The Genetic Regulation of Kranz Anatomy in Maize

Thomas Hughes

Lincoln College
University of Oxford



Thesis submitted for the degree of
Doctor of Philosophy

Trinity Term 2016

Abstract

The genetic regulation of Kranz anatomy in maize

Thomas Hughes, Lincoln College

DPhil, Trinity Term 2016

The C_4 photosynthetic pathway acts to concentrate CO_2 around the enzyme Ribulose-1,5-bisphosphate carboxylase oxygenase (Rubisco), ensuring that it catalyses a carboxylation rather than oxygenation reaction, which in turn suppresses photorespiration. In nearly all cases C_4 photosynthesis is underpinned by characteristic Kranz anatomy, with concentric wreaths of bundle sheath (BS) and mesophyll (M) cells surrounding closely spaced veins. The increased yields associated with the C_4 pathway have lead to the suggestion that C_3 crops such as rice should be engineered to undertake C_4 photosynthesis, however, this goal is currently held back by a lack of understanding about how the development of Kranz anatomy is regulated.

Recently, a number of candidate Kranz regulators have been identified in an RNA-seq study that compared leaf development in maize foliar (Kranz) and husk (non-Kranz) leaves. However, this study did not consider the impact of a recent whole genome duplication in the maize lineage on the gene expression patterns analysed. Therefore, in this thesis maize homeolog gene-pair divergence during early leaf development was assessed. This revealed that expression divergence of homeolog gene-pairs is a significant evolutionary phenomenon.

Functional validation of a subset of Kranz candidates revealed that a *Zmscr1-1; Zmscr1h-1* double mutant exhibited defects in Kranz patterning, including increased formation of extra BS cells and veins with no separating M cells. Furthermore, *Zmnkd1; Zmnkd2* double mutants exhibited a subtle increase in extra BS cell formation. Taken together, this indicates that both *ZmSCR1/ZmSCR1h* and *ZmNKD1/ZmNKD2* function redundantly during Kranz development. No evidence was obtained that two additional genes, *ZmSHR2* and *ZmRVN1*, play a role in Kranz development, and expression of candidate Kranz regulators in rice did not alter leaf anatomy. Together, this work has confirmed roles for a number of genes in Kranz regulation, and has provided insight into the complex regulation underpinning Kranz development in maize.

Acknowledgements

First, I wish to thank Jane Langdale for agreeing to supervise this DPhil, and for providing tremendous supervision and support, as well as introducing me into the world of maize genetics. I also wish to thank Steve Kelly for supervising my first rotation project and for always offering excellent advice throughout the rest of my DPhil.

Throughout the project I have benefitted from the support of members of the Langdale and Kelly labs, as well as the wider department. Special mention must go to the C_4 rice team (past and present): Mara and Jim got me started in the lab, Peng never said an unwise word, Sayuri taught me rice transformation, Zara was always willing to help, Dana arrived at the perfect time (and has tolerated my continual interruptions with an unfathomable amount of patience), David never pretended not to hear me when I asked about trees, Roxaana gave up time to shell maize seed, and Olga has been a tremendous fellow PhD student and friend to share this experience with. In addition, I want to especially thank Julie Bull for being the best lab manager imaginable, John Baker for helping with plant photography, and Gem Toes for always offering encouragement.

I am also enormously grateful to Gail Preston and the BBSRC Doctoral Training Programme, for funding and training, and the Newton Abraham Foundation for their generous funding support, which has enabled me to undertake this DPhil.

A large portion of this work relied on field space in which to grow maize. To this end, I owe a great deal of gratitude to Phil Becraft and Bryan Gontarek, who have gone far beyond the call of duty many times to facilitate this, and Erik Vollbrecht and Ruairidh Sawers who helped enormously with the winter crop. Furthermore, I would like to thank Natalie and Dan Rekemeyer, Jim Cornette, Bruce and Barbara Munson,

and Josh Strable for their incredible Iowa-hospitality to a wandering Brit. Finally, I wish to thank the late Caroline Cornette, who introduced me to many friends in Iowa, and was one of the most interesting and inspirational people I have ever met. She is greatly missed.

I also owe a huge amount of gratitude to Debbie and Tim Eldridge (among others) for being great teachers and working harder than I ever realised, and to Debbie in particular for showing me how interesting plants and photosynthesis are.

There are too many friends who have encouraged me over the past few years to list individually, but a few deserve special mention. First, Sophie, Izzy, Chris, Andy and Christine have been great housemates and even better friends. I also wish to thank Sandy for all of the mountain marathons that he didn't make me do, Jessie, El, Charlotte and Emily for departmental support, Josie, Jerome, Phil and Rhea for board games and red wine, and all of the Spartak Botley players who helped to maintain my sanity during my write-up. In addition, Danny (who kindly let me store maize DNA in his fridge), Dale, Sam and Neirin, despite never missing an opportunity to question what I was doing with my life or make the same tired rice jokes, have always been top-class friends. I also owe an enormous thanks to Anna for being generally amazing and unwaveringly supportive, especially over the past year.

Last but not least, I am incredible grateful to my family. My Mum and Dad have always provided endless encouragement, and Ben, Anna and Gran have always made going home something to look forward to immensely.

Contents

Abstract ..... **i**

Acknowledgements ..... **ii**

Contents ..... **iv**

List of Figures ..... **xii**

List of Tables ..... **xv**

Published work ..... **xvi**

Gene nomenclature ..... **xvii**

Abbreviations ..... **xviii**

Chapter 1 : Introduction **1**

1.1. The majority of plants utilise the C_3 photosynthetic pathway **1**

1.2. The C_4 photosynthetic pathway **3**

1.3. Spatial separation of the C_4 cycle **6**

1.3.1. Single cell C_4 photosynthesis 6

1.3.2. Kranz anatomy underpins C_4 photosynthesis in most species 7

1.3.3. Variations in Kranz form 8

1.4. The evolution of C_4 photosynthesis **11**

1.5. Harnessing the advantages of the C_4 pathway **15**

1.6. The development of Kranz anatomy in monocots **17**

1.6.1. Regulation of vein spacing 20

1.6.2. Regulation of bundle-sheath and mesophyll cell differentiation in C_4 species 23

1.7. Transcriptomics as a tool for identifying regulators of Kranz anatomy **24**

1.8. Radial patterning in the Arabidopsis root **27**

1.9. The role of INDETERMINATE DOMAIN (IDD) transcription factors in plant development **33**

1.10. A proposed model for the development of Kranz anatomy 36

1.11. Model species for studying Kranz anatomy development 39

1.12. The contribution of homeolog gene-pairs to maize leaf development 40

1.13. Thesis aims 41

Chapter 2 : Materials and methods 42

2.1. Plant material and growth conditions 42

2.1.1. Rice 42

2.1.2. Maize 42

2.2. Molecular cloning 43

2.2.1. Primer design 43

2.2.2. Gateway cloning 43

2.2.3. Golden gate cloning 44

2.2.4. Restriction digests 44

2.2.5. DNA sequencing 45

2.2.6. Bacterial transformation, growth and plasmid preparation 45

2.3. Plant transformation and tissue culture 46

2.3.1. Media 46

2.3.2. Rice callus induction 48

2.3.3. *Agrobacterium* mediated transformation of rice 48

2.3.4. Callus selection and regeneration 48

2.3.5. Plant rooting and transfer to soil 49

2.4. Maize mutant identification 49

2.5. DNA extraction and plant genotyping 49

2.5.1. CTAB DNA extraction 49

2.5.2. SDS 96-well plate extraction 50

2.5.3. Polymerase chain reaction 50

2.5.4. Gel electrophoresis 51

2.6. Expression analysis 53

2.6.1. RNA extraction 53
2.6.2. Reverse transcription-PCR 53
2.7. Plant phenotyping 54
2.7.1. Plant photography 54
2.7.2. Fresh sectioning 54
2.7.3. Tissue fixation, embedding and sectioning 55
2.7.4. Leaf clearing 57
2.7.5. GUS staining 57
2.8. Phylogenetic analysis 57
2.8.1. Tree construction 57
2.8.2. Enrichment analysis 58
2.8.3. Calculation of K_a/K_s ratios 58
2.8.4. Monte Carlo resampling tests 59
Chapter 3 : The impact of widespread regulatory neofunctionalisation on homeolog gene evolution following whole genome duplication in maize 60
3.1. Introduction 60
3.1.1. Whole genome duplications provide a source of evolutionary innovation ... 60
3.1.2. Maize underwent a recent whole genome duplication 61
3.2. Results 62
3.2.1. R-NF during maize early leaf development 62
3.2.1.1. Signatures of R-NF are apparent in over 400 homeolog gene-pairs that are expressed during leaf development 62
3.2.1.2. R-NF genes are more likely to have altered expression patterns in foliar leaves than in husk leaves 70
3.2.1.3. Functional annotations of R-NF homeolog genes support selection for distinct biological roles in foliar and husk leaves 71

3.2.2. The interaction between the maize subgenomes and R-NF 72
3.2.2.1. Subgenome-1 homeologs are more likely to be R-NF than subgenome-2 72
3.2.2.2. Biased subgenome expression dominance is only observed in R-NF homeolog gene-pairs 73
3.2.3. Selection acts differently on non-R-NF and R-NF homeolog gene-pairs 76
3.2.3.1. R-NF homeolog gene pairs are under stronger purifying selection and have more similar evolutionary rates than non-R-NF gene pairs 76
3.2.3.2. Dominant homeolog genes exhibit stronger purifying selection than their non-dominant gene-pair 78
3.2.4. Many high-priority candidate Kranz regulators have homeolog genes, which show evidence of expression divergence 78
3.3. Discussion 82
3.3.1. Expression divergence is a prominent feature of maize homeolog gene-pairs 82
3.3.2. Expression dominance is a dynamic feature of homeolog gene-pairs in maize 84
3.3.3. Expression divergence makes protein divergence less likely 85
3.3.4. Understanding homeolog divergence is critical for understanding transcriptional regulation of development in maize 86
Chapter 4 : The role of *ZmSCARECROW1* (*ZmSCR1*) and *ZmSHORTROOT2* (*ZmSHR2*) in the development of Kranz anatomy 89
4.1. Introduction 89
4.2. Results 91
4.2.1. Both *ZmSCR1* and *ZmSHR2* have duplicated gene-pairs 91
4.2.2. Transposon insertion mutants were identified for three of the four maize *ZmSCR1/ZmSHR2* genes 95

