660	8	13	13	Zinc finger (C3HC4-type RING finger) family protein
AT4G36660	8	10	13	Unknown
AT1G77440	8	11	12	Proteasome subunit beta type-3-B; PBC2
AT3G42150	8	9	9	Unknown T27B3.20
AT5G35690	8	11	13	Putative uncharacterized protein
AT4G24740	8	16	14	Serine/threonine-protein kinase AFC2; AFC2
AT3G63260	8	14	17	ATMRK1
AT1G26750	8	11	10	Uncharacterized protein
AT2G47640	8	10	9	Putative small nuclear ribonucleoprotein D2
AT1G28340	8	13	14	Receptor like protein 4; RLP4
AT2G45260	8	10	10	Putative uncharacterized protein
AT5G64816	8	7	6	Uncharacterized protein
AT1G07810	8	8	7	Calcium-transporting ATPase 1, endoplasmic reticulum-type; ECA1
AT1G18440	8	12	11	Peptidyl-tRNA hydrolase, chloroplastic
AT1G20830	8	12	12	Multiple chloroplast division site 1; MCD1
AT2G43235	8	10	8	Unknown
AT2G31560	8	13	15	AT2G31560 protein
AT5G66410	8	8	8	Thioredoxin domain-containing protein PLP3B; PLP3B
AT3G03720	8	10	10	Cationic amino acid transporter 4, vacuolar; CAT4
AT2G18990	8	7	7	Thioredoxin domain-containing protein 9 homolog
AT5G64660	8	7	6	U-box domain-containing protein 27; PUB27
AT4G23190	8	6	4	Cysteine-rich receptor-like protein kinase 11; CRK11
AT5G40650	8	8	10	Succinate dehydrogenase ubiquinone iron-sulfur subunit 2, mitochondrial
AT1G11240	7	8	7	Uncharacterized protein
AT3G13062	7	15	23	Unknown
AT3G61150	7	7	8	Homeobox-leucine zipper protein HDG1; HDG1
AT4G29810	7	13	12	Mitogen-activated protein kinase kinase 2; MKK2
AT1G16870	7	8	6	Mitochondrial 28S ribosomal protein S29-like protein; F6I1.13
AT1G01900	7	7	5	Subtilisin-like protease SBT1.1; SBTI1.1
AT3G55630	7	7	8	Folylpolyglutamate synthase; FPGS3
AT2G41960	7	9	8	Putative uncharacterized protein
AT3G44150	7	11	8	Putative uncharacterized protein
AT1G61010	7	9	8	Cleavage and polyadenylation specificity factor subunit 3-I; CPSF73-I
AT3G26630	7	10	9	Pentatricopeptide repeat-containing protein , chloroplastic; PCMP-A6
AT1G53170	7	9	9	Ethylene-responsive transcription factor 8; ERF8
AT5G18620	7	6	4	Chromatin remodeling factor17; CHR17
AT2G34357	7	8	5	NUC173 domain-containing protein
AT3G11730	7	9	10	Ras-related protein RABD1; RABD1
AT1G29260	7	10	9	Peroxisome biogenesis protein 7; PEX7
AT1G68780	7	10	12	Unknown F14K14.11
AT3G56760	7	8	7	CDPK-related kinase 7; CRK7
AT1G63010	7	11	12	SPX domain-containing membrane protein
AT2G03690	7	6	8	Ubiquinone biosynthesis protein COQ4 homolog, mitochondrial
AT2G14170	7	10	12	Methylmalonate-semialdehyde dehydrogenase acylating, mitochondrial
AT1G56590	7	8	7	AP-3 complex subunit mu; AP3M
AT5G01960	7	10	9	Putative uncharacterized protein
AT5G63040	7	12	11	Unknown
AT2G45200	7	10	9	Golgi SNAP receptor complex member 1-2; GOS12
AT5G63910	7	9	11	Farnesylcysteine lyase; FLCY
AT4G21445	7	7	8	Unknown PRIP
AT1G77080	7	10	11	Agamous-like MADS-box protein AGL27; AGL27
AT3G01510	7	12	11	Phosphoglucan phosphatase LSF1, chloroplastic; LSF1

Continued on next page

Table B.2 –continued from previous page

Gene ID	4d	6d	8d	Annotation
AT2G46450	7	11	12	Probable cyclic nucleotide-gated ion channel 12; CNGC12
AT1G42540	7	8	10	Glutamate receptor 3.3; GLR3.3
AT5G60570	7	8	9	F-box/kelch-repeat protein
AT5G47640	7	13	15	Nuclear transcription factor Y subunit B-2; NFYB2
AT1G33980	7	8	9	Regulator of nonsense transcripts UPF3; UPF3
AT2G39480	7	8	7	ABC transporter B family member 6; ABCB6
AT3G13080	7	6	6	ABC transporter C family member 3; ABCC3
AT4G19450	7	10	14	Major facilitator protein
AT1G53190	7	11	11	Unknown
AT3G55000	7	8	7	Protein TONNEAU 1a; TON1A
AT2G26210	7	12	14	Ankyrin repeat family protein
AT1G10720	7	6	6	BSD domain-containing protein; T16B5.14
AT2G44065	6	8	7	Expressed protein
AT3G05000	6	8	7	Putative uncharacterized protein ; T9J14.5
AT1G62422	6	8	7	F2401.15
AT4G01320	6	6	6	CAAX prenyl protease 1 homolog; FACE1
AT5G10910	6	8	6	AT5g10910/T30N20_180; T30N20_180
AT4G33410	6	9	8	Signal peptide peptidase-like 1; SPPL1
AT1G20100	6	7	4	Putative uncharacterized protein ; T20H2.11
AT1G78180	6	10	10	Probable mitochondrial adenine nucleotide transporter BTL2
AT1G11360	6	9	9	AT1G11360 protein
AT4G39690	6	9	9	AUnknown
AT4G28480	6	9	8	DNAJ heat shock protein; F20O9.160
AT4G02060	6	6	4	DNA replication licensing factor MCM7; MCM7
AT1G53480	6	6	2	Mto 1 responding down 1 protein; T3F20.20
AT1G65410	6	8	8	Protein TRIGALACTOSYLDIACYLGLYCEROL 3, chloroplastic; TGD3
AT5G53940	6	11	11	Protein yippee-like
AT3G52940	6	6	6	Delta(14)-sterol reductase; FK
AT2G19950	6	7	7	Golgin candidate 1; GC1
AT3G60440	6	9	9	Phosphoglycerate mutase family protein
AT3G24350	6	10	11	Syntaxin-32; SYP32
AT2G32970	6	9	8	Uncharacterized protein
AT3G02770	6	10	10	Putative 4-hydroxy-4-methyl-2-oxoglutarate aldolase 1
AT1G08125	6	6	6	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT5G40250	6	9	8	RING-H2 finger protein ATL46; ATL46
AT3G15410	6	7	6	Leucine-rich repeat-containing protein
AT1G61870	6	7	4	Pentatricopeptide repeat-containing protein , mitochondrial; PPR336
AT4G24290	6	11	12	MACPF domain-containing protein
AT5G14370	6	8	7	CCT motif family protein
AT1G04190	6	6	6	Carboxylate clamp-tetratricopeptide repeat protein
AT5G57610	6	10	11	Octicosapeptide/Phox/Bem1p domain-containing protein kinase
AT5G58300	6	7	5	Probable inactive receptor kinase
AT2G04850	6	11	10	Cytochrome b561 and DOMON domain-containing protein
AT3G47910	6	9	8	Ubiquitin carboxyl-terminal hydrolase-related protein
AT1G67890	6	7	7	PAS domain-containing protein tyrosine kinase
AT3G12130	6	9	9	Zinc finger CCCH domain-containing protein 36
AT1G64050	6	8	6	Unknown
AT2G44360	6	11	11	Putative uncharacterized protein
AT5G51200	6	6	5	Nuclear pore complex protein NUP205; NUP205
AT1G19130	6	11	12	Putative uncharacterized protein /F14D16_18
AT1G16000	6	6	5	GAG1
AT3G01770	6	6	6	Transcription factor GTE11; GTE11
AT5G19500	6	10	11	Unknown
AT3G63070	6	7	6	HUA2-like protein 2
AT5G39500	6	8	8	ARF guanine-nucleotide exchange factor GNL1; GNL1
AT3G52950	6	8	7	CBS domain-containing protein CBSCBSPB3; CBSCBSPB3
AT1G22400	6	7	4	UDP-glycosyltransferase 85A1; UGT85A1
AT3G50360	5	6	7	Probable calcium-binding protein CML20; CML20
AT1G49670	5	10	11	Unknown
AT1G47900	5	6	5	Filament-like plant protein 6; FPP6
AT1G55152	5	7	9	Uncharacterized protein
AT1G30330	5	8	8	Auxin response factor 6; ARF6
AT2G14120	5	8	7	Dynamin-related protein 3B; DRP3B

Continued on next page

Table B.2 –continued from previous page

Gene ID	4d	6d	8d	Annotation
AT1G28350	5	7	5	Tyrosine–tRNA ligase 2, cytoplasmic
AT2G47850	5	8	8	Zinc finger CCCH domain-containing protein 32
AT4G15475	5	6	6	F-box/LRR-repeat protein 4; FBL4
AT3G04140	5	8	13	Ankyrin repeat-containing protein; T6K12.24
AT5G54300	5	6	6	Cotton fiber expressed protein 1-like protein
AT1G13190	5	8	6	Unknown
AT1G11280	5	7	6	G-type lectin S-receptor-like serine/threonine-protein kinase
AT2G32170	5	7	8	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT1G76405	5	7	5	Outer envelope pore protein 21B, chloroplastic; OEP21B
AT1G75270	5	6	7	Glutathione S-transferase DHAR2; DHAR2
AT4G28220	5	6	6	External alternative NAD(P)H-ubiquinone oxidoreductase B1, mitochondrial; NDB1
AT5G05140	5	6	5	Probable mediator of RNA polymerase II transcription subunit 26b; MED26B
AT5G62630	5	9	10	HIPL2 protein; HIPL2
AT3G55050	5	6	5	Probable protein phosphatase 2C 48; PP2C6
AT2G38790	5	9	7	Putative uncharacterized protein
AT5G61840	5	9	8	Probable beta-1,4-xylosyltransferase IRX10L; IRX10L
AT1G72320	5	6	5	Pumilio homolog 23; APUM23
AT2G19180	5	9	9	Expressed protein
AT3G23900	5	6	6	RNA recognition motif (RRM)-containing protein
AT3G24770	5	12	10	CLAVATA3/ESR (CLE)-related protein 41; CLE41
AT5G06440	5	7	9	Uncharacterized protein
AT3G05520	5	6	7	F-actin-capping protein subunit alpha
AT5G02800	5	10	10	Serine/threonine-protein kinase CDL1; CDL1
AT3G04080	5	6	6	Apyrase 1; APY1
AT3G07780	5	6	5	Protein OBERON 1; OBE1

To investigate whether the regulation and function of gene orthologs is conserved between maize and arabidopsis, arabidopsis orthologs were inferred by Dr D. Emms using Orthofinder software (Emms & Kelly, 2015) for each of the maize AIL target genes identified in section 4.2.7 (p.143). All of these arabidopsis genes were examined for the presence of the ANT consensus region in a 1 kbp region upstream of the predicted ORF. These arabidopsis orthologous genes were then divided into three groups depending on expression characteristics and whether they had an ANT consensus region upstream of the ORF: group 1 contained 37 arabidopsis orthologs that were AIL targets and were expressed in leaf development (so their regulation and expression pattern was similar to the corresponding maize gene); group 2 contained 125 arabidopsis orthologs that were expressed in leaf development but were not AIL targets (so there expression pattern was similar to maize, but the regulation of gene transcription was not conserved); and group 3 contained 174 genes that were either not expressed during arabidopsis leaf development or the maize gene did not have an arabidopsis ortholog. Two putative Kranz regulator orthologs were identified as having similar expression and AIL regulation properties between maize and arabidopsis. One a putative negative regulator (*GRMZM2G016480*, a programmed cell death interacting protein) and one was a predicted positive regulator (*GRMZM2G171852*; *DLK1H*, DOF family transcription factor).

GO term analysis was conducted on the arabidopsis orthologs, to investigate whether the differences in regulation or expression were in specific biological processes or molecular functions. 11% of gene orthologs between maize and arabidopsis showed conserved expression patterns in leaves and in terms of predicted AIL regulation, and these included all the genes annotated with a role in cell differentiation and cell growth (fig. B.3; p.255). 37% of the gene orthologs were expressed in arabidopsis leaves but did not have a conserved AIL target site, suggesting that these genes have important roles in leaf development in

maize and arabidopsis, but are regulated differently. This group included all the genes annotated as being involved in protein localisation, and the amino acid transporters, lipid, carbohydrate and nucleotide binding protein encoding genes (fig. B.4; p.256). There were also many maize AIL targets that did not have any orthologs in arabidopsis. These data suggest that, in general, AIL regulation of gene expression during leaf development is not highly conserved between arabidopsis and maize, but genes involved in cell growth and differentiation may have a conserved function and regulation in the two species.