4.2.3. The role of *ZmSCR1* and *ZmSCR1h* in maize 99

4.2.3.1. The maize *Zmscr1-1; Zmscr1h-1* double mutant exhibits a reduced growth phenotype not seen in either of the single mutants 99

4.2.3.2. Endodermal cell specification and casparian strip formation in the root are disrupted in the double homozygous *Zmscr1-1; Zmscr1h-1* mutant 102

4.2.3.3. Delayed mid-rib formation causes a droopy leaf phenotype in *Zmscr1-1; Zmscr1h-1* double mutant plants 105

4.2.3.4. Vascular patterning is perturbed in the *Zmscr1-1; Zmscr1h-1* double mutant 105

4.2.3.5. BS and M cell patterning is perturbed in the *Zmscr1-1; Zmscr1h-1* double mutant 112

4.2.3.6. Alterations in foliar leaf patterning are not reflected in the husk leaves of *Zmscr1-1; Zmscr1h-1* double mutants 113

4.2.3.7. *ZmSCR* genes contain an N-terminal domain that shows some homology to that of *AtSCR* 119

4.2.4. The effect of *ZmSCR1* overexpression in rice 121

4.2.4.1. *ZmSCR1* overexpression constructs were cloned for rice transformation 121

4.2.4.2. Rice plants transformed with *ZmSCR1* overexpression constructs exhibited transgene-expression 121

4.2.4.3. No obvious differences in leaf cellular patterning were observed in rice plants overexpressing *ZmSCR1* 123

4.2.5. The role of *ZmSHR2* in maize 129

4.2.5.1. Neither plant nor leaf development is perturbed in the *Zmshr2-1* mutant 129

4.2.5.2. The *Zmscr1-1; Zmshr2-1* double mutant did not show additional phenotypic perturbations 135

4.3. Discussion 138

4.3.1. *ZmSCR1* and *ZmSCR1h* have been recruited from their ancestral C_3 function to redundantly regulate the development of Kranz anatomy.....138
4.3.2. ZmSCR1 and ZmSCR1h are proposed to act alongside ZmSHR to control Kranz development.....141
4.3.3. *ZmSCR1* and *ZmSCR1h* act redundantly in maize root development.....141
4.3.4. Functional validation of a role for *ZmSHR2* in Kranz development will require concurrent loss-of-function analysis of *ZmSHR2h*.....142
4.3.5. *ZmSCR1* and *ZmSCR1h* as confirmed Kranz regulators: prospects for C_4 rice engineering.....143
Chapter 5 : The role of maize IDD transcription factors during leaf development..... 144
5.1. Introduction.....144
5.2. Results.....146
5.2.1. *ZmNKD1* and *ZmRVN1* have maize-specific duplicates.....146
5.2.2. Maize transposon insertion alleles were identified for *ZmNKD1*, *ZmNKD2* and *ZmRVN1*.....150
5.2.3. The role of *ZmNKD1* and *ZmNKD2* in maize.....155
5.2.3.1. *Zmnkd1; Zmnkd2* double mutants exhibit alterations in leaf morphology 155
5.2.3.2. *Zmnkd1; Zmnkd2* double mutants exhibit more frequent occurrences of extra BS cell formation.....155
5.2.3.3. *Zmscr1-1; Zmnkd1-R; Zmnkd2-R* triple mutants did not show more severe leaf patterning defects than single *Zmscr1-1* or double *Zmnkd1-R; Zmnkd2-R* mutants.....160
5.2.4. The role of *ZmRVN1* in maize.....165
5.2.4.1. *Zmrvn1-4* single mutants do not display altered leaf anatomy.....165

5.2.4.2. *Zmrvn1-4; Zmnkd1-R; Zmnkd2-R* triple mutants did not show more severe leaf patterning defects than single *Zmrvn1-4* or double *Zmnkd1-R; Zmnkd2-R* mutants 173
5.2.5. The effect of *ZmNKD1* and *ZmRVN1* expression in rice 175
5.2.5.1. Constitutive expression of *ZmNKD1* and *ZmRVN1* prevented the regeneration of rice callus 175
5.2.5.2. Maize endogenous promoters were cloned for expression of *ZmNKD1* and *ZmRVN1* in rice 177
5.2.5.3. The *ZmNKD1* promoter is a strong shoot specific promoter in rice 177
5.2.5.4. Rice transformed with a $ZmNKD1_{pro}$*::ZmNKD1* construct did not set T1 seed 178
5.2.5.5. The *ZmRVN1* promoter confers early expression in rice leaves only when the *35S* promoter is used upstream to drive the hygromycin resistance gene 180
5.2.5.6. $ZmRVN1_{pro}$*::ZmRVN1* did not perturb rice leaf anatomy.....182
5.2.5.7. Inducible expression of *ZmRVN1* resulted in no alterations to leaf patterning in young rice leaves 186
5.3. Discussion 190
5.3.1. There is extensive redundancy between IDD genes 190
5.3.2. Evidence for functional interaction between *ZmNKD1/2* and *ZmSCR1/1h* 198
5.3.3. The lack of a developmental alterations in rice lines expressing *ZmNKD1* and *ZmRVN1* is surprising given the failure to recover constitutive expression lines 200
Chapter 6 : General discussion 202
6.1. Overview 202
6.2. A revised *SCARECROW* based model for radial patterning in maize leaves 203
6.3. The strengths and limitations of maize as a model species.....210

6.3.1. The maize whole genome duplication adds to potential genetic redundancy in maize....210
6.3.2. Maize transposon insertion resources are a valuable but limited resource for loss-of-function analysis....213
6.3.3. Transgenic approaches are necessary for higher-throughput analysis of Kranz candidate genes and refinement of the current model....214
6.4. How is C_3 leaf development regulated, and how does this differ from C_4 species?....216
6.5. Functional validation of candidate Kranz regulators in rice....217
6.6. Prospects for C_4 rice engineering....219

Appendices....222
Appendix 2: List of maize mutant alleles identified in this study....223
Appendix 4: Enrichment of annotation terms for each R-NF category relative to all annotated maize genes....224
Appendix 6: Ancestral profile inference by Bayesian model comparison....228
Appendix 7: Homeolog genes identified as being under positive selection....229
Appendix 9: Preliminary analysis of constitutive *ZmIDD14* expression in rice....230
Appendix 10: Representative genotyping of 17282 T1 rice lines....231
References....232

List of Figures

Figure 1.1	Cartoon depictions of the C_3 and C_4 metabolic pathways	5
Figure 1.2	Cartoon depictions of rice and maize leaf cross sections	18
Figure 1.3	Cartoon depiction of the current model of endodermis specification in *Arabidopsis* roots	29
Figure 1.4	Kranz development model in maize as proposed by Fouracre *et al.* (2014)	38
Figure 2.1	PCR genotyping strategy for maize mutants	52
Figure 3.1	Identification and quantification of regulatory neofunctionalised genes	64
Figure 3.2	Barplots showing expression dominance of homeolog gene-pairs during leaf development	75
Figure 3.3	Barplots displaying K_a/K_s ratios for homeolog gene-pairs	77
Figure 3.4	Average K_a/K_s ratios for non-R-NF genes where one gene dominates expression	79
Figure 4.1	Maximum likelihood phylogenetic tree of the GRAS transcription factors	92-93
Figure 4.2	Schematic of transposon insertions into three GRAS transcription factors	96
Figure 4.3	Gel images of genotype and expression analysis of GRAS *Mutator* transposon insertion alleles	97
Figure 4.4	Whole plant phenotype of *Zmscr* mutants and WT plants	100
Figure 4.5	Lineplots showing morphological phenotypic perturbations of mature *Zmscr* mutants and WT plants	101
Figure 4.6	Transverse sections of *Zmscr* mutants and WT roots	103
Figure 4.7	Midrib extension along the proximal/distal axis of *Zmscr* mutants and WT foliar leaves	106
Figure 4.8	Fresh transverse sections of foliar leaves from *Zmscr* mutant and WT plants	107-108
Figure 4.9	Quantification of vein-patterning traits in *Zmscr* and WT foliar leaves	110
Figure 4.10	Barplots displaying the frequency of extra BS cell occurrences in *Zmscr* and WT foliar leaves	114

Figure 4.11	Barplots displaying the frequency of occurrences of veins with no separating M cells in *Zmscr* and WT foliar leaves	115
Figure 4.12	Transverse sections of husk leaves from *Zmscr* and WT plants	117
Figure 4.13	Quantification of leaf anatomy traits for *Zmscr* and WT husk leaves	118
Figure 4.14	Alignment of *Arabidopsis* and maize SCR protein sequences	120
Figure 4.15	Cartoon depictions and transgenic line IDs for *ZmSCR1* overexpression constructs	122
Figure 4.16	Gel images showing expression of *ZmSCR1* in transgenic rice lines	124
Figure 4.17	Transverse fresh sections of Kitaake WT and *ZmSCR1* overexpressing rice leaves	126
Figure 4.18	Quantification of Kranz-related anatomical traits in rice lines overexpressing *ZmSCR1*	127
Figure 4.19	Transverse fresh sections of *Zmshr2-1* and *Zmscr1-1; Zmshr2-1* maize foliar leaves	130
Figure 4.20	Boxplots displaying quantification of vein-patterning in *Zmshr2-1* and *Zmscr1-1; Zmshr2-1* foliar leaves	131
Figure 4.21	Barplots displaying the frequency of extra BS cell occurrences between two lateral veins in *Zmscr1-1, Zmshr2-1* and *Zmscr1-1; Zmshr2-1* foliar leaves	133
Figure 4.22	Barplots displaying the frequency of occurrences of veins with no separating M cells between two lateral veins in *Zmscr1-1, Zmshr2-1* and *Zmscr1-1; Zmshr2-1* foliar leaves	136
Figure 4.23	Boxplots displaying normalisation of BS and M cell patterning traits for *Zmscr1-1; Zmshr2-1* double mutants	139
Figure 5.1	Maximum likelihood phylogenetic tree of the IDD transcription factors	148-149
Figure 5.2	Revision of *ZmRVN1h* gene model	151-152
Figure 5.3	Schematic of transposon insertions into IDD transcription factors	154
Figure 5.4	Gel images showing PCR amplified products from WT and homozygous *Zmnkd* mutant alleles	156
Figure 5.5	Whole plant phenotype of *Zmnkd* mutants and WT plants	158
Figure 5.6	Boxplots displaying leaf morphology traits for *Zmnkd* mutants and WT plants	159

Figure 5.7	Transverse fresh sections of *Zmnkd* mutants and WT foliar leaves	161
Figure 5.8	Boxplots displaying vein-patterning traits in *Zmnkd* and WT foliar leaves	162
Figure 5.9	Barplots displaying the frequency of extra BS cell occurrences between two lateral veins in *Zmnkd* and WT foliar leaves	164
Figure 5.10	Transverse fresh sections of *Zmscr1-1; Zmnkd1-R; Zmnkd2-R* triple mutants and WT foliar leaves	166
Figure 5.11	Gel image displaying genotyping and expression analysis of *Zmrvn1-4 Mutator* transposon insertion lines	168
Figure 5.12	Transverse sections of *Zmrvn1-4* mutants and WT foliar leaves	169
Figure 5.13	Boxplots displaying vein-patterning traits in *Zmrvn1-4* and WT foliar leaves	170
Figure 5.14	Barplots displaying BS cell numbers surrounding intermediate veins in *Zmrvn1-4* mutant and WT foliar leaves	171
Figure 5.15	Transverse fresh sections of *Zmrvn1-4; Zmnkd1-R; Zmnkd2-R* triple mutants and WT foliar leaves	174
Figure 5.16	Constructs used to investigate the role of *ZmNKD1* and *ZmRVN1* in rice	176
Figure 5.17	GUS activity driven by the *ZmNKD1* promoter in rice	179
Figure 5.18	Transverse sections of $ZmNKD1_{pro}$*::GUS* and $ZmNKD1_{pro}$*::ZmNKD1* T0 rice leaves	181
Figure 5.19	GUS activity driven by the *ZmRVN1* promoter in rice	183
Figure 5.20	Transverse fresh sections of $ZmRVN1_{pro}$*::ZmRVN1* and WT rice leaves	185
Figure 5.21	Lineplots showing vein density in leaves of $ZmRVN1_{pro}$*::ZmRVN1* rice lines	187
Figure 5.22	Barplots showing the number of M cells separating veins in leaves of $ZmRVN1_{pro}$*::ZmRVN1* rice lines	188
Figure 5.23	Barplots showing the number of BS cells surrounding veins in $ZmRVN1_{pro}$*::ZmRVN1* rice lines	189
Figure 5.24	Images of Kitaake WT and transgenic rice plants following Dex-induction of *ZmRVN1* expression	191
Figure 5.25	Gel images showing RT-PCR amplification products in WT and transgenic inducible *ZmRVN1* lines	192

Figure 5.26	Transverse fresh sections of Kitaake WT and transgenic rice leaves following Dex induction of *ZmRVN1* expression	193
Figure 5.27	Lineplot showing vein density in Kitaake WT and transgenic line rice leaves following Dex induction of *ZmRVN1* expression	194
Figure 5.28	Barplots displaying M and BS cell patterning in Kitaake WT and transgenic rice lines following Dex induction of *ZmRVN1* expression	195
Figure 6.1	Cartoon depictions of Kranz radial patterning models in maize	205

List of Tables

Table 1.1	Summary of the nine distinct forms of Kranz anatomy found in the Poaceae	10
Table 3.1	Frequency of each 4-letter expression profile across expressed homeolog gene-pairs	69
Table 3.2	Subset of candidate Kranz regulators with duplicate genes	81

Published work

Much of the work from Chapter 3 of this study was published as a research paper in Genome Research:

Hughes TE, Langdale JA, Kelly S. 2014. The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. *Genome Research* 24: 1348–55.

The author contributions were as follows: S Kelly came up with the initial project idea and provided supervision. TE Hughes concieved of the methodology concerning R-NF categorisation, partitioned the homeolog gene-pairs accordingly, and undertook all subsequent data analysis. TE Hughes undertook enrichment testing and K_a/K_s ratio calculations using general code previously written by S Kelly. TE Hughes came up with the logic underpinning ancestral expression profile inference, S Kelly formalised this into mathematical equations using Bayes rule (Appendix 6). TE Hughes wrote the paper with input from both S Kelly and JA Langdale.

Much of this paper is reproduced in Chapter 3 of this thesis, however, some additional sections have been added to help put the work into the broader context of the rest of the thesis. In addition, some sections from the paper have been expanded in this thesis to take into account relevant literature published since the original paper came out, and allow more detailed explanation where necessary.

Gene nomenclature

Genes, proteins and mutants from *Arabidopsis thaliana* and *Zea mays* (maize) are referred to throughout this thesis. Conventional genetic nomenclature differs between these species, but for consistency the following nomenclature system is used in this thesis:

- A capitalised letter for genus and lower case letter for species before gene names indicates which species is being referred to.
- Genes are indicated by italicised capital letters.
- Proteins are indicated by non-italicised capital letters.
- Recessive mutant alleles are indicated by italicised lower case letters followed by a dash and an allele reference (normally, but not always, a number). Alleles referred to are homozygous, unless separated by /.
- Higher order mutants (i.e. double or triple mutants) are indicated by the mutant alleles separated by a semicolon.

Examples:

AtSCR – *Arabidopsis thaliana SCARECROW* gene

ZmSCR1 – *Zea mays SCARECROW1* gene

AtSCR – *Arabidopsis thaliana* SCARECROW protein

Zmscr1-1 – Homozygous recessive mutant allele (allele reference = 1) in *ZmSCR1* gene

ZmSCR1/ Zmscr1-1 – Heterozygous for *Zmscr1-1* recessive mutant allele

Zmscr1-1; Zmscr1h-1 – Homozygous double mutant for recessive mutant alleles in both *ZmSCR1* and *ZmSCR1h* genes

Abbreviations

The following abbreviations are used in this thesis:

A – Ascending

ANOVA – Analysis of variance

At – *Arabidopsis thaliana*

Ac/Ds – Activator/Dissociation

BIB – BALD IBIS

BLAST – Basic local alignment search tool

BS – Bundle sheath

CA – Carbonic anhydrase

cDNA – Complementary DNA

CEI – Cortical endodermal initial cell

Co – Cortex

Cs – Casparian strip

D – Descending

DC – Distinctive cell

DE – Differential expression

DEPC – Diethylpyrocarbonate

Dex – Dexamethasone

DNA – Deoxyribonucleic acid

EDTA – Ethylenediaminetetraacetic acid

En – Endodermis

EV – Empty vector

FE – Foliar expanded

FI – Foliar immature

FP – Foliar plastochron

GO – Gene ontology

GUS – β-glucuronidase

HE – Husk exposed

HI – Husk inner

HP – Husk plastochron

HSD – Honest significant difference

Hyg – Hygromycin

IAA – Isoamyl alcohol

IDD – Indeterminate domain

JKD – JACKDAW

L – Litre

M – Mesophyll

MDH – Malate dehydrogenase

MGP – MAGPIE

Mu – *Mutator*

N – Neutral

NAD – Nicotinamide adenine dinucleotide

NAD-ME – NAD dependent malic enzyme

NADP-ME – NAD phosphate dependent malic enzyme

NKD – NAKED ENDOSPERM

NUT – NUTCRACKER

OE – Overexpression

Os – *Oryza sativa*

PEP – Phosphoenolpyruvate

PEPC – PEP carboxylase

PEPCK – PEP carboxykinase

PPDK – Pyruvate orthophosphate dikinase

Pro – Promoter

QC – Quiescent centre

R – Resistance

RNA – Ribonucleic acid

RNAi – RNA interference

RNA-seq – RNA sequencing

R-NF – Regulatory neofunctionalisation

RSEM – RNA-seq by expectation maximisation

R-SF – Regulatory subfunctionalisation

Rubisco – Ribulose-1,5-bisphosphate carboxylase oxygenase

RVN – RAVEN

SCL23 – SCARECROW-LIKE 23

SCR – SCARECROW

SHR – SHORTROOT

SNP – Single nucleotide polymorphism

TAE – Tris Acetate EDTA

TE – Tris-EDTA

Ubi – Ubiquitin

UV – Ultra-violet

WT – Wild type

Zm – *Zea mays*

Chapter 1: Introduction

1.1. The majority of plants utilise the C_3 photosynthetic pathway

Underpinning virtually all life on earth is the process of photosynthesis, whereby plants, algae and photosynthetic bacteria use light energy in order to fix inorganic carbon dioxide (CO_2) into organic carbohydrates. The enzyme ribulose-1,5-bis-phosphate carboxylase/oxygenase (Rubisco) is key to this process, as it catalyses the critical initial carbon fixation reaction between the five-carbon compound ribulose bis-phosphate (RuBP) and one-carbon compound CO_2. This reaction yields two molecules of the three-carbon compound 3-phosphoglycerate (3-PGA), which can then be used both in sucrose synthesis and the regeneration of RuBP (Bassham et al. 1956). This process is termed C_3 photosynthesis, owing to the initial creation of a three-carbon compound.

Despite its ubiquity, C_3 photosynthesis is hampered by the poor ability of Rubisco to distinguish between CO_2 and O_2, leading to competing carboxylation and oxygenation reactions (Lorimer 1981). Instead of yielding two molecules of 3-PGA, the unwanted oxygenation reaction yields one molecule of 3-PGA and one molecule of 2-phosphoglycolate (Lorimer 1981). Phosphoglycolate is toxic and must be transported out of the chloroplasts to be metabolised in peroxisomes and mitochondria, in a complex process termed photorespiration. Whilst photorespiration enables the rescue of some carbon, one quarter is lost as CO_2 during a glycine to serine conversion (Lorimer and Andrews 1973; Sharkey 1988). Both temperature and drought stress increase photorespiration rates. This is due to a combination of three factors: *1)* the selectivity of Rubisco decreases as temperature increases (Ku and Edwards 1977b; Jordan and Ogren 1984; Brooks and Farquhar 1985), *2)* O_2 solubility increases more than CO_2 solubility at high temperatures (Ku and Edwards

1977a), and *3)* drought stress causes stomata to close, reducing the amount of CO_2 around Rubisco (Cornic and Briantais 1991). In combination, this means that photorespiratory losses are an increased problem in hot and dry environments. The impact of climate change and the fact that many of our most important crops utilise C_3 photosynthesis, has resulted in considerable interest in combatting photorespiratory losses in order to increase crop yields (Ogren 2003; Peterhansel et al. 2010).