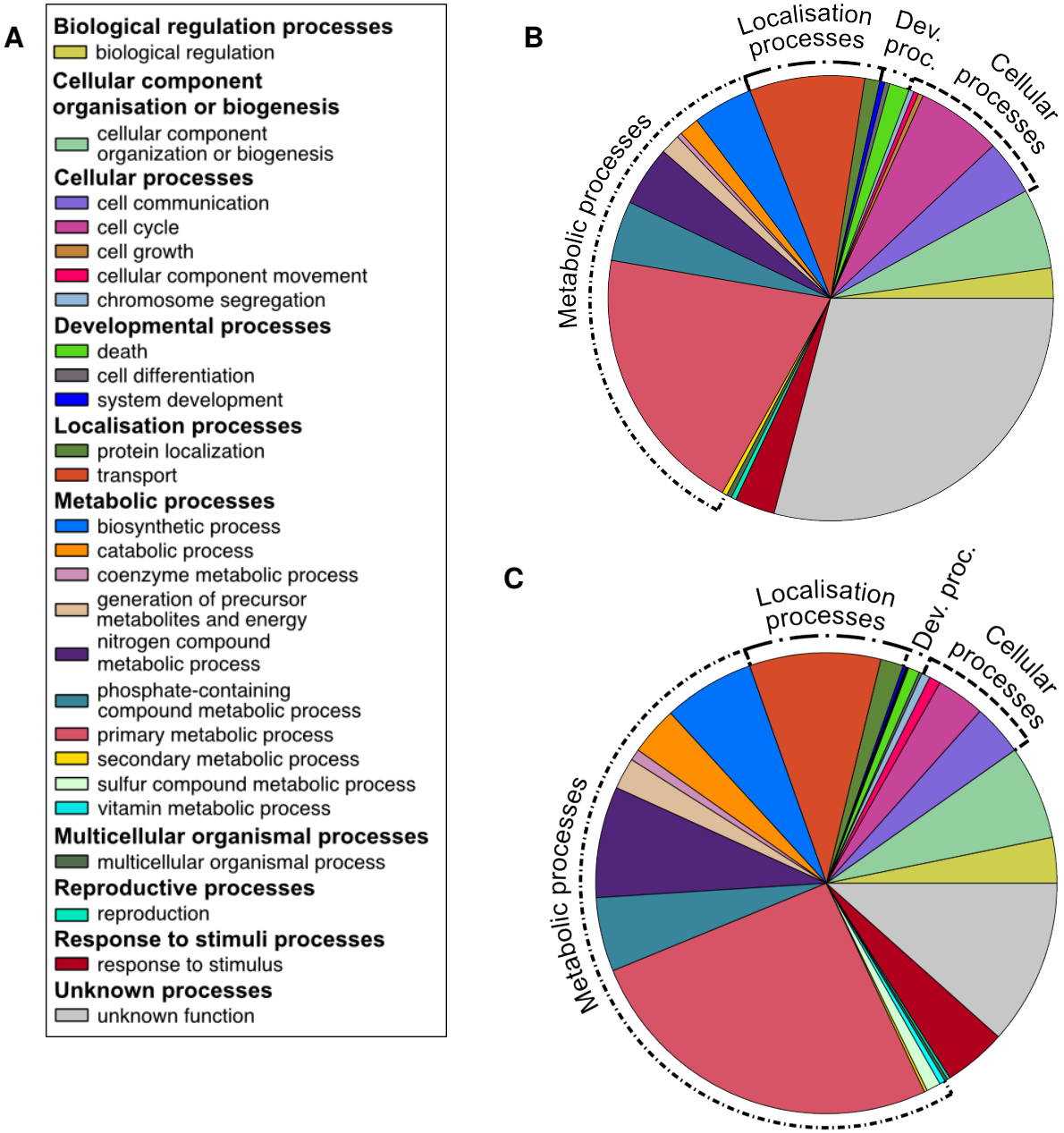


Figure B.1: Biological process gene ontology of putative AIL protein target genes expressed during maize and arabidopsis leaf development.

A) Categories of gene ontology (GO) annotations related to biological processes present in the putative AIL target gene lists. **B-C)** Pie charts showing biological process gene ontology of putative AIL target genes expressed during leaf development in- maize (262 genes) (B) and arabidopsis (453 genes) (C).

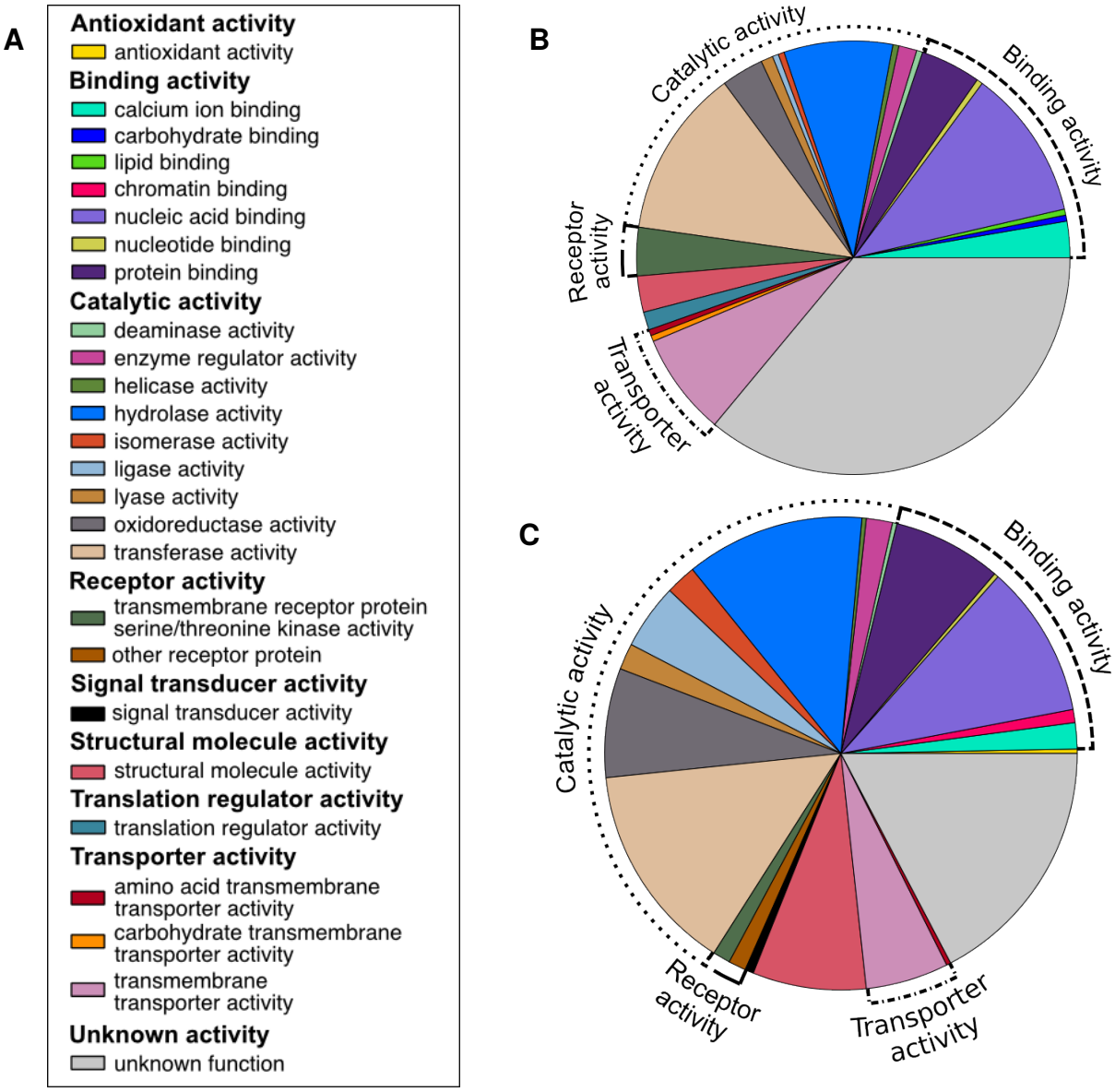


Figure B.2: Molecular function gene ontology of putative AIL protein target genes expressed during maize and arabidopsis leaf development.

A) Categories of gene ontology (GO) annotations related to molecular function present in the putative AIL target gene lists. **B-C)** Pie charts showing molecular function gene ontology of putative AIL target genes expressed during leaf development in- maize (262 genes) (B) and arabidopsis (453 genes) (C).

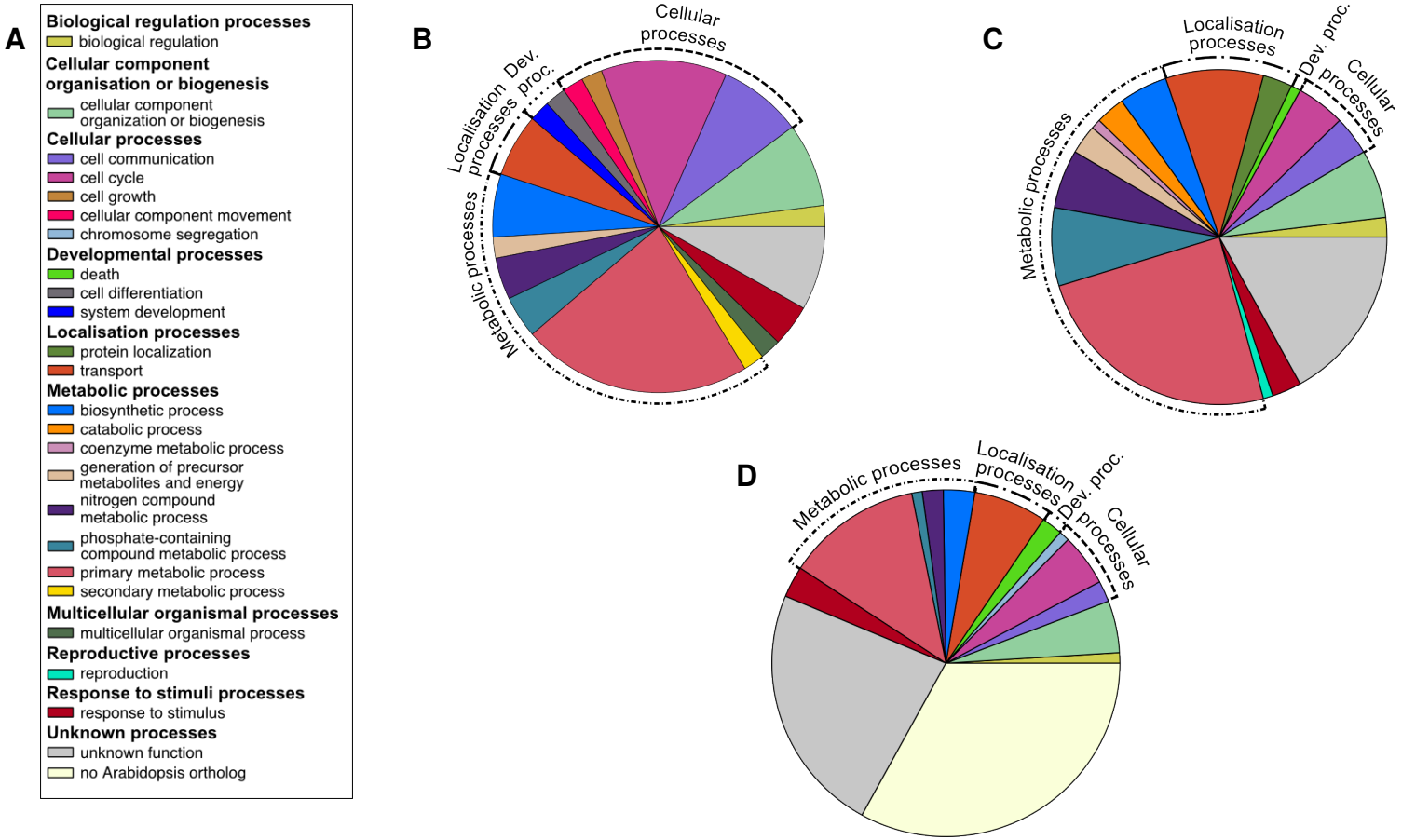


Figure B.3: Biological process gene ontology of arabidopsis orthologs to maize AIL target genes, grouped by conservation of expression and regulation in leaf development. A) Gene ontology (GO) annotation categories related to biological processes for AIL target genes. Pie charts of GO terms for arabidopsis genes that show **B-C)** conserved expression- and AIL regulation (37 genes) (B), but no AIL regulation (125 genes) (C), **D)** no leaf expression or there was no orthology (174 genes), compared to maize AIL target genes.

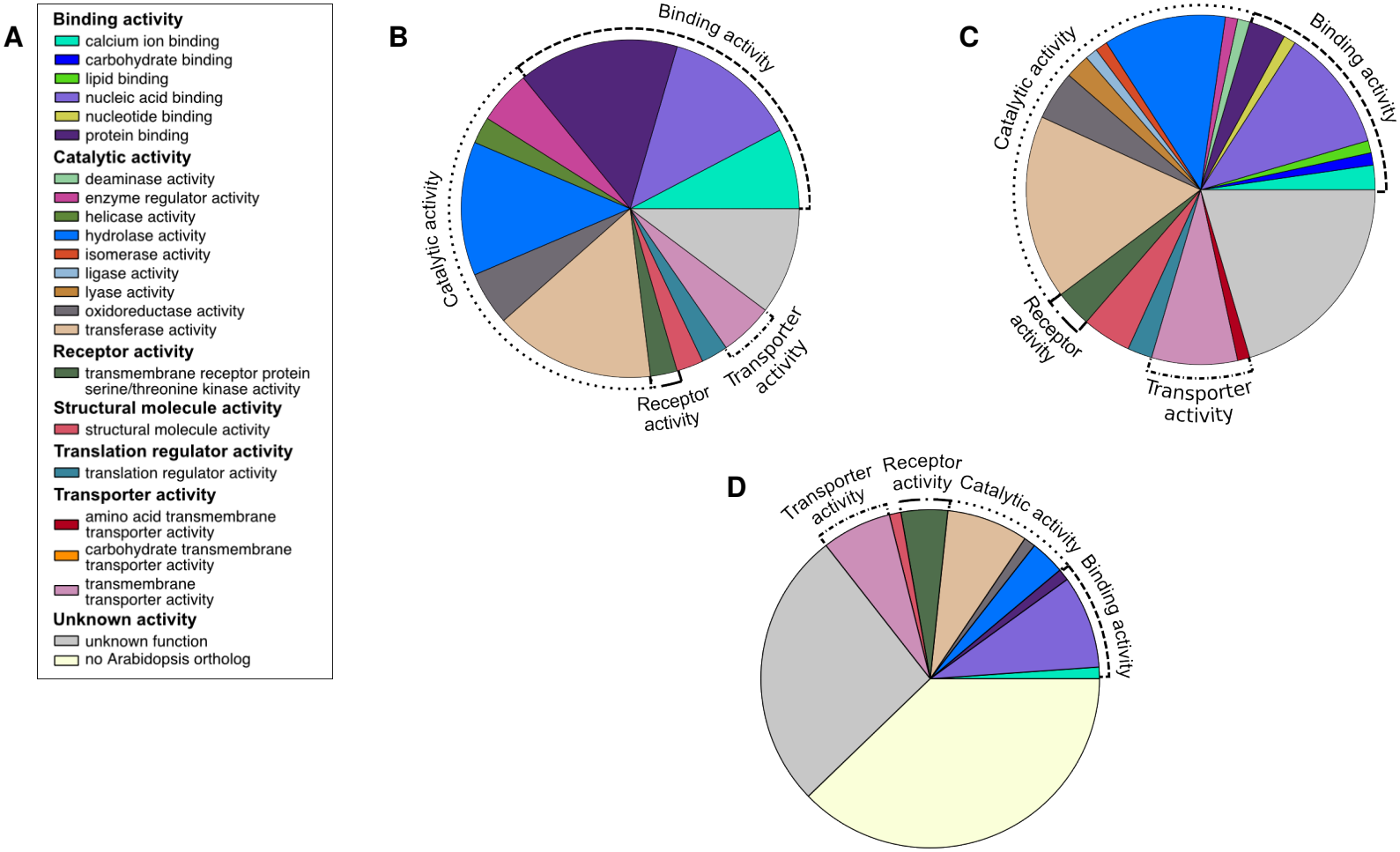


Figure B.4: Biological process gene ontology of arabidopsis orthologs to maize AIL target genes, grouped by conservation of expression and regulation in leaf development. A) Gene ontology (GO) annotation categories related to molecular functions of AIL target genes. Pie charts of GO terms for arabidopsis genes that show **B-C)** conserved expression- and AIL regulation (37 genes) (B), but no AIL regulation (125 genes) (C), **D)** no leaf expression or there was no orthology (174 genes), compared to maize AIL target genes.

B.3 Preliminary results: OsAIL5 promotes vein initiation in rice

ZmAIL5 was identified as a putative negative regulator of Kranz anatomy in that gene expression in foliar leaves was downregulated relative to husk leaves during development. It was therefore predicted that downregulation of the ortholog *OsAIL5* in rice may result in Kranz-like characteristics. This was tested using RNAi in rice.

B.3.0.1 *ZmUBI$_{pro}$::OsAIL5i* transformation gave multiple regenerants but low survival rates and seed set

To investigate the effect of suppressing OsAIL5 activity on rice leaf development, a 305 bp region of the gene was cloned into a hairpin RNAi vector. A BLAST search was used to ensure that the DNA region chosen did not contain any contiguous 21 bp regions that were also present in other rice genes, thereby minimising the risk of off-target gene silencing. The *ZmUBI$_{pro}$::OsAIL5i* construct was transformed into wild-type rice callus, with the same procedure as in section 4.2.5.1 (p.130). Although nine lines regenerated, only three lines had good seed set (table B.3; p.258).

B.3.0.2 Surviving *UBI$_{pro}$::OsAIL5i* lines showed no or low levels of gene silencing

T_2 individuals from lines *OsAIL5i-1*, *OsAIL5i-3* and *OsAIL5i-4* were advanced for further testing. All individuals from line *OsAIL5i-1* showed poor growth which was not correlated with transgene presence, and therefore this line was not taken forward for expression analysis (data not shown). Lines *OsAIL5i-3* and *OsAIL5i-4* were genotyped (fig. B.5, A; p.259) and did not show a significant deviation from a

3:1 segregation of the transgene, tested using a chi-square test of independence (*UBI_{pro}::OsAIL5i-3*: $\chi^2(1, N = 14) = 0.2$, $p>0.05$; *UBI_{pro}::OsAIL5i-4*: $\chi^2(1, N = 10) = 0.8$, $p>0.05$), indicating single insertion events. To test for gene silencing, semi-Q RT-PCR was carried out with cDNA samples reverse-transcribed using RNA samples from P1-P5 of young tillers (fig. B.5, B; p.259). *UBIQUITIN* expression was used as the normalising control, and *OsAIL5* gene expression was tested. Very little gene silencing was seen, with only three individuals in the *OsAIL5i-3* line displaying some suppression of gene expression.

B.3.0.3 Suppression of OsAIL5 activity results in a trend for lower vein density

A characteristic of Kranz anatomy is an increased frequency of veins across the leaf blade compared to non-Kranz leaves. To determine whether suppression of *OsAIL5* gives Kranz-like vein patterning in rice, leaf phenotypes were examined in siblings from an RNAi line that accumulated either normal or reduced levels of *OsAIL5* transcripts. Plants with reduced *OsAIL5* transcript levels showed decreased vein density (significant to $p<0.1$) due to fewer veins per leaf, without a corresponding decrease in leaf width (fig. B.6, A-C; p.261). A decrease in vein

Line	Transgene	T0 survival	T0 seed set
OsAIL5i-1	Yes	Yes	Good
OsAIL5i-2	Yes	Yes	Poor
OsAIL5i-3	Yes	Yes	Good
OsAIL5i-4	Yes	Yes	Good
OsAIL5i-5	Yes	Yes	Poor
OsAIL5i-6	Yes	No	N/A
OsAIL5i-7	Yes	Yes	Poor
OsAIL5i-8	Yes	No	N/A
OsAIL5i-9	Yes	No	N/A

Table B.3: *UBI_{pro}1::OsAIL5i* transformation regeneration summary.

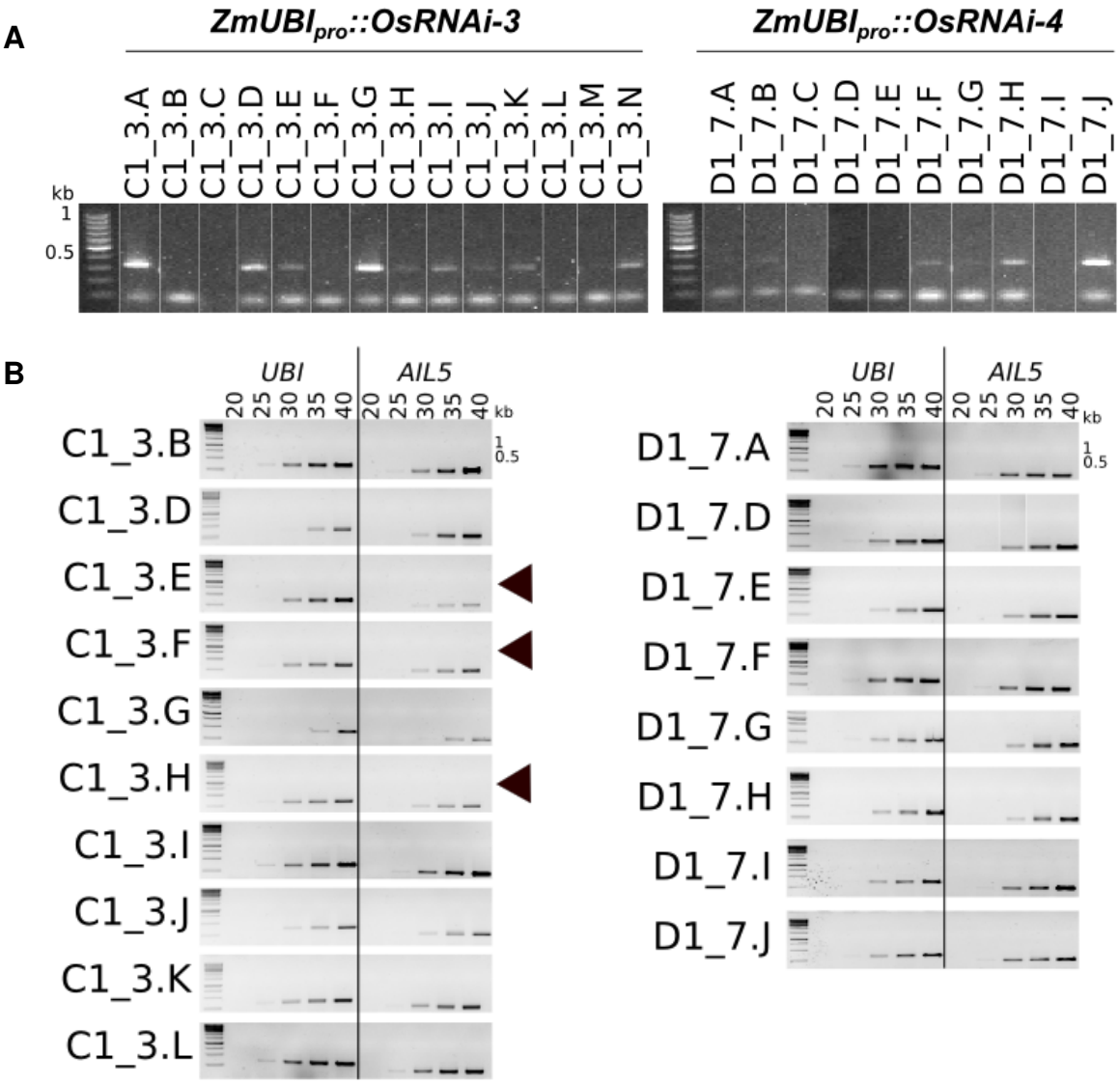


Figure B.5: Composite of gel images showing *ZmUBI$_{pro}$::OsAIL5i* genotyping and expression analysis. T_2 individuals from two *ZmUBI$_{pro}$::OsAIL5i* lines were analysed for- **A)** Transgene insertion using PCR on genomic DNA. **B)** *OsAIL5* expression using semi-quantitative RT-PCR from a subset of genotyped individuals. *OsUBI* amplification was used as a normalising control and DNA amplification was measured after 20, 25, 30, 35 and 40 PCR cycles. Each PCR well ID is a unique identifier for one individual which corresponds with the RT-PCR ID. ID code refers to independent transformation events (C1/D1), the T_1 family (.#) and the T_2 individual in the family (last letter). Black arrows indicate individuals with reduced *OsAIL5* transcript levels.

density can be caused by an increase in mesophyll cell size or number, or by an increase in vascular bundle size. To evaluate these possibilities, measurements were made between the second and third lateral veins (counting out from the mid-rib). Notably, there was no significant difference in mesophyll cell number or size (fig. B.6, E-F), or in vascular bundle size (fig. B.7, A; p.262). As such, it must be concluded that the observable decrease in vein density was caused by differences in either the midrib or margin regions of the leaf.

Vascular bundle characteristics were also measured in the RNAi line siblings. To discover whether decreased *OsAIL5* transcript levels perturb vein development, intermediate vascular bundle size, number of cells, and bundle sheath cell width measurements were taken from veins located between the second and third lateral veins (counting out from the mid-rib) and lateral vein heights were taken from across the whole leaf. No significant differences were identified between lines with normal and reduced *OsAIL5* transcript levels (fig. B.7, A-D; p.262).

B.3.0.4 *OsAIL5* summary

In summary, the *OsAIL5i* RNAi construct impacted plant survival and seed set, and there may have been selective pressure against gene silencing. Rice plants that showed suppression of *OsAIL5* expression displayed reduced vein initiation. These data suggest that the protein is an important positive regulator of vein initiation during leaf development, and may play a role in floral development.

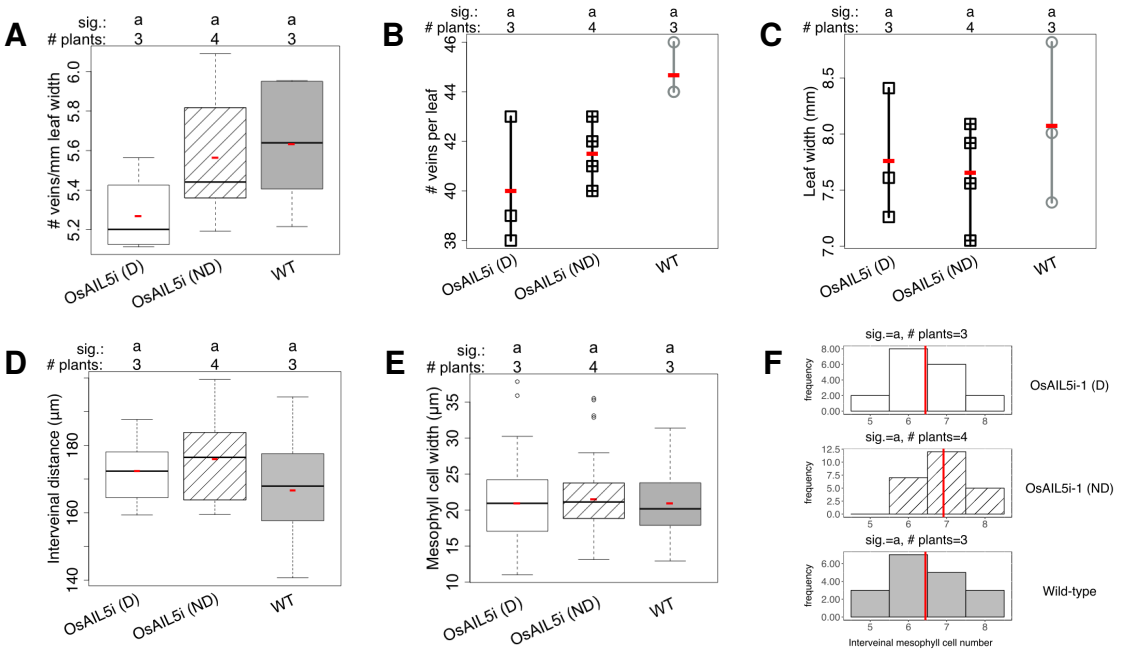


Figure B.6: Reduced vein density trend in the *OsAIL5i-3* RNAi rice line. A) Boxplot showing vein density (veins/mm leaf width). **B-C)** Strip charts showing-number of veins per leaf (B) and leaf width (mm) (C). **D-E)** Boxplots showing-distance between intermediate veins (µm) (D) and mesophyll cell widths (µm) (E). **F)** Histogram showing number of mesophyll cells between adjacent intermediate veins. "D"- transgenic plants with knock-down of *OsAIL5*, "ND"- transgenic plants with WT *OsAIL5* levels. "Sig."- different letters indicate significant differences in group means ($p<0.05$). # plants- number of individual plants used for analysis. **Boxplots**: Box- data within 25^{th}-75^{th} percentiles, with black horizontal line- median, red dash- mean, whiskers- lowest and highest value still within 1.5 IQR, open circle- value outside 1.5 IQR. **Strip charts**: Square/circle- individual data point, vertical line- range of data, red dash- mean. **Histograms**: Red vertical line-mean.