Theoretically, photorespiratory losses can be overcome either through improving the specificity of Rubisco for CO_2, or by reducing O_2 concentrations around Rubisco. Considerable variation in Rubisco CO_2 specificity has been found between organisms, but it has been proposed that there is a trade off, with gains in specificity being offset by concurrent decreases in the catalytic rate of the reaction (Jordan and Ogren 1981; Bainbridge et al. 1995; Tcherkez et al. 2006). Nevertheless, computational models have suggested that introducing Rubisco variants with higher CO_2 specificities into C_3 crops could increase canopy photosynthesis by up to 27% (Zhu et al. 2004; Long et al. 2006). Recently, the potential for introducing different Rubisco variants into crops has been demonstrated in tobacco by the successful introduction and correct assembly of the cyanobacteria *Synechococcus elongates* Rubisco large and small subunits. However, the relatively low CO_2 specificity of the *Synechococcus elongates* Rubisco meant that despite being capable of autotrophic growth the resulting plants were unable to survive under normal atmospheric CO_2 concentrations (Lin et al. 2014). This is presumably due to the absence of a carbon-concentrating carboxysome in tobacco, which in cyanobacteria enables Rubisco to exhibit lower specificities and higher catalytic rates (Lin et al. 2014).

Whilst there are numerous studies underway exploring Rubisco manipulation, given the apparent constraint between specificity and catalytic rate, it may be that alternative strategies for reducing photorespiratory losses are required. One

possibility is to engineer alternative photorespiratory bypass pathways. A number of attempts of this kind have been made in *Arabidopsis thaliana* (hereafter referred to as *Arabidopsis*), with two strategies providing evidence that the approach is feasible (Kebeish et al. 2007; Maier et al. 2012; Peterhansel et al. 2013). In both cases, the metabolic bypasses were designed such that CO_2 was released in the chloroplast rather than in the mitochondria, and as such it was hypothesised that this CO_2 could be refixed by Rubisco rather than lost (Peterhansel et al. 2013). Both approaches lead to an increase in plant growth and photosynthesis, although it was not clear whether this was due to increased CO_2 refixation or an improved overall energy balance (Kebeish et al. 2007; Maier et al. 2012; Peterhansel et al. 2013). Whilst these results are promising, the bypass strategy does not address the high oxygenation rates associated with Rubisco that are the underlying cause of photorespiration. Therefore, an alternative approach has been proposed to increase the effective CO_2 concentration (and thus reduce the effective O_2 concentration) around Rubisco. This suggestion is supported by the diverse set of carbon concentrating mechanisms employed by a range of photosynthetic organisms.

1.2. The C_4 photosynthetic pathway

A variety of strategies to concentrate CO_2 around Rubisco can be found in nature. These include the carbon-concentrating mechanisms of many cyanobacteria and green algae, and both crassulacean acid metabolism (CAM) and C_4 photosynthesis in flowering plants (Kaplan and Reinhold 1999; Vats et al. 2011; Edwards and Ogburn 2012). Of these divergent strategies, C_4 photosynthesis is the most agronomically important, because productive crop species such as maize and sugarcane are C_4, as are many weeds. Research into C_4 photosynthesis has seen a resurgence in the past five years, in part due to ambitious attempts to engineer C_3 crops such as rice to use the C_4 pathway (Langdale 2011).

Evidence for a non-C_3 photosynthetic pathway was first obtained when ^{14}C feeding experiments in both maize and sugarcane revealed that the four-carbon compounds malate and aspartate were among the first labelled products (Karpilov 1960; Kortschak et al. 1965). Subsequent experimental work confirmed these findings, and lead to a model for the C_4 dicarboxylic acid pathway being proposed (Hatch and Slack 1966; Hatch 2002). The C_4 pathway is now known to begin with a carbon fixation reaction that is catalysed by the oxygen-insensitive enzyme phosphoenolpyruvate carboxylase (PEPC), rather than Rubisco (Bowes and Ogren 1972). In maize and sugarcane, the cycle relies on six core enzymes. First, CO_2 is converted to HCO_3^- by carbonic anhydrase (CA) in the cytosol of mesophyll (M) cells. This HCO_3^- is then combined by PEPC with the three-carbon compound phosphoenolpyruvate (PEP), to form the four-carbon compound oxaloacetate, which is in turn transported into the chloroplast where it is converted into malate by Nicotinamide Adenine Dinucleotide Phosphate malate dehydrogenase (NADP-MDH). Malate is then transported out of the M cell chloroplast, and diffuses through plasmodesmata into an adjacent bundle-sheath (BS) cell where it is then transported into a Rubisco containing chloroplast. NADP-Malic Enzyme (NADP-ME) then decarboxylates malate into the three carbon compound pyruvate, with the release of one molecule of CO_2. This raises the CO_2 concentration around Rubisco to saturating levels, which in turn represses the oxygenation reaction and thus inhibits photorespiration. Finally, pyruvate is transported out of the BS chloroplast and diffuses through plasmodesmata into the M, where it is transported into a M chloroplast and converted to PEP by Pyruvate phosphate dikinase (PPDK). This PEP is returned to the M cytosol, thus restarting the cycle (Figure 1.1) (A full review of C_4 biochemistry is provided by Furbank 2016).

Whilst this cycle is representative of a large number of C_4 species, there are a number of variations. For example, three distinct decarboxylation enzymes, named

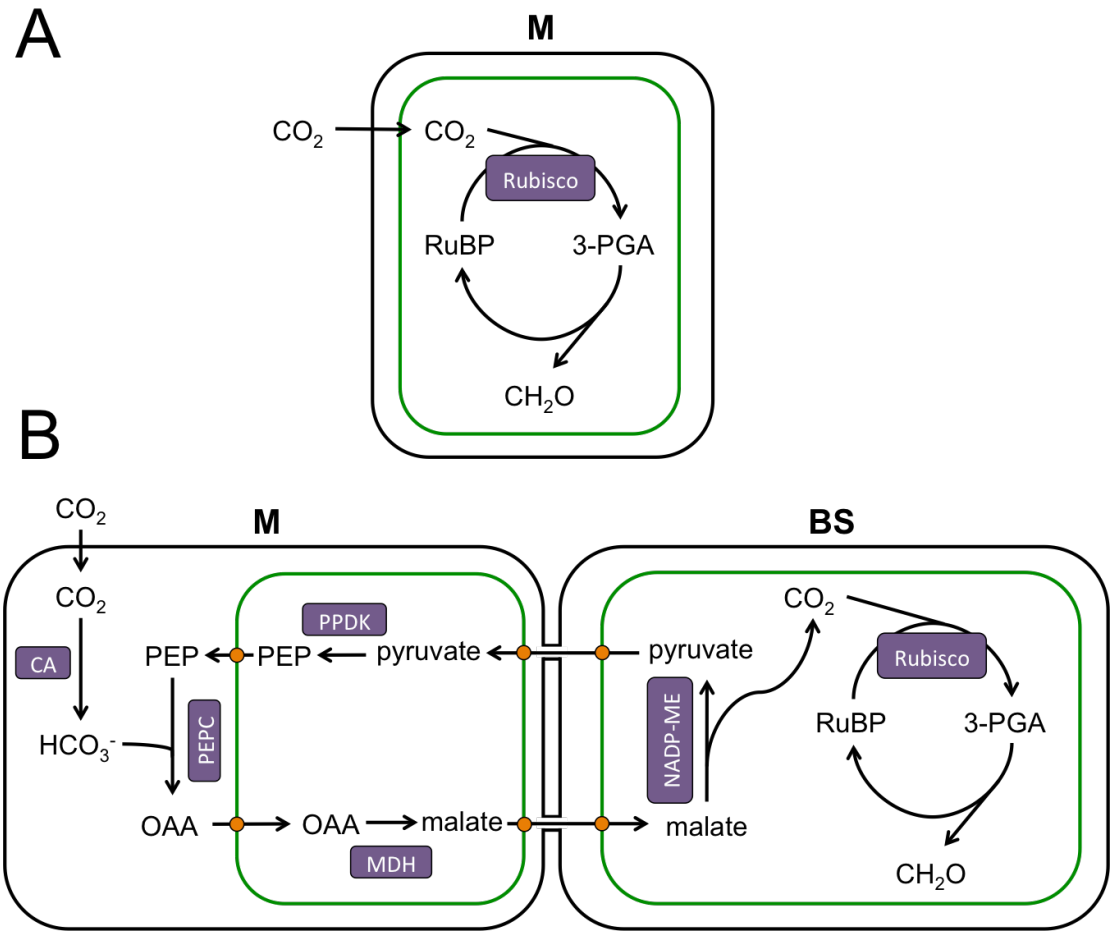


Figure 1.1 Cartoon depictions of the C_3 and C_4 metabolic pathways. A) The C_3 cycle operating in a M cell, **B)** The NADP-ME C_4 cycle compartmentalised between M and BS cells. Outer black lines represent cell membranes, green lines represent chloroplasts, purple rectangles represent enzymes, black arrows represent either chemical reactions or transport steps. Active transport across the chloroplast membrane is represented by orange circles. Channels between M and BS cells in (B) represent plasmodesmata. Adapted from Langdale (2011).

NADP-ME, NAD-ME and PEP-CK, are known to function in C_4 cycles, with species previously subdivided into three biochemical subgroups depending on the decarboxylation enzyme used (Drincovich and Lara 2011). Recent data, however, suggests that significant flexibility exists between these pathways, and it is now thought that only the NADP-ME and NAD-ME subtypes should be considered distinct, with both exhibiting secondary PEP-CK activity (Furbank 2011; Wang et al. 2014b). In addition, some species convert oxaloacetate to aspartate rather than malate, which once in the BS is transported into mitochondria before being converted back to oxaloacetate and then to malate for decarboxylation. In this version, instead of pyruvate being shuttled back to the M, it is first converted to alanine in the BS cytosol. Despite these points of variation, efficient function of the C_4 pathway relies on the spatial separation of the initial carboxylation catalysed by PEPC, and the site of decarboxylation around Rubisco.

1.3. Spatial separation of the C_4 cycle

1.3.1. Single cell C_4 photosynthesis

C_4 photosynthesis often relies on the spatial separation of initial carboxylation and decarboxylation between M and BS cells, as described for maize and sugarcane above. However, a number of species have emerged that carry out C_4 in a single cell (Reiskind et al. 1997; Voznesenskaya et al. 2001, 2002). One of the first to be discovered was *Borszczowia aralocaspica*, a dicot in the Chenopodiaceae family, which contains multiple C_4 species exhibiting conventional Kranz anatomy (Freitag and Stichler 2000; Voznesenskaya et al. 2001). *B. aralocaspica* forms only a single layer of palisade cells, rather than the two distinct cell layers thought to be essential for C_4 function, with dimorphic chloroplasts forming at the distal and proximal ends of the cell (Voznesenskaya et al. 2001, 2003; reviewed by Erlinghaeuser et al. 2016). Since this discovery, further single-celled C_4 variants have been identified in both terrestrial and aquatic plant lineages (reviewed by Edwards et al. 2004; Bowes

2011). Interestingly, examples of inducible single-cell C_4 species are known, for example *Hydrilla verticillata* is able to shift to a C_4 cycle in response to low CO_2 conditions (Bowes and Salvucci 1984; Reiskind et al. 1997). Therefore, it is now clear that the spatial separation of initial carbon fixation and decarboxylation necessary for a functional C_4 cycle can be achieved within the same cell, however, the majority of C_4 species (and all domesticated C_4 crops) achieve this spatial separation via distinct cell-types.