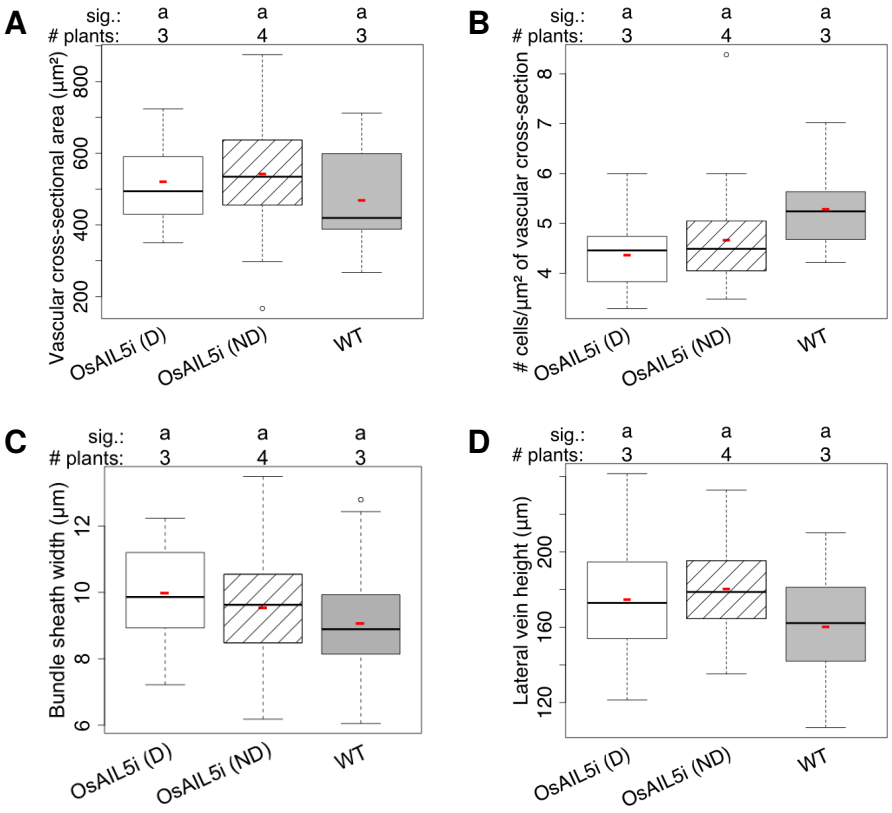


Figure B.7: No significant vascular bundle differences in the *OsAIL5i-3* RNAi transgenic rice line. Boxplots showing- **A)** intermediate vascular bundle cross-sectional area (µm^2), **B)** cell number per µm^2 cross-sectional intermediate vein area, **C)** bundle sheath cell widths around intermediate veins (µm), **D)** height of lateral veins (µm). "D"- transgenic plants with knock-down of *OsAIL5*, "ND"- transgenic plants with WT *OsAIL5* levels. "Sig."- different letters indicate significant differences in group means ($p<0.05$). # plants- number of individual plants used for analysis. **Boxplots**: Box- data within 25th-75th percentiles, with black horizontal line- median, red dash- mean, whiskers- lowest and highest value still within 1.5 IQR, open circle- value outside 1.5 IQR.

C

Digital appendix: Materials and methods supplementary data

The tables listed below can be found in the file 'Digital_appendix_C.xlsx'. A CD is included at the back of this thesis.

C.1 Gene cloning for plant transformation

C.1.1 Table C.1.1. Gene and primer details for Gateway gene cloning

C.1.2 Table C.1.2. Gateway and Level 2 Golden Gate plasmid details and resistance markers

C.1.3 Table C.1.3. Gateway plasmid Sanger sequencing primers

C.1.4 Table C.1.4. Golden Gate plasmids

C.1.5 Table C.1.5. Colony PCR, genotyping and expression analysis primer pairs

C.2 In situ hybridisation

C.2.1 Table C.2.1. *In situ* probe synthesis PCR primers, RNA probe hydrolysis time and SSC wash concentration

C.2.2 Table C.2.2 Tissue-Tek VIP 5 vacuum infiltration processor cycle details

References

Ahmad, M., Rim, Y., Chen, H. & Kim, J.Y. (2012). Functional characterization of *Arabidopsis* Dof transcription factor AtDof4.1. *Russian Journal of Plant Physiology*, 60, 116–123.

Aida, M., Beis, D., Heidstra, R., Willemsen, V., Blilou, I., Galinha, C., Nussaume, L., Noh, Y.S., Amasino, R. & Scheres, B. (2004). The PLETHORA Genes Mediate Patterning of the *Arabidopsis* Root Stem Cell Niche. *Cell*, 119, 109–120.

Anne, P., Azzopardi, M., Gissot, L., Beaubiat, S., Hématy, K. & Palauqui, J.C. (2016). OCTOPUS Negatively Regulates BIN2 to Control Phloem Differentiation in *Arabidopsis thaliana*. *Current Biology*, 25, 2584–2590.

Antre, A., Bhat, R. & Pushpalatha, N. (2013). An Ac/Ds-mediated Insertional Mutagenesis System for Functional Genomics in Sorghum. *Molecular Plant Breeding*, 4, 220–227.

Appelhagen, I., Lu, G.H., Huep, G., Schmelzer, E., Weisshaar, B. & Sagasser, M. (2011). TRANSPARENT TESTA1 interacts with R2R3-MYB factors and affects early and late steps of flavonoid biosynthesis in the endothelium of *Arabidopsis thaliana* seeds. *Plant Journal*, 67, 406–419.

Arruda, P. & Gerhardt, I.R. (2016). Increasing cell wall deposition and biomass density in plants.

Ashburner, M., Ball, C A., Blake, J.A., Botstein, D., Butler, H ., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M. & Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 25, 25–29.

Aubry, S., Brown, N.J. & Hibberd, J.M. (2011). The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway. *Journal of experimental botany*, 62, 3049–3059.

Baima, S., Nobili, F., Sessa, G., Lucchetti, S., Ruberti, I. & Morelli, G. (1995). The expression of the ATHB-8 homeobox gene is restricted to provascular cells in *Arabidopsis thaliana*. *Development*, 121, 4171–4182.

Baima, S., Possenti, M., Matteucci, A., Wisman, E., Altamura, M.M., Ruberti, I. & Morelli, G. (2001). The *Arabidopsis* ATHB-8 HD-Zip Protein Acts as a Differentiation-Promoting Transcription Factor of the Vascular Meristems. *Plant Physiology*, 126, 643–655.

Baumann, K. (1999). The DNA Binding Site of the Dof Protein NtBBF1 Is Essential for Tissue-Specific and Auxin-Regulated Expression of the rolB Oncogene in Plants. *The Plant Cell*, 11, 323–334.

van den Bergh, E., Külahoglu, C., Bräutigam, A., Hibberd, J.M., Weber, A.P.M., Zhu, X.G. & Eric Schranz, M. (2014). Gene and genome duplications and the origin of C_4 photosynthesis: Birth of a trait in the Cleomaceae. *Current Plant Biology*, 1, 2–9.

Berry, J.A., Downton, W.J.S. & Tregunna, E.B. (1970). The photosynthetic carbon metabolism of *Zea mays* and Gomphrena globosa: the location of the CO_2 fixation and the carboxyl transfer reactions. *Canadian Journal of Botany*, 48, 777–786.

Bishopp, A., Help, H., El-Showk, S., Weijers, D., Scheres, B., Friml, J., Benková, E., Mähönen, A. & Helariutta, Y. (2011). A Mutually Inhibitory Interaction between Auxin and Cytokinin Specifies Vascular Pattern in Roots. *Current Biology*, 21, 917–926.

Björkman, O. (1968). Carboxydismutase Activity in Shade-adapted and Sun-adapted Species of Higher Plants. *Physiologia Plantarum*, 21, 1–10.

Bonke, M., Thitamadee, S., Mahonen, A.P., Hauser, M.T. & Helariutta, Y. (2003). APL regulates vascular tissue identity in *Arabidopsis*. *Nature*, 426, 181–186.

Bräutigam, A., Kajala, K., Wullenweber, J., Sommer, M., Gagneul, D., Weber, K.L., Carr, K.M., Gowik, U., Maß, J., Lercher, M.J., Westhoff, P., Hibberd, J.M. & Weber, A.P.M. (2011). An mRNA Blueprint for C_4 Photosynthesis Derived from Comparative Transcriptomics of Closely Related C_3 and C_4 Species. *Plant Physiology*, 155, 142–156.

Brown, R.H., Bouton, J.H., Rigsby, L. & Rigler, M. (1983). Photosynthesis of Grass Species Differing in Carbon Dioxide Fixation Pathways : VIII. Ultrastructural Characteristics of Panicum Species in the Laxa Group. *Plant physiology*, 71, 425–431.

Brown, W.V. (1958). Leaf Anatomy in Grass Systematics. *Botanical Gazette*, 119, 170–178.

Brown, W.V. (1975). Variations in anatomy, associations, and origins of Kranz tissue. *American Journal of Botany*, pp. 395–402.

Brutnell, T.P., Sawers, R.J., Mant, A. & Langdale, J.A. (1999). BUNDLE SHEATH DEFECTIVE2, a novel protein required for post-translational regulation of the rbcL gene of maize. *The Plant cell*, 11, 849–864.

van Campen, J., Yaapar, M.N., Narawatthana, S., Lehmeier, C., Wanchana, S., Thakur, V., Chater, C., Kelly, S., Rolfe, S.A., Quick, W.P. & Fleming, A.J. (2016). Combined Chlorophyll Fluorescence and Transcriptomic Analysis Identifies the P3/P4 Transition as a Key stage in Rice Leaf Photosynthetic Development. *Plant physiology*, 170, 1655–1674.

Carland, F. & Nelson, T. (2009). CVP2- and CVL1-mediated phosphoinositide signaling as a regulator of the ARF GAP SFC/VAN3 in establishment of foliar vein patterns. *The Plant Journal*, 59, 895–907.

Carlsbecker, A., Lee, J.Y., Roberts, C.J., Dettmer, J., Lehesranta, S., Zhou, J., Lindgren, O., Moreno-Risueno, M.A., Vatén, A., Thitamadee, S., Campilho, A., Sebastian, J., Bowman, J.L., Helariutta, Y. & Benfey, P.N. (2010). Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. *Nature*, 465, 316–321.

Chandler, J.W. & Werr, W. (2015). Cytokinin-auxin crosstalk in cell type specification. *Trends in Plant Science*, 20, 291–300.

Chang, Y.M., Liu, W.Y., Shih, A.C.C., Shen, M.N., Lu, C.H., Lu, M.Y.J., Yang, H.W., Wang, T.Y., Chen, S.C.C., Chen, S.M., Li, W.H. & Ku, M.S.B. (2012). Characterizing Regulatory and Functional Differentiation between Maize Mesophyll and Bundle Sheath Cells by Transcriptomic Analysis. *Plant Physiology*, 160, 165–177.

Chen, H., Ahmad, M., Rim, Y., Lucas, W.J. & Kim, J.Y. (2013). Evolutionary and molecular analysis of Dof transcription factors identified a conserved motif for intercellular protein trafficking. *The New phytologist*, 198, 1250–60.

Cho, S.H., Yoo, S.C., Zhang, H., Pandeya, D., Koh, H.J., Hwang, J.Y., Kim, G.T. & Paek, N.C. (2013). The rice *narrow leaf2* and *narrow leaf3* loci encode WUSCHEL-related homeobox 3A (OsWOX3A) and function in leaf, spikelet, tiller and lateral root development. *New Phytologist*, 198, 1071–1084.

Christin, P.A., Arakaki, M., Osborne, C.P. & Edwards, E.J. (2015). Genetic enablers underlying the clustered evolutionary origins of C_4 photosynthesis in angiosperms. *Molecular Biology and Evolution*, 32, 846–858.

Christin, P.A., Besnard, G., Samaritani, E., Duvall, M.R., Hodkinson, T.R., Savolainen, V. & Salamin, N. (2008). Oligocene CO_2 Decline Promoted C_4 Photosynthesis in Grasses. *Current Biology*, 18, 37–43.

Christin, P.A., Osborne, C.P., Chatelet, D.S., Columbus, J.T., Besnard, G., Hodkinson, T.R., Garrison, L.M., Vorontsova, M.S. & Edwards, E.J. (2013). Anatomical enablers and the evolution of C_4 photosynthesis in grasses. *Proceedings of the National Academy of Sciences*, 110, 1381–1386.

Christin, P.A., Salamin, N., Kellogg, E.A., Vicentini, A. & Besnard, G. (2009). Integrating phylogeny into studies of C_4 variation in the grasses. *Plant physiology*, 149, 82–87.

Cigan, M.A., Unger-Wallace, E. & Haug-Collet, K. (2005). Transcriptional gene silencing as a tool for uncovering gene function in maize. *The Plant Journal*, 43, 929–940.

Coenen, C. & Lomax, T.L. (1997). Auxin-cytokinin interactions in higher plants: old problems and new tools. *Trends in Plant Science*, 2, 351–356.

Collingridge, P.W. & Kelly, S. (2012). MergeAlign: improving multiple sequence alignment performance by dynamic reconstruction of consensus multiple sequence alignments. *BMC Bioinformatics*, 13, 1–10.

Covshoff, S. & Hibberd, J.M. (2012). Integrating C_4 photosynthesis into C_3 crops to increase yield potential. *Current Opinion in Biotechnology*, 23, 209–214.

Cowperthwaite, M., Park, W., Xu, Z., Yan, X., Maurais, S.C. & Dooner, H.K. (2002). Use of the transposon Ac as a gene-searching engine in the maize genome. *The Plant cell*, 14, 713–726.

Cox, K.H., DeLeon, D.V., Angerer, L.M. & Angerer, R.C. (1984). Detection of mRNAs in sea urchin embryos by *in situ* hybridization using asymmetric RNA probes. *Developmental biology*, 101, 485–502.

Cribb, L., Hall, L.N. & Langdale, J.A. (2001). Four mutant alleles elucidate the role of the G2 protein in the development of C_4 and C_3 photosynthesizing maize tissues. *Genetics*, 159, 787–797.

Cui, H. & Benfey, P.N. (2009). Interplay between SCARECROW, GA and LIKE HETEROCHROMATIN PROTEIN 1 in ground tissue patterning in the *Arabidopsis* root. *The Plant Journal*, 58, 1016–1027.

Cui, H., Kong, D., Liu, X. & Hao, Y. (2014). SCARECROW, SCR-LIKE 23 and SHORT-ROOT control bundle sheath cell fate and function in *Arabidopsis thaliana*. *The Plant Journal*, 78, 319–27.

Cui, H., Levesque, M.P., Vernoux, T., Jung, J.W., Paquette, A.J., Gallagher, K.L., Wang, J.Y., Blilou, I., Scheres, B. & Benfey, P.N. (2007). An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. *Science*, 316, 421–5.

Davis, M.W. (2013). ApE Plasmid Editor. www.biologylabs.utah.edu/jorgensen/wayned/ape/

De Rybel, B., Mahonen, A.P., Helariutta, Y. & Weijers, D. (2016). Plant vascular development: from early specification to differentiation. *Nature Reviews Molecular Cell Biology*, 17, 30–40.

De Rybel, B., Möller, B., Yoshida, S., Grabowicz, I., de Reuille, P.B., Boeren, S., Smith, R.S., Borst, J.W. & Weijers, D. (2013). A bHLH complex controls embryonic vascular tissue establishment and indeterminate growth in *Arabidopsis*. *Developmental Cell*, 24, 426–437.

Deng, W. & Roberts, S.G.E. (2006). Core promoter elements recognized by transcription factor IIB. *Biochemical Society Transactions*, 34, 1051–1053.

Dengler, N.G., Dengler, R.E. & Hattersley, P.W. (1985). Differing Ontogenetic Origins of PCR ("Kranz") Sheaths in Leaf Blades of C_4 Grasses (Poaceae). *American Journal of Botany*, 72, 284–302.

Dewitte, W., Riou-Khamlichi, C., Scofield, S., Healy, J.M.S., Jacqmard, A., Kilby, N.J. & Murray, J.A.H. (2003). Altered Cell Cycle Distribution, Hyperplasia, and Inhibited Differentiation in *Arabidopsis* Caused by the D-Type Cyclin CYCD3. *The Plant Cell*, 15, 79–92.

Dewitte, W., Scofield, S., Alcasabas, A.A., Maughan, S.C., Menges, M., Braun, N., Collins, C., Nieuwland, J., Prinsen, E., Sundaresan, V. & Murray, J.A.H. (2007). *Arabidopsis* CYCD3 D-type cyclins link cell proliferation and endocycles and are rate-limiting for cytokinin responses. *Proceedings of the National Academy of Sciences*, 104, 14537–14542.

Dhondt, S., Van Haerenborgh, D., Van Cauwenbergh, C., Merks, R.M.H., Philips, W., Beemster, G.T.S. & Inze, D. (2012). Quantitative analysis of venation patterns of *Arabidopsis* leaves by supervised image analysis. *The Plant Journal*, 69, 553–563.

Di Laurenzio, L., Wysocka-Diller, J., Malamy, J.E., Pysh, L., Helariutta, Y., Freshour, G., Hahn, M.G., Feldmann, K.A. & Benfey, P.N. (1996). The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell*, 86, 423–433.

Diaz, I., Martinez, M., Isabel-LaMoneda, I., Rubio-Somoza, I. & Carbonero, P. (2005). The DOF protein, SAD, interacts with GAMYB in plant nuclei and activates transcription of endosperm-specific genes during barley seed development. *The Plant Journal*, 42, 652–62.

Diaz, I., Vicente-Carbajosa, J., Abraham, Z., Martínez, M., Isabel-La Moneda, I. & Carbonero, P. (2002). The GAMYB protein from barley interacts with the DOF transcription factor BPBF and activates endosperm-specific genes during seed development. *The Plant Journal*, 29, 453–464.

Donner, T.J., Sherr, I. & Scarpella, E. (2009). Regulation of preprocambial cell state acquisition by auxin signaling in *Arabidopsis* leaves. *Development*, 136, 3235–3246.

Döring, F., Streubel, M., Bräutigam, A. & Gowik, U. (2016). Most photorespiratory genes are preferentially expressed in the bundle sheath cells of the C_4 grass Sorghum bicolor . *Journal of Experimental Botany*, 67, 3053–3064.

Downton, W.J.S. (1970). Preferential C_4-dicarboxylic acid synthesis, the postillumination CO_2 burst, carboxyl transfer step, and grana configurations in plants with C_4-photosynthesis. *Canadian Journal of Botany*, 48, 1795–1800.

Edwards, G.E. & Voznesenskaya, E.V. (2011). C_4 Photosynthesis: Kranz Forms and Single-Cell C_4 in Terrestrial Plants. In: *C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms* (eds. Raghavendra, S.A. & Sage, F.R.). Springer Netherlands, Dordrecht, chap. 4, pp. 29–61.

Efroni, I., Han, S.K., Kim, H., Wu, M.F., Steiner, E., Birnbaum, K., Hong, J., Eshed, Y. & Wagner, D. (2013). Regulation of Leaf Maturation by Chromatin-Mediated Modulation of Cytokinin Responses. *Developmental Cell*, 24, 438–445.

Elo, A., Immanen, J., Nieminen, K. & Helariutta, Y. (2009). Stem cell function during plant vascular development. *Seminars in Cell & Developmental Biology*, 20, 1097–1106.

Emms, D.M., Covshoff, S., Hibberd, J.M. & Kelly, S. (2016). Independent and Parallel Evolution of New Genes by Gene Duplication in Two Origins of C_4 Photosynthesis Provides New Insight into the Mechanism of Phloem Loading in C_4 Species. *Molecular Biology and Evolution*, 33, 1796–1806.

Emms, D.M. & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 16, 1–14.

Engler, C., Kandzia, R. & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PloS ONE*, 3, e3647.

Esau, K. (1943). Ontogeny of the vascular bundle in *Zea mays*. *Hilgardia*, 15, 325–368.

Esau, K. (1965). *Plant anatomy*. 2nd edn. Wiley, New York.

Etchells, J.P., Provost, C.M., Mishra, L. & Turner, S.R. (2013). WOX4 and WOX14 act downstream of the PXY receptor kinase to regulate plant vascular proliferation independently of any role in vascular organisation. *Development*, 140, 2224–2234.

Etchells, J.P., Provost, C.M. & Turner, S.R. (2012). Plant vascular cell division is maintained by an interaction between PXY and ethylene signalling. *PLoS Genetics*, 8, e1002997.

Etchells, J.P. & Turner, S.R. (2010). The PXY-CLE41 receptor ligand pair defines a multifunctional pathway that controls the rate and orientation of vascular cell division. *Development*, 137, 767–774.

FAO (2009). How to Feed the World in 2050. Tech. rep., FAO, Rome.

Feng, S., Yue, R., Tao, S., Yang, Y., Zhang, L., Xu, M., Wang, H. & Shen, C. (2015). Genome-wide identification, expression analysis of auxin-responsive GH3 family genes in maize (*Zea mays* L.) under abiotic stresses. *Journal of Integrative Plant Biology*, 57, 783–795.

Fisher, A.P. & Sozzani, R. (2016). Uncovering the networks involved in stem cell maintenance and asymmetric cell division in the *Arabidopsis* root. *Current Opinion in Plant Biology*, 29, 38–43.

Fladung, M. (1994). Genetic Variants of Panicum maximum (Jacq.) in C_4 Photosynthetic Traits. *Journal of Plant Physiology*, 143, 165–172.

Foster, A.S. (1952). Foliar Venation in Angiosperms from an Ontogenetic Standpoint. *American Journal of Botany*, 39, 752–766.

Fouracre, J.P., Ando, S. & Langdale, J.A. (2014). Cracking the Kranz enigma with systems biology. *Journal of Experimental Botany*, 65, 3327–3339.

Fukao, Y. (2012). Protein-Protein Interactions in Plants. *Plant and Cell Physiology*, 53, 617–625.

Furbank, R.T., Jenkins, C.L.D. & Hatch, M.D. (1989). CO_2 concentrating mechanism of C_4 photosynthesis permeability of isolated bundle sheath cells to inorganic carbon. *Plant Physiology*, 91, 1364–1371.

Furuta, K.M., Yadav, S.R., Lehesranta, S., Belevich, I., Miyashima, S., Heo, J.O., Vatén, A., Lindgren, O., De Rybel, B., Van Isterdael, G., Somervuo, P., Lichtenberger, R., Rocha, R., Thitamadee, S., Tähtiharju, S., Auvinen, P., Beeckman, T., Jokitalo, E. & Helariutta, Y. (2014). *Arabidopsis* $NAC_4$5/86 direct sieve element morphogenesis culminating in enucleation. *Science*, 345, 933–937.

Galinha, C., Hofhuis, H., Luijten, M., Willemsen, V., Blilou, I., Heidstra, R. & Scheres, B. (2007). PLETHORA proteins as dose-dependent master regulators of *Arabidopsis* root development. *Nature*, 449, 1053–1057.

Gallavotti, A., Yang, Y., Schmidt, R.J. & Jackson, D. (2008). The Relationship between Auxin Transport and Maize Branching. *Plant Physiology*, 147, 1913–1923.

Gandotra, N., Coughlan, S.J. & Nelson, T. (2013). The *Arabidopsis* leaf provascular cell transcriptome is enriched in genes with roles in vein patterning. *The Plant Journal*, 74, 48–58.

Gardiner, J., Donner, T.J. & Scarpella, E. (2011). Simultaneous activation of SHR and ATHB8 expression defines switch to preprocambial cell state in *Arabidopsis* leaf development. *Developmental Dynamics*, 240, 261–270.

Gardiner, J., Sherr, I. & Scarpella, E. (2010). Expression of DOF genes identifies early stages of vascular development in *Arabidopsis* leaves. *International Journal of Developmental Biology*, 54, 1389–1396.

Gibson, S.I. (2005). Control of plant development and gene expression by sugar signaling. *Current Opinion in Plant Biology*, 8, 93–102.

Gilchrist, E. & Haughn, G. (2010). Reverse genetics techniques: engineering loss and gain of gene function in plants. *Briefings in functional genomics*, 9, 103–110.

Giulini, A., Wang, J. & Jackson, D. (2004). Control of phyllotaxy by the cytokinin-inducible response regulator homologue ABPHYL1. *Nature*, 430, 1031–1034.

Gonzalez, D. (2015). *Plant transcription factors : evolutionary, structural and functional aspects*. Academic Press, Amsterdam.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. & Rokhsar, D.S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic acids research*, 40, D1178–86.

Gowik, U., Bräutigam, A., Weber, K.L., Weber, A.P.M. & Westhoff, P. (2011). Evolution of C_4 Photosynthesis in the Genus Flaveria: How Many and Which Genes Does It Take to Make C_4? *The Plant Cell*, 23, 2087–2105.

Grass Phylogeny Working Group II (2012). New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytologist*, 193, 304–312.

Haberlandt, G. (1882). Vergleichende anatomie des assimilatorischen gewebesystems der pflanzen. *Jb. Wiss. Bot.*, 13, 74–188.

Hagemann, W. & Gleissberg, S. (1996). Organogenetic capacity of leaves: The significance of marginal blastozones in angiosperms. *Plant Systematics and Evolution*, 199, 121–152.

Hall, L.N., Rossini, L., Cribb, L. & Langdale, J.A. (1998). GOLDEN 2: A Novel Transcriptional Regulator of Cellular Differentiation in the Maize Leaf. *The Plant Cell*, 10, 925–936.

Hancock, J.F. (2012). *Plant evolution and the origin of crop species*. 3rd edn. CABI, Cambridge, MA.

Hatch, M.D. (1987). C 4 photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics*, 895, 81–106.

Hatch, M.D. & Slack, C.R. (1966). Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. *The Biochemical journal*, 101, 103–111.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A.P. & Lercher, M.J. (2013). Predicting C_4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape. *Cell*, 153, 1579–1588.

Helariutta, Y., Fukaki, H., Wysocka-Diller, J., Nakajima, K., Jung, J., Sena, G., Hauser, M.T. & Benfey, P.N. (2000). The SHORT-ROOT gene controls radial patterning of the *Arabidopsis* root through radial signaling. *Cell*, 101, 555–567.

Hibberd, J.M. & Covshoff, S. (2010). The regulation of gene expression required for C_4 photosynthesis. *Annual review of plant biology*, 61, 181–207.

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

Hirsch, S., Kim, J., Muñoz, A., Heckmann, A.B., Downie, J.A. & Oldroyd, G.E.D. (2009). GRAS Proteins Form a DNA Binding Complex to Induce Gene Expression during Nodulation Signaling in Medicago truncatula. *The Plant Cell*, 21, 545–557.

Horstman, A., Willemsen, V., Boutilier, K. & Heidstra, R. (2014). AINTEGUMENTA-LIKE proteins: hubs in a plethora of networks. *Trends in Plant Science*, 19, 146–157.

Huang, C.F., Chang, Y.M., Lin, J.J., Yu, C.P., Lin, H.H., Liu, W.Y., Yeh, S., Tu, S.L., Wu, S.H., Ku, M.S.B. & Li, W.H. (2016). Insights into the regulation of C_4 leaf development from comparative transcriptomic analysis. *Current Opinion in Plant Biology*, 30, 1–10.

Hughes, A.L. (1994). The Evolution of Functionally Novel Proteins after Gene Duplication. *Proceedings of the Royal Society of London B: Biological Sciences*, 256, 119–124.

Hughes, T.E., Langdale, J.A. & Kelly, S. (2014). The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. *Genome research*, 24, 1348–1355.

Ilegems, M., Douet, V., Meylan-Bettex, M., Uyttewaal, M., Brand, L., Bowman, J.L. & Stieger, P.A. (2010). Interplay of auxin, KANADI and Class III HD-ZIP transcription factors in vascular tissue formation. *Development*, 137, 975–984.

Inoue, H., Nojima, H. & Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, 96, 23–28.

Ioio, R.D., Nakamura, K., Moubayidin, L., Perilli, S., Taniguchi, M., Morita, M.T., Aoyama, T., Costantino, P. & Sabatini, S. (2008). A genetic framework for the control of cell division and differentiation in the root meristem. *Science*, 322, 1380–1384.

Isabel-LaMoneda, I., Diaz, I., Martinez, M., Mena, M. & Carbonero, P. (2003). SAD: A new DOF protein from barley that activates transcription of a cathepsin B-like thiol protease gene in the aleurone of germinating seeds. *Plant Journal*, 33, 329–340.

Ishiwata, A., Ozawa, M., Nagasaki, H., Kato, M., Noda, Y., Yamaguchi, T., Nosaka, M., Shimizu-Sato, S., Nagasaki, A., Maekawa, M., Hirano, H.Y. &

Sato, Y. (2013). Two WUSCHEL-related homeobox Genes, *narrow leaf2* and *narrow leaf3*, Control Leaf Width in Rice. *Plant and Cell Physiology*, 54, 779–792.

Jackson, D. (1992). *In situ* hybridisation in plants. In: *Molecular plant pathology : a practical approach Vol I* (eds. Gurr, S.J., McPherson, M.J. & Bowles, D.J.). Oxford, pp. 163–174. 1st edn.