1.3.2. Kranz anatomy underpins C_4 photosynthesis in most species

The majority of C_4 species separate photosynthetic reactions between M and BS cells. In C_3 leaves, veins are separated by an encircling BS cell layer, between 6 and 20 M cells and the BS cell layer of the next vein. Rubisco is primarily located in the M cells, where C_3 carbon fixation occurs. In C_4 leaves vein density is increased such that veins are often separated by the BS cell layers and just two M cells. This arrangement, with concentric rings of BS and M cells surrounding each vein, leads to a 'wreath-like' appearance, and hence is termed Kranz anatomy (the German word for wreath) (Haberlandt 1896). In addition to the specific cellular arrangement, Kranz BS cells are often enlarged compared to their C_3 counterparts, likely reflecting their important photosynthetic role (Nelson and Dengler 1992). In C_3 species, proposed BS cell roles include loading of metabolites into the vasculature, structural support and cavitation repair (Koroleva et al. 2000; Griffiths et al. 2013). The increase in vein density, reduction in M cell number, and increase in BS cell size that are characteristic of Kranz anatomy, act to minimise diffusional distances between M and BS cells. The resultant BS:M ratio of approximately one reflects the joint contribution of both cell-types to carbon fixation in C_4 plants.

Two further anatomical innovations often associated with Kranz anatomy are the development of dimorphic chloroplasts, and the presence of a suberin lamella (Laetsch and Price 1969; Kirchanski 1975; Hattersley and Browning 1981). In C_3

species, very few chloroplasts are found in BS cells, however, C_4 BS cells contain many enlarged chloroplasts that are distinct from C_4 M cell chloroplasts (Rhoades and Carvalho 1944; Bisalputra et al. 1969). BS chloroplasts in NADP-ME C_4 species often contain few grana (thylakoid stacks in the chloroplast stroma) with less photosystem II, whereas M chloroplasts form extensive grana with both photosystems present (Laetsch and Price 1969; Woo 1970; Sheen and Bogorad 1988). This is thought to enable the extra ATP required in the BS of NADP-ME C_4 species to be provided via cyclic electron flow around photosystem I (Takabayashi et al. 2005). A further feature of C_4 dimorphic chloroplasts is that BS chloroplasts are arranged either centrifugally or centripetally, whereas M chloroplasts are randomly positioned (Brown 1960). Centripetal chloroplast positioning is thought to increase the diffusion distance to the M cells, which prevents CO_2 leakage out of the BS (Hattersley and Browning 1981; von Caemmerer and Furbank 2003). In C_4 species with centrifugal BS chloroplasts, it has been suggested that leakage is reduced by the presence of a suberin layer between the BS and M cells (Hattersley and Browning 1981). In summary, Kranz anatomy features a number of specialisations, all of which act to enable effective partitioning of the C_4 cycle between two distinct cell-types, and thus enable efficient C_4 cycle operation.

1.3.3. Variations in Kranz form

The Kranz anatomy described in section 1.2.3 is representative of the classical NADP-ME subtype, which is found in both maize and sugarcane. There is, however, a great deal of variation in Kranz forms between species. These variations are discussed below in relation to the Poaceae grasses (of which maize and sugarcane are both members), however, they also apply equally to other C_4 lineages, both monocot and dicot.

There are four subfamilies in the Poaceae that contain C_4 species: the Panicaoidaea, Micrairoideae, Chloridoideae and Aristidoideae, with at least 9 distinct Kranz forms (Duvall et al. 2007; Christin et al. 2008; Edwards and Voznesenskaya 2011). These 9 Kranz forms are termed classical NADP-ME, classical NAD-ME, classical PEP-CK (all named after the decarboxylation enzyme believed to operate in each biochemical subtype), Arundinelloid, Aristidoid, Stipagrostoid, Eriachneoid, Neurachneoid and Triodioid (reviewed by Edwards and Voznesenskaya 2011). The precise anatomical differences for all of these subtypes are summarised in Table 1.1, with common points of variation and similarity discussed below.

Five key areas of Kranz variation were highlighted by Edwards and Voznesenskaya (2011). The first of these is the number of BS cell layers formed. Whilst the majority of Kranz subtypes have a single layer of BS cells, the Aristidoid and Neurachneoid subtypes form a double layer (Brown 1958; Dengler et al. 1985). In both, the inner ring resembles the typical C_4 Kranz BS cells, whereas the outer ring contains either mesophyll-like chloroplasts with extensive grana (Aristidoid) or few chloroplasts (Neurachneoid) (Hattersley et al. 1986; Voznesenskaya et al. 2005). The second point of variation is the presence or absence of a mestome sheath (MS), which is formed in some subtypes between the vasculature and the BS (Edwards and Voznesenskaya 2011). The third variable is the presence or absence of a suberin layer between the BS and M, which is present in many of the subtypes, but is absent in some, including the classical NAD-ME subtype (Hattersley and Browning 1981). The final two points of variation relate to chloroplast positioning in the BS and chloroplast dimorphism between the BS and M. In the classical NAD-ME subtype, chloroplasts are located centripetally, whereas in NADP-ME species they are located centrifugally (Brown 1960). In NADP-ME species BS chloroplasts are agranal and M chloroplasts granal, however, this pattern is reversed in the NAD-ME subtype which forms granal BS chloroplasts and agranal M chloroplasts (Gutierrez et al. 1974). This

Subtype	Summary
Classical NADP-ME	No mestome sheath, BS chloroplasts centrifugal and lacking grana, M chloroplasts have extensive grana, suberin layer normally present.
Classical NAD-ME	Mestome sheath present, BS chloroplasts centripetal and have grana, M chloroplasts agranal, suberin layer absent.
Classical PEP-CK	Mestome sheath present, BS and M chloroplasts have grana, BS chloroplasts usually centrifugal.
Arundinelloid	Similar to classical NADP-ME with the additional formation of distinctive cells (DCs), similar to BS but not associated with a vascular bundle.
Aristidoid	No mestome sheath, two BS cell layers (inner Kranz), Kranz BS chloroplasts agranal and centrifugal. No suberin layer.
Stipagrostoid	Mestome sheath present and formed of larger thin-walled cells, BS chloroplasts agranal and centripetal, no suberin layer.
Eriachneoid	Similar mestome sheath to stripagrostoid, BS chloroplasts with grana and centrifugal
Neurachneoid	No mestome sheath, double BS layer (inner Kranz), BS chloroplasts with grana centrifugal or peripheral, suberin layer present
Triodioid	Thin-walled mestome sheath, BS extensions present, BS chloroplasts with grana and centrifugal, suberin layer absent.

Table 1.1 Summary of the nine distinct forms of Kranz anatomy found in the Poaceae. Adapted from Edwards and Voznesenskaya (2011).

difference is thought to reflect the differing M and BS ATP requirements of the NADP-ME and NAD-ME pathways (Takabayashi et al. 2005). One further anatomical feature of note is the presence of 'distinctive cells' in the Arundinelloid subtype, which were first described in Arundinella but subsequently found also in Arthraxon and Microstegium (Tateoka 1958; Ueno 1995). Although not fully understood, distinctive cells are similar to BS cells, both in terms of ultrastructure and C_4 enzyme expression, despite not being associated with a vascular bundle (Crookston and Moss 1973; Ueno 1995; Wakayama et al. 2003, 2006). Even within the grasses, therefore, there is large variation in Kranz forms, presumably reflecting the multiple evolutionary origins of C_4 in this lineage (see section 1.4).

Given the observed variation in Kranz forms, what can be considered a fundamental Kranz trait? The main aspect that does not vary across C_4 lineages is the presence of concentric layers of an outer M cell layer and inner BS cell layer(s). The outer M cell layer is ideally placed to fix CO_2 that diffuses into the leaf through stomata, whereas the inner BS layer allows for decarboxylation at the site of Rubisco. Either through an increase in vein density and BS cell size, or the presence of distinctive cells between veins, all Kranz subtypes have an increased BS:M ratio relative to C_3 species, and it is this which is the key anatomical shift associated with Kranz anatomy.

1.4. The evolution of C_4 photosynthesis

Despite the variations exhibited in both biochemistry and anatomy, C_4 photosynthesis remains one of the most striking examples of convergent evolution. The latest estimates suggest that there are 66-68 angiosperm lineages containing C_4 species, with eudicots and monocots accounting for 36 and 30-32 of the lineages respectively (Sage et al. 2011a, 2012; Grass Phylogeny Working Group II 2012). Whereas the evolutionary independence of a small number of these lineages is difficult to validate,

the majority indisputably represent independent origins of C_4. There has been remarkable convergence on largely the same suite of biochemical reactions and anatomical specialisations. For example, Kranz anatomy is found in all but four of the lineages (Sage et al. 2011a). In addition to C_4, around 21 lineages contain C_3-C_4 intermediate species, with 9 of these containing no C_4 representatives (Sage et al. 2011a). Recently, the application of modelling approaches to study C_4 evolution revealed that these C_3-C_4 intermediate species are most likely on the way to evolving C_4 photosynthesis, rather than representing evolutionary dead-ends (Heckmann et al. 2013; Williams et al. 2013). Specifically, it was found that when appropriate enabling factors are present, C_4 evolution follows a smooth single peaked Mt. Fuji like fitness trajectory, whereby each step towards C_4 increases overall fitness (Heckmann et al. 2013). However, by assessing and modelling C_4 evolution in different phylogenetic backgrounds, Williams *et al.* (2013) found that whilst similar overall changes were necessary to evolve C_4, there was significant flexibility in the exact order of C_4 trait acquisition (Williams et al. 2013). Together, these studies have provided a great deal of insight into the mechanisms underpinning C_4 evolution.