Jacobs, W.P. (1952). The Role of Auxin in Differentiation of Xylem Around a Wound. *American Journal of Botany*, 39, 301–309.

Jankovsky, J.P., Smith, L.G. & Nelson, T. (2001). Specification of bundle sheath cell fates during maize leaf development: roles of lineage and positional information evaluated through analysis of the tangled1 mutant. *Development*, 128, 2747–2753.

Javelle, M., Marco, C.F. & Timmermans, M. (2011). *in situ* Hybridization for the Precise Localization of Transcripts in Plants. *Journal of Visualized Experiments*, p. e3328.

Jenkins, M.T. (1926). A Second Gene Producing Golden Plant Color in Maize. *The American Naturalist*, 60, 484–488.

Ji, J., Strable, J., Shimizu, R., Koenig, D., Sinha, N. & Scanlon, M.J. (2010). WOX4 Promotes Procambial Development. *Plant Physiology*, 152, 1346–1356.

Jiao, Y., Burke, J., Chopra, R., Burow, G., Chen, J., Wang, B., Hayes, C., Emendack, Y., Ware, D. & Xin, Z. (2016). A Sorghum Mutant Resource as an Efficient Platform for Gene Discovery in Grasses. *The Plant Cell*, 28, 1551–1562.

Jin, J., Zhang, H., Kong, L., Gao, G. & Luo, J. (2014). PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Research*, 42, D1182–D1187.

Johnston, R., Leiboff, S. & Scanlon, M.J. (2015). Ontogeny of the sheathing leaf base in maize (*Zea mays*). *New Phytologist*, 205, 306–315.

Kajala, K., Covshoff, S., Karki, S., Woodfield, H., Tolley, B.J., Dionora, M.J.A., Mogul, R.T., Mabilangan, A.E., Danila, F.R. & Hibberd, J.M. (2011). Strategies for engineering a two-celled C_4 photosynthetic pathway into rice. *Journal of experimental botany*, 62, 3001–3010.

Kang, J. & Dengler, N. (2002). Cell cycling frequency and expression of the homeobox gene ATHB-8 during leaf vein development in *Arabidopsis*. *Planta*, 216, 212–219.

Kang, J., Mizukami, Y., Wang, H., Fowke, L. & Dengler, N.G. (2007). Modification of cell proliferation patterns alters leaf vein architecture in *Arabidopsis thaliana*. *Planta*, 226, 1207–18.

Kelley, D.R., Arreola, A., Gallagher, T.L. & Gasser, C.S. (2012). ETTIN (ARF3) physically interacts with KANADI proteins to form a functional complex essential for integument development and polarity determination in *Arabidopsis*. *Development*, 139, 1105–1109.

Kennedy, R.A. & Laetsch, W.M. (1974). Plant species intermediate for C_3, C_4 photosynthesis. *Science*, 184, 1087–1089.

Kersey, P.J., Allen, J.E., Armean, I., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen, M., Davis, P., Falin, L.J., Grabmueller, C., Humphrey, J., Kerhornou, A., Khobova, J., Aranganathan, N.K., Langridge, N., Lowy, E., McDowall, M.D., Maheswari, U., Nuhn, M., Ong, C.K., Overduin, B., Paulini, M., Pedro, H., Perry, E., Spudich, G., Tapanari, E., Walts, B., Williams, G., TelloâĂŞRuiz, M., Stein, J., Wei, S., Ware, D., Bolser, D.M., Howe, K.L., Kulesha, E., Lawson, D., Maslen, G. & Staines, D.M. (2016). Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Research*, 44, D574–D580.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. & Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14, 1–13.

Kim, J., Jung, J.H., Reyes, J.L., Kim, Y.S., Kim, S.Y., Chung, K.S., Kim, J.A., Lee, M., Lee, Y., Narry Kim, V., Chua, N.H. & Park, C.M. (2005). microRNA-directed cleavage of ATHB15 mRNA regulates vascular development in *Arabidopsis* inflorescence stems. *The Plant Journal*, 42, 84–94.

Kim, S., Soltis, P.S., Wall, K. & Soltis, D.E. (2006). Phylogeny and Domain Evolution in the APETALA2-like Gene Family. *Molecular Biology and Evolution*, 23, 107–120.

Kinsman, E.A. & Pyke, K.A. (1998). Bundle sheath cells and cell-specific plastid development in *Arabidopsis* leaves. *Development*, 125, 1815–1822.

Kitomi, Y., Ito, H., Hobo, T., Aya, K., Kitano, H. & Inukai, Y. (2011). The auxin responsive AP2/ERF transcription factor CROWN ROOTLESS5 is involved in crown root initiation in rice through the induction of OsRR1, a type-A response regulator of cytokinin signaling. *The Plant Journal*, 67, 472–484.

Klucher, K.M., Chow, H., Reiser, L. & Fischer, R.L. (1996). The AINTEGUMENTA gene of *Arabidopsis* required for ovule and female gametophyte development is related to the floral homeotic gene APETALA2. *The Plant cell*, 8, 137–53.

Knappe, S., Löttgert, T., Schneider, A., Voll, L., Flügge, U.I. & Fischer, K. (2003). Characterization of two functional phosphoenolpyruvate/phosphate translocator (PPT) genes in *Arabidopsis*-AtPPT1 may be involved in the provision of signals for correct mesophyll development. *The Plant Journal*, 36, 411–420.

Koizumi, K., Sugiyama, M. & Fukuda, H. (2000). A series of novel mutants of *Arabidopsis thaliana* that are defective in the formation of continuous vascular network: calling the auxin signal flow canalization hypothesis into question. *Development*, 127, 3197– 3204.

Krebs, J., Mueller-Roeber, B. & Ruzicic, S. (2010). A novel bipartite nuclear localization signal with an atypically long linker in DOF transcription factors. *Journal of plant physiology*, 167, 583–586.

Krizek, B.A. (1999). Ectopic expression of AINTEGUMENTA in *Arabidopsis* plants results in increased growth of floral organs. *Developmental genetics*, 25, 224–36.

Krizek, B.A. (2003). AINTEGUMENTA utilizes a mode of DNA recognition distinct from that used by proteins containing a single AP2 domain. *Nucleic Acids Research*, 31, 1859–1868.

Krizek, B.A. (2011). Aintegumenta and Aintegumenta-like 6 regulate auxin-mediated flower development in *Arabidopsis*. *BMC Research Notes*, 4, 176.

Krizek, B.A. & Sulli, C. (2006). Mapping sequences required for nuclear localization and the transcriptional activation function of the *Arabidopsis* protein AINTEGUMENTA. *Planta*, 224, 612–621.

Külahoglu, C., Denton, A.K., Sommer, M., Maß, J., Schliesky, S., Wrobel, T.J., Berckmans, B., Gongora-Castillo, E., Buell, C.R., Simon, R., De

Veylder, L., Bräutigam, A. & Weber, A.P.M. (2014). Comparative Transcriptome Atlases Reveal Altered Gene Expression Modules between Two Cleomaceae C_3 and C_4 Plant Species. *The Plant Cell*, 26, 3243–3260.

Kurakawa, T., Ueda, N., Maekawa, M., Kobayashi, K., Kojima, M., Nagato, Y., Sakakibara, H. & Kyozuka, J. (2007). Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature*, 445, 652–655.

Laetsch, W. (1971). Chloroplast structural relationships in leaves of C_4 plants. In: *Photosynthesis and photorespiration : proceedings of a conference held at Australian National University, Canberra, Australia, 23 November-5 December, 1970* (eds. Hatch, M.D., Osmond, C.B. & Slatyer, R.O.). Wiley-Interscience, New York, pp. 323–349.

Laetsch, W.M. (1974). The C_4 Syndrome: A Structural Analysis. *Annual Review of Plant Physiology*, 25, 27–52.

Langdale, J.A. (1994). *In situ* Hybridization. In: *The Maize Handbook* (eds. Freeling, M. & Walbot, V.). Springer, New York, NY, pp. 165–180. 1st edn.

Langdale, J.A. & Kidner, C.A. (1994). bundle sheath defective, a mutation that disrupts cellular differentiation in maize leaves. *Development*, 120, 673– 681.

Langdale, J.A., Lane, B., Freeling, M. & Nelson, T. (1989). Cell lineage analysis of maize bundle sheath and mesophyll cells. *Developmental Biology*, 133, 128–139.

Langdale, J.A. & Nelson, T. (1991). Spatial regulation of photosynthetic development in C_4 plants. *Trends in Genetics*, 7, 191–196.

Langdale, J.A., Rothermel, B.A. & Nelson, T. (1988a). Cellular pattern of photosynthetic gene expression in developing maize leaves. *Genes & Development*, 2, 106–115.

Langdale, J.A., Zelitch, I., Miller, E. & Nelson, T. (1988b). Cell position and light influence C_4 versus C_3 patterns of photosynthetic gene expression in maize. *The EMBO journal*, 7, 3643–3651.

Lauterbach, M., Billakurthi, K., Kadereit, G., Ludwig, M., Westhoff, P. & Gowik, U. (2016). C_3 cotyledons are followed by C_4 leaves: intra-individual transcriptome analysis of *Salsola soda* (Chenopodiaceae). *Journal of Experimental Botany*. doi: 10.1093/jxb/erw343

Le Hir, R. & Bellini, C. (2013). The plant-specific dof transcription factors family: new players involved in vascular system development and functioning in *Arabidopsis*. *Frontiers in plant science*, 4, 164.

Leegood, R.C. (2008). Roles of the bundle sheath cells in leaves of C_3 plants. *Journal of Experimental Botany*, 59, 1663–1673.

Leegood, R.C., Lea, P.J., Adcock, M.D. & Häusler, R.E. (1995). The regulation and control of photorespiration. *Journal of Experimental Botany*, 46, 1397–1414.

Letunic, I & Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees *Nucleic Acids Research*, 44, W242–245.

Levene, H. (1960). Robust tests for equality of variances1. *Contributions to probability and statistics: Essays in honor of Harold Hotelling*, 2, 278–292.

Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S.L., Kebrom, T.H., Provart, N., Patel, R., Myers, C.R., Reidel, E.J., Turgeon, R., Liu, P., Sun, Q., Nelson, T. & Brutnell, T.P. (2010). The developmental dynamics of the maize leaf transcriptome. *Nature Genetics*, 42, 1060–1067.

Life Technologies (2012). TRIzol Reagent. Tech. Rep. 15596026.PPS, Carlsbad, USA.

Liu, L., White, M.J. & MacRae, T.H. (1999). Transcription factors and their genes in higher plants. *European Journal of Biochemistry*, 262, 247–257.

Ljung, K. (2013). Auxin metabolism and homeostasis during plant development. *Development*, 140, 943– 950.

Lundquist, P.K., Rosar, C., Bräutigam, A. & Weber, A.P.M. (2014). Plastid Signals and the Bundle Sheath: Mesophyll Development in Reticulate Mutants. *Molecular Plant*, 7, 14–29.

Lynch, M. & Conery, J.S. (2000). The Evolutionary Fate and Consequences of Duplicate Genes. *Science*, 290, 1151– 1155.

Ma, L., Sang, X., Zhang, T., Yu, Z., Li, Y., Zhao, F., Wang, Z., Wang, Y., Yu, P., Wang, N., Zhang, C., Ling, Y., Yang, Z. & He, G. (2016). ABNORMAL VASCULAR BUNDLES regulates cell proliferation and procambium cell establishment during aerial organ development in rice. *New Phytologist*, doi:10.1111/nph.14142.

Mähönen, A.P., Bishopp, A., Higuchi, M., Nieminen, K.M., Kinoshita, K., Törmäkangas, K., Ikeda, Y., Oka, A., Kakimoto, T. & Helariutta, Y. (2006). Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development. *Science*, 311, 94–98.

Mähönen, A.P., Bonke, M., Kauppinen, L., Riikonen, M., Benfey, P.N. & Helariutta, Y. (2000). A novel two-component hybrid molecule regulates vascular morphogenesis of the *Arabidopsis* root. *Genes & development*, 14, 2938–2943.

Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M.J., Weber, A.P.M., Westhoff, P. & Gowik, U. (2014). The role of photorespiration during the evolution of C_4 photosynthesis in the genus Flaveria. *eLife*, 3, e02478.

Marshall, D.M., Muhaidat, R., Brown, N.J., Liu, Z., Stanley, S., Griffiths, H., Sage, R.F. & Hibberd, J.M. (2007). *Cleome*, a genus closely related to *Arabidopsis*, contains species spanning a developmental progression from C_3 to C_4 photosynthesis. *The Plant Journal*, 51, 886–896.

Mattsson, J., Ckurshumova, W. & Berleth, T. (2003). Auxin signaling in *Arabidopsis* leaf vascular development. *Plant Physiology*, 131, 1327–1339.

McAbee, J.M., Hill, T.A., Skinner, D.J., Izhaki, A., Hauser, B.A., Meister, R.J., Venugopala Reddy, G., Meyerowitz, E.M., Bowman, J.L. & Gasser, C.S. (2006). ABERRANT TESTA SHAPE encodes a KANADI family member, linking polarity determination to separation and growth of *Arabidopsis* ovule integuments. *The Plant Journal*, 46, 522–531.

McCarty, D.R., Settles, A.M., Suzuki, M., Tan, B.C., Latshaw, S., Porch, T., Robin, K., Baier, J., Avigne, W., Lai, J., Messing, J., Koch, K.E. & Hannah, L.C. (2005). Steady-state transposon mutagenesis in inbred maize. *The Plant Journal*, 44, 52–61.

Mi, H., Lazareva-Ulitsky, B., Loo, R.,Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A.,Doremieux, O., Campbell, M.J., Kitano, H. & Thomas, P.D. (2005) The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Research*, 33, D284–D288.

Mi, H., Muruganujan. A., Casagrande, J.T & Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system *Nature Protocols*, 8, 1551–1566.

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T. & Thomas, P.D. (2016). PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic acids research*, 44, D336–342.

Miki, D. & Shimamoto, K. (2004). Simple RNAi vectors for stable and transient suppression of gene function in rice. *Plant and Cell Physiology*, 45, 490–495.

Mitsuda, N. & Ohme-Takagi, M. (2009). Functional analysis of transcription factors in *Arabidopsis*. *Plant and Cell Physiology*, 50, 1232–1248.

Mizukami, Y. & Fischer, R.L. (2000). Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. *Proceedings of the National Academy of Sciences*, 97, 942–947.

Mok, D.W.S. & Mok, M.C. (2001). Cytokinin metabolism and action. *Annual Review of Plant Physiology and Plant Molecular Biology*, 52, 89–118.

Monson, R. (2003). Gene Duplication, Neofunctionalization, and the Evolution of C_4 Photosynthesis. *International Journal of Plant Sciences*, 164, S43–S54.

Moore, I., Samalova, M. & Kurup, S. (2006). Transactivated and chemically inducible gene expression in plants. *The Plant Journal*, 45, 651–683.

Moore, R.C. & Purugganan, M.D. (2005). The evolutionary dynamics of plant duplicate genes. *Current Opinion in Plant Biology*, 8, 122–128.

Moreno-Risueno, M.A., Sozzani, R., YardÄśmcÄś, G.G., Petricka, J.J., Vernoux, T., Blilou, I., Alonso, J., Winter, C.M., Ohler, U., Scheres, B. & Benfey, P.N. (2015). Transcriptional control of tissue formation throughout root development. *Science*, 350, 426– 430.

Mouchel, C.F., Osmont, K.S. & Hardtke, C.S. (2006). BRX mediates feedback between brassinosteroid levels and auxin signalling in root growth. *Nature*, 443, 458–461.

Mudunkothge, J.S. & Krizek, B.A. (2012). Three *Arabidopsis* AIL/PLT genes act in combination to regulate shoot apical meristem function. *The Plant Journal*, 71, 108–121.

Murray, M.G. & Thompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, 8, 4321–4326.

Nelson, T. (2011). The grass leaf developmental gradient as a platform for a systems understanding of the anatomical specialization of C_4 leaves. *Journal of Experimental Botany*, 62, 3039–3048.

Nelson, T. & Langdale, J.A. (1989). Patterns of leaf development in C_4 plants. *The Plant Cell*, 1, 3.

Nishimura, A., Ito, M., Kamiya, N., Sato, Y. & Matsuoka, M. (2002). OsPNH1 regulates leaf development and maintenance of the shoot apical meristem in rice. *The Plant Journal*, 30, 189–201.

Nole-Wilson, S. & Krizek, B.A. (2000). DNA binding properties of the *Arabidopsis* floral development protein AINTEGUMENTA. *Nucleic Acids Research*, 28, 4076–4082.

Nole-Wilson, S., Tranby, T.L. & Krizek, B.A. (2005). AINTEGUMENTA-like (AIL) genes are expressed in young tissues and may specify meristematic or division-competent states. *Plant molecular biology*, 57, 613–28.

Nomura, M., Katayama, K., Nishimura, A., Ishida, Y., Ohta, S., Komari, T., Miyao-Tokutomi, M., Tajima, S. & Matsuoka, M. (2000). The promoter of rbcS in a C_3 plant (rice) directs organ-specific, light-dependent expression in a C_4 plant (maize), but does not confer bundle sheath cell-specific expression. *Plant Molecular Biology*, 44, 99–106.

Ogasawara, H., Kaimi, R., Colasanti, J. & Kozaki, A. (2011). Activity of transcription factor JACKDAW is essential for SHR/SCR-dependent activation of SCARECROW and MAGPIE and is modulated by reciprocal interactions with MAGPIE, SCARECROW and SHORT ROOT. *Plant Molecular Biology*, 77, 489–499.

Ohashi-Ito, K. & Bergmann, D.C. (2006). *Arabidopsis* FAMA Controls the Final Proliferation/Differentiation Switch during Stomatal Development. *The Plant Cell*, 18, 2493–2505.

Ohmori, Y., Tanaka, W., Kojima, M., Sakakibara, H. & Hirano, H.Y. (2013). WUSCHEL-RELATED HOMEOBOX4 Is Involved in Meristem Maintenance and Is Negatively Regulated by the CLE Gene FCP1 in Rice. *The Plant Cell*, 25, 229–241.

Ohmori, Y., Yasui, Y. & Hirano, H.Y. (2014). Overexpression analysis suggests that FON2-LIKE CLE PROTEIN1 is involved in rice leaf development. *Genes & Genetic Systems*, 89, 87–91.

Ohno, S. (1970). *Evolution by gene duplication*. 1st edn. Allen & Unwin, London.

Ort, D.R., Merchant, S.S., Alric, J., Barkan, A., Blankenship, R.E., Bock, R., Croce, R., Hanson, M.R., Hibberd, J.M., Long, S.P., Moore, T.A., Moroney, J., Niyogi, K.K., Parry, M.A.J., Peralta-Yahya, P.P., Prince, R.C., Redding, K.E., Spalding, M.H., van Wijk, K.J., Vermaas, W.F.J., von Caemmerer, S., Weber, A.P.M., Yeates, T.O., Yuan, J.S. & Zhu, X.G. (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proceedings of the National Academy of Sciences*, 112, 8529–8536.

Ouwerkerk, P.B., de Kam, R.J., Hoge, J.H. & Meijer, A.H. (2001). Glucocorticoid-inducible gene expression in rice. *Planta*, 213, 370–378.

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J. & Buell, C.R. (2007). The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Research*, 35, D883–D887.

Pan, B.Z., Chen, M.S., Ni, J. & Xu, Z.F. (2014). Transcriptome of the inflorescence meristems of the biofuel plant *Jatropha curcas* treated with cytokinin. *BMC Genomics*, 15, 1–21.

Pan, Y., Chen, X.Q., Xie, H., Deng, L., Li, X.L., Zhang, X.D., Han, L.X., Yang, F.P., Xue, J. & Zhang, L.Q. (2015). A maize bundle sheath defective mutation mapped on chromosome 1 between SSR markers umc1395 and umc1603. *Journal of Integrative Agriculture*, 14, 1949–1957.

Pelechano, V. & Steinmetz, L.M. (2013). Gene regulation by antisense transcription. *Nature Reviews Genetics*, 14, 880–893.

Petricka, J.J., Clay, N.K. & Nelson, T.M. (2008). Vein patterning screens and the *defectively organized tributaries* mutants in *Arabidopsis thaliana*. *The Plant Journal*, 56, 251–263.

Pick, T.R., Brautigam, A., Schluter, U., Denton, A.K., Colmsee, C., Scholz, U., Fahnenstich, H., Pieruschka, R., Rascher, U., Sonnewald, U. & Weber, A.P.M. (2011). Systems Analysis of a Maize Leaf Developmental Gradient

Redefines the Current C_4 Model and Provides Candidates for Regulation. *The Plant Cell*, 23, 4208–4220.

Poethig, R.S. & Szymkowiak, E.J. (1995). Clonal analysis of leaf development in maize. *Maydica*, 40, 67–76.

Prigge, M.J., Otsuga, D., Alonso, J.M., Ecker, J.R., Drews, G.N. & Clark, S.E. (2005). Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in *Arabidopsis* development. *The Plant cell*, 17, 61–76.

R Core Team (2013). R: A language and environment for statistical computing.

Randall, R.S., Miyashima, S., Blomster, T., Zhang, J., Elo, A., Karlberg, A., Immanen, J., Nieminen, K., Lee, J.Y., Kakimoto, T., Blajecka, K., Melnyk, C.W., Alcasabas, A., Forzani, C., Matsumoto-Kitano, M., Mähönen, A.P., Bhalerao, R., Dewitte, W., Helariutta, Y. & Murray, J.A.H. (2015a). AINTEGUMENTA and the D-type cyclin CYCD3;1 regulate root secondary growth and respond to cytokinins. *Biology Open*. doi: 10.1242/bio.013128

Randall, R.S., Sornay, E., Dewitte, W. & Murray, J.A.H. (2015b). AINTEGUMENTA and the D-type cyclin CYCD3;1 independently contribute to petal size control in *Arabidopsis*: evidence for organ size compensation being an emergent rather than a determined property. *Journal of Experimental Botany*, 66, 3991–4000.

Rao, X., Lu, N., Li, G., Nakashima, J., Tang, Y. & Dixon, R.A. (2016). Comparative cell-specific transcriptomics reveals differentiation of C_4 photosynthesis pathways in switchgrass and other C_4 lineages. *Journal of Experimental Botany*, 67, 1649–1662.

Rashotte, A.M., Mason, M.G., Hutchison, C.E., Ferreira, F.J., Schaller, G.E. & Kieber, J.J. (2006). A subset of *Arabidopsis* AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway. *Proceedings of the National Academy of Sciences*, 103, 11081–11085.

Rawsthorne, S., Hylton, C.M., Smith, A.M. & Woolhouse, H.W. (1988). Photorespiratory metabolism and immunogold localization of photorespiratory enzymes in leaves of C_3 and C_3-C_4 intermediate species of Moricandia. *Planta*, 173, 298–308.

Rizal, G., Thakur, V., Dionora, J., Karki, S., Wanchana, S., Acebron, K., Larazo, N., Garcia, R., Mabilangan, A., Montecillo, F., Danila, F., Mogul, R., Pablico, P., Leung, H., Langdale, J.A., Sheehy, J., Kelly, S. & Quick, W.P. (2015). Two forward genetic screens for vein density mutants in sorghum converge on a cytochrome P450 gene in the brassinosteroid pathway. *The Plant Journal*, 84, 257–266.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., and Lander, E.S., Getz, G. & Mesirov, J.P.(2011) *Nature biotechnology*, 29, 24–26

Rodriguez-Villalon, A., Gujas, B., Kang, Y.H., Breda, A.S., Cattaneo, P., Depuydt, S. & Hardtke, C.S. (2014). Molecular genetic framework for protophloem formation. *Proceedings of the National Academy of Sciences*, 111, 11551–11556.

Rodriguez-Villalon, A., Gujas, B., van Wijk, R., Munnik, T. & Hardtke, C.S. (2015). Primary root protophloem differentiation requires balanced phosphatidylinositol-4,5-biphosphate levels and systemically affects root branching. *Development*, 142, 1437–1446.

Rosar, C., Kanonenberg, K., Nanda, A.M., Mielewczik, M., Bräutigam, A., Novák, O., Strnad, M., Walter, A. & Weber, A.P.M. (2012). The Leaf Reticulate Mutant *dov1* Is Impaired in the First Step of Purine Metabolism. *Molecular Plant*, 5, 1227–1241.

Rossini, L., Cribb, L., Martin, D.J. & Langdale, J.A. (2001). The Maize Golden2 Gene Defines a Novel Class of Transcriptional Regulators in Plants. *The Plant Cell*, 13, 1231–1244.

Royo, J., Gómez, E., Barrero, C., Muñiz, L.M., Sanz, Y. & Hueros, G. (2009). Transcriptional activation of the maize endosperm transfer cell-specific gene BETL1 by ZmMRP-1 is enhanced by two C_2H2 zinc finger-containing proteins. *Planta*, 230, 807–818.

Sachs, T. (1981). The Control of the Patterned Differentiation of Vascular Tissues. Academic Press, vol. 9, pp. 151–262.

Sage, R.F., Christin, P.A. & Edwards, E.J. (2011a). The C_4 plant lineages of planet Earth. *Journal of Experimental Botany*, 62, 3155–3169.

Sage, R.F., Khoshravesh, R. & Sage, T.L. (2014). From proto-Kranz to C_4 Kranz: building the bridge to C_4 photosynthesis. *Journal of Experimental Botany*, 65, 3341–3356.

Sage, R.F., Sage, T.L. & Kocacinar, F. (2012). Photorespiration and the evolution of C_4 photosynthesis. *Annual review of plant biology*, 63, 19–47.

Sage, R.F. & Zhu, X.G. (2011). Exploiting the engine of C_4 photosynthesis. *Journal of Experimental Botany*, 62, 2989–3000.

Sage, T.L., Busch, F.A., Johnson, D.C., Friesen, P.C., Stinson, C.R., Stata, M., Sultmanis, S., Rahman, B.A., Rawsthorne, S. & Sage, R.F. (2013). Initial Events during the Evolution of C_4 Photosynthesis in C_3 Species of Flaveria. *Plant Physiology*, 163, 1266–1276.

Sage, T.L., Sage, R.F., Vogan, P.J., Rahman, B., Johnson, D.C., Oakley, J.C. & Heckel, M.A. (2011b). The occurrence of C_2 photosynthesis in Euphorbia subgenus Chamaesyce (Euphorbiaceae). *Journal of experimental botany*, 62, 3183–3195.

Sakaguchi, J. & Fukuda, H. (2008). Cell differentiation in the longitudinal veins and formation of commissural veins in rice (*Oryza sativa*) and maize (*Zea mays*). *Journal of plant research*, 121, 593–602.

Sambrook, J. (2001). *Molecular cloning : a laboratory manual*. 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Scacchi, E., Osmont, K.S., Beuchat, J., Salinas, P., Navarrete-Gómez, M., Trigueros, M., Ferrándiz, C. & Hardtke, C.S. (2009). Dynamic, auxin-responsive plasma membrane-to-nucleus movement of *Arabidopsis* BRX. *Development*, 136, 2059– 2067.

Scacchi, E., Salinas, P., Gujas, B., Santuari, L., Krogan, N., Ragni, L., Berleth, T. & Hardtke, C.S. (2010). Spatio-temporal sequence of cross-regulatory events in root meristem growth. *Proceedings of the National Academy of Sciences*, 107, 22734–22739.

Scarpella, E. (2003). The RADICLELESS1 gene is required for vascular pattern formation in rice. *Development*, 130, 645–658.

Scarpella, E., Boot, K.J.M., Rueb, S. & Meijer, A.H. (2002). The Procambium Specification Gene OsHOX1 Promotes Polar Auxin Transport Capacity and Reduces Its Sensitivity toward Inhibition . *Plant Physiology*, 130, 1349–1360.

Scarpella, E., Francis, P. & Berleth, T. (2004). Stage-specific markers define early steps of procambium development in *Arabidopsis* leaves and correlate termination of vein formation with mesophyll differentiation. *Development*, 131, 3445–3455.

Scarpella, E., Marcos, D., Friml, J. & Berleth, T. (2006). Control of leaf vascular patterning by polar auxin transport. *Genes & Development*, 20, 1015–1027.

Schlereth, A., Möller, B., Liu, W., Kientz, M., Flipse, J., Rademacher, E.H., Schmid, M., Jürgens, G. & Weijers, D. (2010). MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. *Nature*, 464, 913–916.

Schmitz-Linneweber, C. & Small, I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends in Plant Science*, 13, 663–670.

Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9, 671–675.

Schulze, S., Westhoff, P. & Gowik, U. (2016). Glycine decarboxylase in C_3, C_4 and C_3-C_4 intermediate species. *Current Opinion in Plant Biology*, 31, 29–35.