Given the advantage of C_4 in conditions where photorespiration is a problem, it is not surprising that the first C_4 origins occurred alongside a drop in atmospheric CO_2 around 30 million years ago (Pagani et al. 2005; Christin et al. 2008). Although this fall in CO_2 has long being proposed as the primary driving force behind C_4 evolution (Ehleringer et al. 1991), many origins have been dated much later, suggesting a more complicated picture. The CO_2 concentration at which C_4 becomes advantageous over C_3 is higher at higher temperatures, and as such it was predicted that C_4 origins should correlate with temperature (Edwards and Smith 2010). Instead, 18 out of 20 origins in the grasses correlated far better with a reduction in mean annual precipitation, and a reduction in forest cover that resulted in more open grassland (Edwards and Smith 2010). This observation supports a hypothesis in

which lower water availability caused by reduced precipitation, and exposure to higher temperatures because of open habitats, selected for C_4 as a water conserving method (Osborne et al. 2012). In the eudicots, a similarly complex picture is emerging, with C_4 evolution in the Chenopodiaceae most abundant in drought tolerant saline resistant lineages (Kadereit et al. 2012). Rather than being fully explained by one ecological driver, it is instead more likely that C_4 was promoted by a range of environmental factors that increased photorespiration, depending on the physiological and environmental context of the lineage in question (Christin and Osborne 2014).

The independent origins of C_4 are not evenly distributed across the angiosperm phylogeny, with many clades that have experienced seemingly conducive environmental conditions for C_4 evolution, containing no C_4 species. Instead, C_4 origins are clustered in certain clades. One pertinent example is the PACMAD clade (named as such because it contains the Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, and Danthonioideae subfamilies), which accounts for all 18 of the C_4 lineages found in the grasses (Christin et al. 2008; Sage et al. 2011a). The sister BEP clade (Bambusoideae, Ehrhartoideae and Pooideae) contains no C_4 lineages. This distinction suggests that a trait was selected for at the base of the PACMAD clade that subsequently enabled the evolution of C_4. It has long been suggested that anatomical pre-adaptations are one of the enablers that makes the predicted smooth Mt Fuji-like fitness landscape of C_4 evolution accessible (Sage 2001; Heckmann et al. 2013). To test whether certain anatomical traits made C_4 evolution more likely in the PACMAD grasses, Christin *et al.* (2013) measured a range of anatomical traits across 157 PACMAD and BEP species. It was found that the probability of C_4 evolution correlated with an increased proportional BS cell area of the leaf (Christin et al. 2013). Despite a reduction in the distance between BS cells (which resulted in an increased relative BS cell area) in species at the base of both

the PACMAD and the BEP clades, BS cell size was subsequently reduced in the BEP clade but not in the PACMAD clade. Therefore, Christin *et al.* (2013) hypothesised that this reduction in BS cell size (and the associated lower relative BS area of the leaf), meant that the potential fitness gains associated with C_4 photosynthesis were inaccessible to BEP species (Christin et al. 2013; Heckmann et al. 2013).

There is further evidence in the eudicots that supports the selection of related anatomical traits prior to the evolution of C_4. The genera *Flaveria* and *Heliotropium* have been particularly informative, because they contain a number of C_3-C_4 intermediate species. In *Flaveria*, phylogenetic relationships have been resolved such that the most closely related C_3 species (*F. pringlei* and *F. robusta*) to the C_3-C_4 intermediate and C_4 species are known (McKown et al. 2005; Kapralov et al. 2011). *F. robusta* exhibits higher vein densities than C_3 *Flaveria* species that are more distantly related to the intermediates (McKown and Dengler 2007). Further analysis revealed that both *F. pringlei* and *F. robusta* formed 'proto-Kranz' anatomy, with higher BS chloroplast and mitochondria numbers, and proportionally more of these organelles localised to the cell periphery (Sage et al. 2013). These anatomical changes are hypothesised to facilitate the evolution of a C_2 photorespiratory pump, which is thought to be an essential metabolic stepping stone towards toward C_4 photosynthesis (Sage et al. 2012; Heckmann et al. 2013; Sage et al. 2013; Schulze et al. 2013; Williams et al. 2013; Mallmann et al. 2014). These conclusions are supported by work in *Heliotropium*, which also contains C_3 species that form 'proto-Kranz' anatomy similar to that revealed in *Flaveria* (Muhaidat et al. 2011). Furthermore, increased vein density has been found in a range of C_3 species closely related to various independent C_4 lineages across the eudicots (Marshall et al. 2007; McKown and Dengler 2007; Voznesenskaya et al. 2007; Christin et al. 2011; Muhaidat et al. 2011; Sage et al. 2011b). As increased vein density is known to

increase the BS:M ratio, these findings support work in grasses that found increased relative BS area is an important anatomical pre-adaptation for evolving C_4.

In addition to anatomical traits, a number of other factors have been proposed as being important for facilitating C_4 evolution. One example is the acquisition of novel regulatory sequences, or trans-acting factors, that promote appropriate cellular compartmentation of C_4 pathway components (Brown et al. 2011; Sage et al. 2012). Another is the duplication of key genes, which would have allowed functional divergence without the accumulation of deleterious mutations (Monson 2003; Wang et al. 2009; Sage et al. 2012). Many of the C_4 cycle genes have undergone recent duplication (Rondeau et al. 2005; Christin and Besnard 2009; Wang et al. 2009; Sage et al. 2012), and whole genome duplication events at the base of the grasses are proposed to have provided much of the novel genetic material facilitating C_4 evolution (Paterson et al. 2004; Wang et al. 2009).

1.5. Harnessing the advantages of the C_4 pathway

By concentrating CO_2 around Rubisco and thus suppressing deleterious oxygenation reactions, C_4 crops exhibit up to 50% higher yield than their C_3 counterparts (Sheehy et al. 2007; Sage and Zhu 2011). This increased carbon capture efficiency also results in higher water use efficiency relative to C_3 plants (up to 50% higher) because lower stomatal conductance is needed to achieve the same photosynthetic rate, and higher nitrogen use efficiency (up to 40% higher) because less nitrogen is allocated to Rubisco (Brown 1978; Sage and Pearcy 1987; Brown 1999). These advantages are particularly pronounced in hot, dry environments where photorespiratory costs are highest. Despite making up only ~3% of the estimated 250,000 land plant species, C_4 plants are responsible for 21% of the total primary productivity on the planet (Lloyd and Farquhar 1994; Sage 2004). However, only two of the twelve most valuable crop species utilise the C_4 pathway: maize and sugar-cane (Sage and Zhu

2011). To exploit the productivity of C_4 photosynthesis for food security in the future, either more C_4 crops must be consumed, or existing C_3 crops need to be engineered to use the C_4 pathway (Sage and Zhu 2011). Despite the technological challenge, engineering the C_4 pathway into C_3 species may be the most feasible option given long-standing cultural preferences in many areas of the world for C_3 grains such as rice (Sage and Zhu 2011). Roughly half of the world's population relies on rice as a staple cereal crop, and unlike wheat or maize nearly all of the global annual rice crop is consumed directly by humans (Sheehy et al. 2007). Given predicted population increases, rice yields need to increase by 50% by 2050, but yield increases through traditional breeding have plateaued at less than 1% a year (Ray et al. 2012). Thus, attempts are being made to engineer the C_4 pathway into rice (Hibberd et al. 2008).

Given that the C_4 cycle can operate in a single-cell (Reiskind et al. 1997; Voznesenskaya et al. 2001), and that rice currently uses one cell-type to photosynthesise, it might be expected that engineering a single-cell C_4 cycle would be the most feasible strategy for C_4 rice. Indeed, this has already been attempted, with C_4 enzymes from grasses overproduced in rice M cells, with the aim of producing a *Hydrilla verticillatai*-like C_4 cycle (Taniguchi et al. 2008; Miyao et al. 2011). Unfortunately, this approach did not result in a functional CO_2 concentrating mechanism (Taniguchi et al. 2008). A counter-argument to the single-cell approach is that all domesticated C_4 crop species exhibit conventional Kranz anatomy, and in general Kranz-C_4 species are far more widespread than their single-cell counterparts. There are no examples of single-cell C_4 species closely related to rice, whilst there are many examples of Kranz C_4 grasses. Therefore, in order to fully harness the benefits of the C_4 cycle it will most likely be necessary to introduce Kranz type leaf anatomy into rice.

Engineering a full Kranz-type C_4 photosynthetic pathway into rice is a considerable challenge, involving complex changes in leaf anatomy and biochemistry. However,

encouragement comes from the convergent evolution of C_4 photosynthesis (Sage 2004; Sage et al. 2011a) (see section 1.4). This has raised hopes that although C_4 represents a complex trait, the switch from C_3 to C_4 might be relatively simple (Hibberd et al. 2008). Further encouragement has been provided by the discovery that aspects of C_4 photosynthesis operate in the stems of C_3 plants (Hibberd and Quick 2002). PEPC operates in the roots of C_3 plants, resulting in the presence of malate in the xylem stream (Arnozis et al. 1988), and BS cells immediately adjacent to the shoot vasculature of tobacco plants are capable of decarboxylating the malate, with the released CO_2 fixed into starch (Hibberd and Quick 2002). These results point to a spatially separated C_4 cycle operating in C_3 plants, which suggests that engineering the C_4 cycle into rice may not be an impossible goal. However, attempts to engineer a Kranz-type C_4 photosynthetic pathway are currently held back by our limited of understanding of how Kranz anatomy develops in C_4 species.

1.6. The development of Kranz anatomy in monocots

In C_4 monocot leaves, the increase in vein density associated with Kranz anatomy is achieved through the development of an additional class of intermediate veins that are not observed in C_3 monocot leaves. These have been inconsistently referred to previously as either ‘small’ or ‘minor’ intermediate veins (Sharman 1942; Russell and Evert 1985; Bosabalidis et al. 1994; Nelson and Dengler 1997), but are here termed rank-2 intermediate veins. C_3 and C_4 monocots both form lateral veins and rank-1 intermediate veins (Figure 1.2). Lateral veins (which include the midvein) have at least two large metaxylem elements, as well as protoxylem, metaphloem and protophloem elements (though these may be squashed at maturity) (Sharman 1942; Russell and Evert 1985). In addition, lateral veins form sclerenchyma tissue that extends both adaxially and abaxially to the epidermis, resulting in BS cells being present only on the left and right side of the bundle (Sharman 1942; Russell and Evert 1985). Lateral veins are separated by several rank-1 intermediates (Nelson

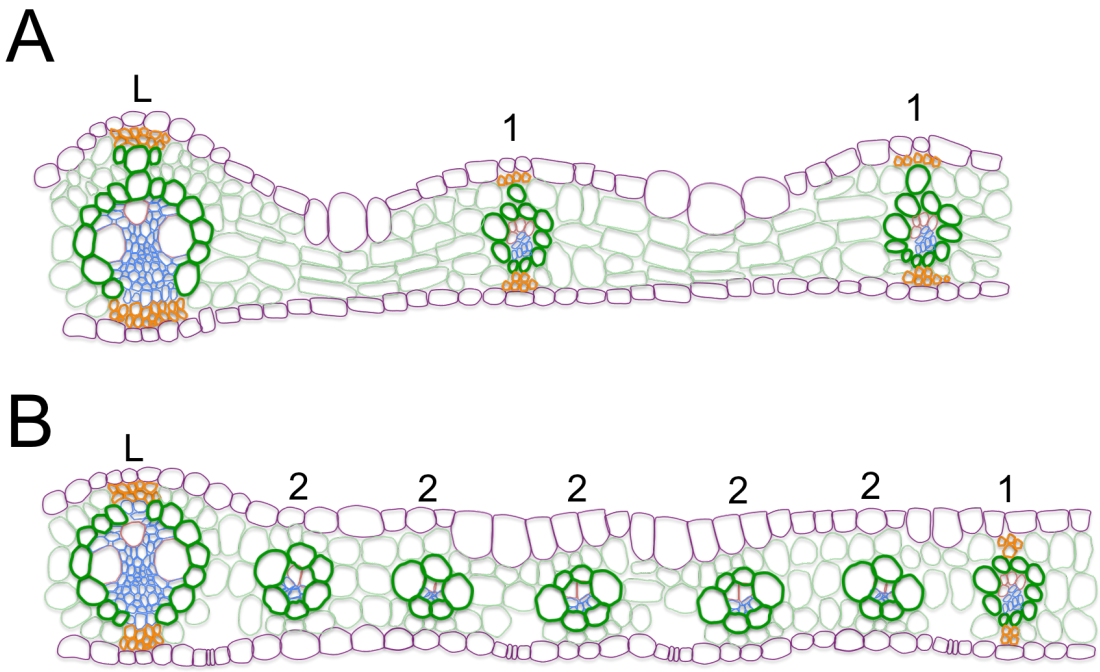


Figure 1.2 Cartoon depictions of rice and maize leaf cross sections. A) Rice **B)** Maize. In both sections purple represents the epidermis, dark green represents BS cells, light green represents M cells, blue represents vasculature, red represents metaxylem elements and orange represents sclerenchyma. L indicates a lateral vein, 1 indicates a rank-1 intermediate vein, and in (B) 2 indicates a rank-2 intermediate vein. Cartoons based on a template provided by Peng Wang.

and Dengler 1997). Rank-1 intermediates also form metaxylem (although they are not large), metaphloem, and may form protophloem (Russell and Evert 1985). They do not, however, form protoxylem elements. Sclerenchyma is present but now links from the BS to the epidermis, such that the BS cells form a continuous ring around the vasculature, and is not necessarily present on both the adaxial and abaxial sides of the leaf (Russell and Evert 1985). In C_3 leaves between 6 and 20 M cells separate these rank-1 intermediate veins. However, in C_4 leaves, vein density is increased through the development of several rank-2 intermediate veins between the rank-1 intermediates (Figure 1.2) (Nelson and Dengler 1997). Rank-2 intermediates have only metaxylem and metaphloem, and form no sclerenchyma (Sharman 1942; Russell and Evert 1985; Bosabalidis et al. 1994). Therefore, the challenge of engineering C_4 into a C_3 leaf requires an understanding of how the development of rank-2 intermediate veins is regulated.

Numerous studies have described the ontogeny of vascular development in maize (Sharman 1942; Esau 1943; Nelson and Dengler 1997). After a new primordium is initiated, procambial tissue that will form the midvein is laid down acropetally (from base to tip) (Esau 1943; Hall and Langdale 1996). Parallel lateral veins are formed approximately one plastochron (the time interval between primordia initiation) later than mid-vein differentiation, developing both acropetally and basipetally (from tip to base) (Nelson and Dengler 1997). In contrast, both rank-1 and rank-2 intermediate veins develop just basipetally. Lateral vein initiation and extension is completed within the first few plastochrons following leaf initiation, however, intermediate vein initiation continues beyond this (Nelson and Dengler 1997). Crucially, as intermediate veins are initiated and differentiate, they do so in a fixed position relative to neighbouring veins, and appear to use existing veins as a positional landmark (Nelson and Dengler 1997). Moreover, vein spacing is tightly regulated, with a linear

relationship between vein number and leaf blade width found in grasses (Russell and Evert 1985; Dannenhoffer et al. 1990).

The second critical phase of Kranz development, after procambium specification and vein formation, is the radial patterning of C_4 BS and M cells. Previous clonal analysis revealed that BS and M cells differentiated according to position rather than cell lineage (Langdale et al. 1989). Furthermore, BS and M cells differentiate and accumulate the relevant C_4 enzymes in coordination with the vein they surround, (i.e. cells around the midvein develop first and those around rank-2 intermediate veins develop later), raising the possibility that BS and M cell differentiation is somehow linked to vasculature differentiation (Mayfield and Taylor 1984; Langdale et al. 1987). This suggestion is supported by cellular differentiation patterns in the husk leaves of maize. Husk leaves have low vein density, with as many as 20 M cells separating each pair of veins, however BS and M cells that are immediately adjacent to a vein develop as C_4 (Langdale et al. 1988; Pengelly et al. 2011). Taken together, these observations indicate a model whereby the development of C_4-like BS and M cells is likely caused by a signal originating in the vasculature (Langdale and Nelson 1991).

The challenge to understand how Kranz anatomy develops is thus two-fold: 1) Understanding how rank-2 intermediate veins are initiated; and 2) Understanding the nature of the signal originating in the veins that promotes the radial patterning of C_4 BS and M cells.

1.6.1. Regulation of vein spacing

Vascular tissue differentiates from procambium, and thus vein density is determined by the intervals at which procambium is specified across developing primordia. Insight into general mechanisms of vascular development has been provided by studies in C_3 dicots (predominantly in *Arabidopsis*), which have revealed that the phytohormone auxin plays a crucial role in determining where a vein will be initiated.

It was demonstrated in the 1950's that following wounding of a *Coleus blumei* (another C_3 dicot) stem, a leaf-derived signal triggered xylem regeneration, and that this response could also be promoted by exogenous auxin application (Jacobs 1952). It is now known that auxin is synthesised in young leaves, with free-auxin levels peaking at the tips of developing primordia (Aloni et al. 2003). Notably, auxin localisation patterns predict sites of procambium differentiation in *Arabidopsis* leaves (Mattsson et al. 2003; Scarpella et al. 2010). Auxin flow and localisation is mediated by the auxin efflux PIN-FORMED (PIN) transporters, which determine the direction of auxin flow via their asymmetric cellular distribution (Wisniewska et al. 2006). The *Arabidopsis* genome encodes eight PIN proteins (Krecek et al. 2009), of which *PIN1* is thought to be the most important in relation to vein formation, due to its restricted expression at sites where subsequent procambium differentiation occurs (Scarpella et al. 2006). Thus, altered regulation of PIN1 localisation is likely to be a key step in the development of higher vein densities in Kranz exhibiting C_4 plants.

In *Arabidopsis*, mutant screens have identified several lines with defective vein-patterning (Koizumi et al. 2000; Deyholos et al. 2000; Petricka et al. 2008). Five of these mutants were named *DEFECTIVELY ORGANISED TRIBUTARIES (DOT)*, with the mutated genes shown to encode a range of proteins (Petricka et al. 2008). One of these mutants, *dot5*, exhibited altered secondary and higher-order vein formation, and showed enhanced sensitivity to the inhibition of polar auxin transport (Petricka et al. 2008). The mutated gene encoded a C2H2 transcription factor, and reintroduction of a functional copy into the *dot5* background fully rescued the phenotype (Petricka et al. 2008). Whilst these results are informative, as yet no *Arabidopsis* mutants have been identified that exhibit higher overall vein density, suggesting that vein patterning is tightly regulated and therefore difficult to perturb.

Much less is known about vascular development in C_3 monocots than in C_3 dicots. By far the best-studied C_3 monocot is rice, from which most insight concerning C_3

monocot vein patterning has been obtained. A number of rice populations have been identified with variation in vein density (Smillie et al. 2012; Feldman et al. 2014), however, only a limited number of mutants with vascular patterning defects have been characterised. The first to be identified was *Osradicleless1,* which exhibits altered vein patterning in leaves, attributed to a delay in procambium formation (Scarpella et al. 2003). Furthermore, *radicleless1* mutants have reduced sensitivity to auxin and increased sensitivity to cytokinin (Scarpella et al. 2003). A conserved role for auxin in C_3 monocot vascular development was further confirmed by analysis of the rice *narrowleaf1 (nal1)* mutant (Qi et al. 2008). Mutant *nal1* leaves form fewer veins than wild-type (WT) (although there is a corresponding decrease in leaf width), and the culms exhibited fewer veins with altered distribution (Qi et al. 2008). Interestingly, the NAL1 protein is expressed in the vasculature, and the *nal1* mutant has reduced polar auxin transport capacity (Qi et al. 2008). These observations support the hypothesis that as in C_3 dicots, auxin is critical in regulating vein patterning in rice.

Although C_3 species can provide information into the broad mechanisms underpinning vascular differentiation, work in C_4 species is essential to reveal the mechanism by which rank-2 intermediate veins are formed. Previous work in maize confirmed a role for auxin in vascular development (as in C_3 species), in that maize seedlings grown on polar auxin transport inhibitors formed hypertrophic vascular bundles (Tsiantis et al. 1999). Unfortunately, however, screens in C_4 species have thus far failed to identify vein patterning mutants (Hall and Langdale 1996; Nelson and Dengler 1997). A mutant with increased interveinal distances relative to wild-type was identified in the C_4 monocot *Panicum maximum*, but it was not characterised in detail and the line has since been lost (Fladung 1994). More recently, two independent large-scale mutant screens using *Sorghum bicolor* (hereafter referred to as sorghum), a C_4 grass closely related to maize, identified two

mutants with reduced vein density (Rizal et al. 2015). However, the pleiotropic effect of both mutations, and the fact that both mapped to a cytochrome P450 gene involved in brassinosteroid biosynthesis, pointed to a defect in general brassinosteroid-regulated development rather than a specific perturbation to procambium initiation.