Shahmuradov, I.A. & Solovyev, V.V. (2015). Nsite, NsiteH and NsiteM computer tools for studying transcription regulatory elements. *Bioinformatics*, 31, 3544–3545.

Shapiro, S.S. & Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591–611.

Sharkey, T.D. (1988). Estimating the rate of photorespiration in leaves. *Physiologia Plantarum*, 73, 147–152.

Sharman, B.C. (1942). Developmental anatomy of the shoot of *Zea mays* L. *Annals of Botany*, 6, 245–282.

Sharp, R.E. & LeNoble, M.E. (2002). ABA, ethylene and the control of shoot and root growth under water stress. *Journal of Experimental Botany*, 53, 33–37.

Sharpe, R.M., Mahajan, A., Takacs, E.M., Stern, D.B. & Cahoon, A.B. (2011). Developmental and cell type characterization of bundle sheath and mesophyll chloroplast transcript abundance in maize. *Current genetics*, 57, 89–102.

Sieburth, L.E., Muday, G.K., King, E.J., Benton, G., Kim, S., Metcalf, K.E., Meyers, L., Seamen, E. & Van Norman, J.M. (2006). SCARFACE Encodes an ARF-GAP That Is Required for Normal Auxin Efflux and Vein Patterning in *Arabidopsis*. *The Plant Cell*, 18, 1396–1411.

Skirycz, A., Reichelt, M., Burow, M., Birkemeyer, C., Rolcik, J., Kopka, J., Zanor, M.I., Gershenzon, J., Strnad, M., Szopa, J., Mueller-Roeber, B. & Witt, I. (2006). DOF transcription factor AtDof1.1 (OBP2) is part of a regulatory network controlling glucosinolate biosynthesis in *Arabidopsis*. *The Plant Journal*, 47, 10–24.

Slewinski, T.L., Anderson, A.A., Zhang, C. & Turgeon, R. (2012). SCARECROW Plays a Role in Establishing Kranz Anatomy in Maize Leaves. *Plant and Cell Physiology*, 53, 2030–2037.

Smillie, I.R.A., Pyke, K.A. & Murchie, E.H. (2012). Variation in vein density and mesophyll cell architecture in a rice deletion mutant population. *Journal of Experimental Botany*, 63, 4563–4570.

Smith, G.H. (1934). Anatomy of the Embryonic Leaf. *American Journal of Botany*, 21, 194–209.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.

Strabala, T.J., O'Donnell, P.J., Smit, A.M., Ampomah-Dwamena, C., Martin, E.J., Netzler, N., Nieuwenhuizen, N.J., Quinn, B.D., Foote, H.C.C. & Hudson, K.R. (2006). Gain-of-Function Phenotypes of Many CLAVATA3/ESR Genes, Including Four New Family Members, Correlate with Tandem Variations in the Conserved CLAVATA3/ESR Domain. *Plant Physiology*, 140, 1331–1344.

TAIR (2016) The Arabidopsis Information Resource (TAIR) pattern matching www.arabidopsis.org/cgi-bin/patmatch/nph-patmatch.pl

Taiz, L., Zeiger, E., Møller, I.M. & Murphy, A. (2015). *Plant physiology and development*. 6th edn. Sinauer Associates, Inc., Stamford, CT.

Tausta, S.L., Li, P., Si, Y., Gandotra, N., Liu, P., Sun, Q., Brutnell, T.P. & Nelson, T. (2014). Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C_4-related processes. *Journal of Experimental Botany*, 65, 3543–3555.

Tokunaga, H., Kojima, M., Kuroha, T., Ishida, T., Sugimoto, K., Kiba, T. & Sakakibara, H. (2012). *Arabidopsis lonely guy* (LOG) multiple mutants reveal a central role of the LOG-dependent pathway in cytokinin activation. *The Plant Journal*, 69, 355–365.

Truernit, E., Bauby, H., Belcram, K., Barthelemy, J. & Palauqui, J.C. (2012). OCTOPUS, a polarly localised membrane-associated protein, regulates phloem differentiation entry in *Arabidopsis thaliana*. *Development*, 139, 1306–1315.

Tukey, J.W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, pp. 99–114.

Ulmasov, T., Murfett, J., Hagen, G. & Guilfoyle, T.J. (1997). Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *The Plant Cell*, 9, 1963–1971.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. & Rozen, S.G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, 40, 1–12.

Urban, M.A., Nelson, D.M., Jiménez-Moreno, G., Châteauneuf, J.J., Pearson, A. & Hu, F.S. (2010). Isotopic evidence of C_4 grasses in southwestern Europe during the Early Oligocene-Middle Miocene. *Geology*, 38, 1091–1094.

Van Norman, J.M., Breakfield, N.W. & Benfey, P.N. (2011). Intercellular Communication during Plant Development. *The Plant Cell*, 23, 855–864.

Vicente-Carbajosa, J., Moose, S.P., Parsons, R.L. & Schmidt, R.J. (1997). A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proceedings of the National Academy of Sciences*, 94, 7685–7690.

Vicentini, A., Barber, J.C., Aliscioni, S.S., Giussani, L.M. & Kellogg, E.A. (2008). The age of the grasses and clusters of origins of C_4 photosynthesis. *Global Change Biology*, 14, 2963–2977.

Vogan, P.J., Frohlich, M.W. & Sage, R.F. (2007). The functional significance of C_3-C_4 intermediate traits in *Heliotropium* L. (Boraginaceae): gas exchange perspectives. *Plant, cell & environment*, 30, 1337–1345.

Voll, L., Hausler, R.E., Hecker, R., Weber, A., Weissenbock, G., Fiene, G., Waffenschmidt, S. & Flugge, U.I. (2003). The phenotype of the *Arabidopsis cue1* mutant is not simply caused by a general restriction of the shikimate pathway. *The Plant Journal*, 36, 301–317.

Vollbrecht, E., Duvick, J., Schares, J.P., Ahern, K.R., Deewatthanawong, P., Xu, L., Conrad, L.J., Kikuchi, K., Kubinec, T.A., Hall, B.D., Weeks, R., Unger-Wallace, E., Muszynski, M., Brendel, V.P. & Brutnell, T.P. (2010). Genome-Wide Distribution of Transposed Dissociation Elements in Maize. *The Plant Cell*, 22, 1667–1685.

Voznesenskaya, E.V., Franceschi, V.R., Kiirats, O., Freitag, H. & Edwards, G.E. (2001). Kranz anatomy is not essential for terrestrial C_4 plant photosynthesis. *Nature*, 414, 543–546.

Voznesenskaya, E.V., Koteyeva, N.K., Chuong, S.D.X., Ivanova, A.N., Barroca, J., Craven, L.A. & Edwards, G.E. (2007). Physiological, anatomical and biochemical characterisation of photosynthetic types in genus *Cleome* (Cleomaceae). *Functional Plant Biology*, 34, 247–267.

Wang, L., Czedik-Eysenberg, A., Mertz, R.A., Si, Y., Tohge, T., Nunes-Nesi, A., Arrivault, S., Dedow, L.K., Bryant, D.W., Zhou, W., Xu, J., Weissmann, S., Studer, A., Li, P., Zhang, C., LaRue, T., Shao, Y., Ding, Z., Sun, Q., Patel, R.V., Turgeon, R., Zhu, X., Provart, N.J., Mockler, T.C., Fernie, A.R., Stitt, M., Liu, P. & Brutnell, T.P. (2014a). Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice. *Nat Biotech*, 32, 1158–1165.

Wang, M.B., Upadhyaya, N.M., Brettell, R.I.S. & Waterhouse, P.M. (1997). Intron-mediated improvement of a selectable marker gene for plant transformation using Agrobacterium tumefaciens. *Journal of Genetics and Breeding*, 51, 325–334.

Wang, P., Fouracre, J., Kelly, S., Karki, S., Gowik, U., Aubry, S., Shaw, M.K., Westhoff, P., Slamet-Loedin, I.H., Quick, W.P., Hibberd, J.M. & Langdale,

J.A. (2013a). Evolution of GOLDEN2-LIKE gene function in C_3 and C_4 plants. *Planta*, 237, 481–495.

Wang, P., Kelly, S., Fouracre, J. & Langdale, J.A. (2013b). Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C_4 Kranz anatomy. *Plant Journal*, 75, 656–670.

Wang, P., Vlad, D. & Langdale, J.A. (2016). Finding the genes to build C_4 rice. *Current Opinion in Plant Biology*, 31, 44–50.

Wang, W., Haberer, G., Gundlach, H., Gläßer, C., Nussbaumer, T., Luo, M.C., Lomsadze, A., Borodovsky, M., Kerstetter, R.A., Shanklin, J., Byrant, D.W., Mockler, T.C., Appenroth, K.J., Grimwood, J., Jenkins, J., Chow, J., Choi, C., Adam, C., Cao, X.H., Fuchs, J., Schubert, I., Rokhsar, D., Schmutz, J., Michael, T.P., Mayer, K.F.X. & Messing, J. (2014b). The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. *Nature Communications*, 5. doi:10.1038/ncomms4311

Welch, D., Hassan, H., Blilou, I., Immink, R., Heidstra, R. & Scheres, B. (2007). *Arabidopsis* JACKDAW and MAGPIE zinc finger proteins delimit asymmetric cell division and stabilize tissue boundaries by restricting SHORT-ROOT action. *Genes & development*, 21, 2196–204.

Werner, T., Motyka, V., Laucou, V., Smets, R., Van Onckelen, H. & Schmülling, T. (2003). Cytokinin-deficient transgenic *Arabidopsis* plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. *Plant Cell*, 15, 2532–2550

Whitlock, M. & Schluter, D. (2009). *The analysis of biological data*. Roberts, Greenwood Village, Colo. 1st edn.

Williams, B.P., Burgess, S.J., Reyna-Llorens, I., Knerova, J., Aubry, S., Stanley, S. & Hibberd, J.M. (2016). An Untranslated cis-Element Regulates the Accumulation of Multiple C_4 Enzymes in *Gynandropsis gynandra* Mesophyll Cells. *The Plant Cell*, 28, 454–465.

Williams, B.P., Johnston, I.G., Covshoff, S. & Hibberd, J.M. (2013). Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis. *eLife*, 2, e00961.

Williams-Carrier, R., Stiffler, N., Belcher, S., Kroeger, T., Stern, D.B., Monde, R.A., Coalter, R. & Barkan, A. (2010). Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy Mutator lines of maize. *The Plant Journal*, 63, 167–177.

Woo, H.R., Koo, H.J., Kim, J., Jeong, H., Yang, J.O., Lee, I.H., Jun, J.H., Choi, S.H., Park, S.J., Kang, B., Kim, Y.W., Phee, B.K., Kim, J.H., Seo, C., Park, C., Kim, S.C., Park, S., Lee, B., Lee, S., Hwang, D., Nam, H.G. & Lim, P.O. (2016). Programming of Plant Leaf Senescence with Temporal and Inter-Organellar Coordination of Transcriptome in *Arabidopsis*. *Plant physiology*, 171, 452–467.

Xu, J., Bräutigam, A., Weber, A.P.M. & Zhu, X.G. (2016). Systems analysis of cis-regulatory motifs in C_4 photosynthesis genes using maize and rice leaf transcriptomic data during a process of de-etiolation. *Journal of Experimental Botany*, 67, 5105–5117.

Yan, T., Yoo, D., Berardini, T.Z., Mueller, L.A., Weems, D.C., Weng, S., Cherry, J.M. & Rhee, S.Y. (2005). PatMatch: a program for finding patterns in peptide and nucleotide sequences. *Nucleic Acids Research*, 33, W262–W266.

Yanagisawa, S. (1995). A novel DNA-binding zinc finger motif that may form a single. *Nucleic Acids Research*, 23, 3403–3410.

Yanagisawa, S. (2001). The transcriptional activation domain of the plant-specific Dof1 factor functions in plant, animal, and yeast cells. *Plant and Cell Physiology*, 42, 813–822.

Yang, Y., Zhong, J., Ouyang, Y.D. & Yao, J. (2013). The integrative expression and co-expression analysis of the AGO gene family in rice. *Gene*, 528, 221–235.

Yardley, S. (2013). *Characterisation of an AINTEGUMENTA-like mutant in maize*. Undergraduate project, University of Oxford.

Yi, G., Neelakandan, A.K., Gontarek, B.C., Vollbrecht, E. & Becraft, P.W. (2015). The naked endosperm Genes Encode Duplicate INDETERMINATE Domain Transcription Factors Required for Maize Endosperm Cell Patterning and Differentiation. *Plant physiology*, 167, 443–56.

Yu, N.I., Lee, S.A., Lee, M.H., Heo, J.O., Chang, K.S. & Lim, J. (2010). Characterization of SHORT-ROOT function in the *Arabidopsis* root vascular system. *Molecules and cells*, 30, 113–119.

Zeng, M., Hu, B., Li, J., Zhang, G., Ruan, Y., Huang, H., Wang, H. & Xu, L. (2016). Stem cell lineage in body layer specialization and vascular patterning of rice root and leaf. *Science Bulletin*, 61, 847–858.

Zhang, H., Lin, X., Han, Z., Qu, L.J. & Chai, J. (2016). Crystal structure of PXY-TDIF complex reveals a conserved recognition mechanism among CLE peptide-receptor pairs. *Cell Research*, 26, 543–555

Zheng, X., Deng, W., Luo, K., Duan, H., Chen, Y., McAvoy, R., Song, S., Pei, Y. & Li, Y. (2007). The cauliflower mosaic virus (CaMV) 35S promoter sequence alters the level and patterns of activity of adjacent tissue- and organ-specific gene promoters. *Plant cell reports*, 26, 1195–1203.

Zhou, J., Wang, X., Lee, J.Y. & Lee, J.Y. (2013). Cell-to-cell movement of two interacting AT-hook factors in *Arabidopsis* root vascular tissue patterning. *The Plant cell*, 25, 187–201.

Zhuang, J., Deng, D.X., Yao, Q.H., Zhang, J., Xiong, F., Chen, J.M. & Xiong, A.S. (2010). Discovery, phylogeny and expression patterns of AP2-like genes in maize. *Plant Growth Regulation*, 62, 51–58.

Zürcher, E., Tavor-Deslex, D., Lituiev, D., Enkerli, K., Tarr, P.T. & Müller, B. (2013). A Robust and Sensitive Synthetic Sensor to Monitor the Transcriptional Output of the Cytokinin Signaling Network in Planta. *Plant Physiology*, 161, 1066–1075.