1.6.2. Regulation of bundle-sheath and mesophyll cell differentiation in C_4 species

Until recently, our understanding of the molecular determinants underpinning BS and M cell differentiation was limited to the Golden2-like (GLK) transcription factors, which regulate dimorphic chloroplast development in C_4 species (Langdale and Kidner 1994; Rossini et al. 2001; Wang et al. 2013a). These were first identified in maize, through analysis of the *bundle sheath defective1 (bsd1)* mutant, which exhibits impaired bundle sheath chloroplast development, and are now known to encode a paralogous gene-pair where one gene is preferentially expressed in the M and one in the BS of maize leaves (Langdale and Kidner 1994; Hall 1998; Rossini et al. 2001). However, the GLK transcription factors do not regulate BS and M cell patterning, because the two cell-types develop at the appropriate time and in the appropriate position within the leaf.

More recently, the combined use of transcriptome analysis, mutant studies and comparison to analogous developmental processes in *Arabidopsis,* has lead to confirmed roles for two maize genes, *ZmSCARECROW1* (*ZmSCR1)* and *ZmSHORTROOT1 (ZmSHR1)*, in BS and M cell patterning in maize (Slewinski et al. 2012; Wang et al. 2013b; Slewinski et al. 2014). In addition to *ZmSCR1* and *ZmSHR1*, a number of other promising candidate Kranz regulators have emerged, which must now be functionally validated.

1.7. Transcriptomics as a tool for identifying regulators of Kranz anatomy

Traditional approaches to gene discovery have offered little insight into the genetic regulators that underpin the development of Kranz anatomy (Langdale 2011). As a result, it seems increasingly unlikely that the previously hypothesised Kranz 'master switch' will be identified (Westhoff and Gowik 2010). Instead, Kranz development likely involves multiple genes and extensive regulatory redundancy, which would explain why forward genetic screens have yielded limited results. Untangling this complexity is a significant challenge, however, recent advances in the field of systems biology have lead to renewed optimism. Over the past few years, there have been a range of systems-based approaches that have provided fresh insight into the genetic regulation that underpins the development of Kranz anatomy (Fouracre et al. 2014). Systems-based approaches are hugely varied, but at the simplest level involve the holistic study of a biological system, rather than the traditional reductionist approach. Whilst they can include a wide range of emerging technologies, the use of transcriptomics has been of particular importance in the context of Kranz anatomy development (Fouracre et al. 2014).

In monocots, leaf development proceeds basipetally, meaning that a developmental progression is represented along the length of a single leaf (Nelson 2011). Transcriptome analysis along the developmental gradient of maize foliar leaves has enabled the transcriptional networks underlying the development of C_4 photosynthesis to be characterised (Li et al. 2010; Pick et al. 2011; Tausta et al. 2014; Wang et al. 2014a). In total, 938 transcription factors are differentially expressed along the leaf, and 180 differ between BS and M cells in the mature leaf (Li et al. 2010). Whilst this analysis provided important insight into C_4 photosynthesis, its use was limited by the lack of a C_3 comparison. More recently, however, this problem has been rectified by a study which integrated transcriptome profiles along

the developmental gradient of both maize and rice leaves, in order to compare how gene expression is regulated between C_4 and C_3 species (Wang et al. 2014a). A number of candidate *cis*-regulatory elements that may drive photosynthetic gene expression, and/or cell-preferential expression, have been identified from this work. Moreover, 118 transcription factors were identified that are differentially expressed in maize BS and M cells (Wang et al. 2014a). In addition, separate BS and M transcriptomes have been compared at three distinct developmental points along the leaf (Tausta et al. 2014). This has provided greater resolution of how the cell-specific accumulation of transcripts changes as leaf development proceeds. Difficulties associated with BS and M cell isolation in rice (and C_3 species more generally) have thus far prevented a comparison with C_3 species, however, these transcription factors are strong candidates for having undergone recruitment from an ancestral C_3 function into a C_4 regulatory role (Wang et al. 2014a). Interrogation of these datasets will continue to provide fresh understanding of how C_4 photosynthesis is regulated.

The extent to which the leaf gradient datasets can be used to elucidate novel regulators of Kranz anatomy is limited by the fact that vein spacing and radial-patterning of BS and M cells was established prior to the earliest developmental stages analysed. Elucidation of true Kranz regulators thus requires an analysis of earlier developmental stages. One recent study undertook both histological and transcriptomic analyses of maize embryonic leaves (Liu et al. 2013). However, leaves of different ages were pooled for transcriptome analysis and thus limited insight was gained into the precise gene expression profiles that underpin Kranz development.

To date, only one study has been specifically designed to capture developmental regulators of Kranz anatomy (Wang et al. 2013b). This study took advantage of the two distinct leaf types in maize: the foliar leaves, which act as the main photosynthetic tissue and exhibit conventional Kranz anatomy, and the husk leaves

which surround the developing ear and exhibit more C_3-like anatomy, with up to 20 M cells separating veins (Langdale et al. 1988). Transcriptome samples were taken at 5 distinct developmental stages in both foliar and husk leaves: Shoot apical meristem plus plastochron-1 (P1) and P2 primordia (FP and HP), P3 and P4 primordia (FP3/4 and HP3/4), P5 primordia (FP5 and HP5), immature leaf (FI and HI) and fully expanded leaf (FE and HE) (Wang et al. 2013b). In both foliar and husk primordia, a similar developmental trajectory takes place between P1 and P4, whereby the mid-vein is initiated (P2), followed by subsequent development of lateral veins (P2-P4). This developmental trajectory diverges between foliar and husk primordia after P4. By P5 in foliar primordia, rank-2 intermediate veins have developed, resulting in the characteristic high vein density associated with Kranz anatomy, and each vascular bundle is separated by only 1 or 2 M cells (Wang et al. 2013b). In contrast, by P5 in husk primordia, cell division and expansion of M cells results in widely-spaced veins that are separated by up to 10 M cells. A comparison of expression at these stages in foliar and husk leaves, therefore, enabled the identification of putative Kranz regulators.

Both putative positive and putative negative Kranz regulators were identified by Wang *et al.* (2013). A positive regulator of Kranz anatomy should be expressed at higher levels in foliar primordia than in husk primordia, and be expressed prior to and/or during P5. In contrast, a negative regulator should be expressed at higher levels in husk primordia than in both foliar primordia and husk expanded leaves (as anatomy is fully established here). Using these criteria (alongside a number of additional conditions), genes were filtered by Wang *et al.* (2013) to identify those whose transcript profiles were consistent with that of a hypothesised positive or negative regulator of Kranz anatomy development. As part of this process, all genes were assigned to clusters based on their expression patterns across the three primordia samples (P, P3/4 and P5) in both foliar and husk leaves. In total, 283

candidate positive regulators, and 142 candidate negative regulators, of Kranz development were identified (Wang et al. 2013b). Of these, 71 were predicted to encode transcription factors. In addition, all genes were allocated to expression clusters based on their transcript profiles in the P, P3/4 and P5 samples. This analysis revealed an additional cohort of transcription factors, all of which were allocated to the 'FA2' expression cluster. This cluster was designed to capture those genes with ascending transcript levels between FP and FP3/4, and maintenance of these levels between FP3/4 and FP5 (Wang et al. 2013b). Manual annotation of this combined pool of transcription factors revealed a subset of high-priority candidates (Fouracre et al. 2014), a number of which support the recruitment of the SCR/SHR pathway into leaves.

1.8. Radial patterning in the Arabidopsis root

To understand how the SCR/ SHR pathway may operate in developing leaves, it is first necessary to understand how it functions in the root, where it was first identified and characterised. In both roots and shoots of *Arabidopsis*, specific cell-types are arranged radially around the vasculature. In the root, the vasculature (stele) is immediately surrounded by a layer of endodermal cells, which is in turn surrounded by a layer of cortical cells (Dolan et al. 1993). Together these cell layers form the ground-tissue. The endodermis and cortex are derived from a cortex endodermis initial cell (CEI), which lies adjacent to the quiescent centre (QC) (Dolan et al. 1993). Two asymmetric cell divisions give rise to the distinct endodermis and cortex layers. First, the CEI divides anticlinally to form a daughter cell and another CEI. Second, the resultant daughter cell divides periclinally to form the progenitors of the endodermis and cortex. The radial patterning of the ground-tissue layer therefore relies on the precise regulation of these asymmetric cell divisions.

Two *Arabidopsis* mutants with defective ground-tissue formation, named *Atscarecrow (Atscr)* and *Atshortroot (Atshr)*, have provided insight into how the endodermis and cortex layers are specified (Benfey et al. 1993; Scheres et al. 1995). Both *Atscr* and *Atshr* form only a single cell layer of ground tissue. In *Atscr* this layer differentiates with features of both the endodermis and the cortex (Laurenzio et al. 1996), whereas in *Atshr* the layer comprises only cortex cells (Benfey et al. 1993). The mutations responsible for the observed phenotypes were subsequently mapped to two genes named *AtSCR* and *AtSHR* (Laurenzio et al. 1996; Helariutta et al. 2000), both of which were among the first members of the plant-specific GRAS transcription factor family to be identified (Pysh and Wysocka‐Diller 1999; Helariutta et al. 2000). Based on the observed phenotypes, both AtSCR and AtSHR are proposed to regulate the asymmetric division of the CEI daughter cell, with AtSHR then promoting endodermal cell identity.

A number of experiments have revealed that the AtSHR protein is mobile, and furthermore, that its movement is fundamental to the radial patterning process. In wild-type roots *AtSHR* transcripts accumulate only in stele cells, however, AtSHR protein moves one cell layer out into either the adjacent CEI cell, CEI daughter cell or endodermis layer (Helariutta et al. 2000; Nakajima et al. 2001; Cui et al. 2007; Gallagher and Benfey 2009). In contrast, *AtSCR* mRNA and protein mainly accumulate adjacent to the stele in the CEI cell, CEI daughter cell and the endodermis layer, and the protein is not mobile (Laurenzio et al. 1996). Protein-protein interaction assays have demonstrated that AtSHR and AtSCR are able to physically interact, and transcription of *AtSCR* is promoted by AtSHR. Based on these results a model of ground-tissue formation was proposed whereby the AtSHR/AtSCR complex promotes the asymmetric division of the CEI daughter cell into the endodermis and cortex progenitors cells (Cui et al. 2007). Subsequently, upon moving into the endodermis from the stele, AtSHR is bound by AtSCR and

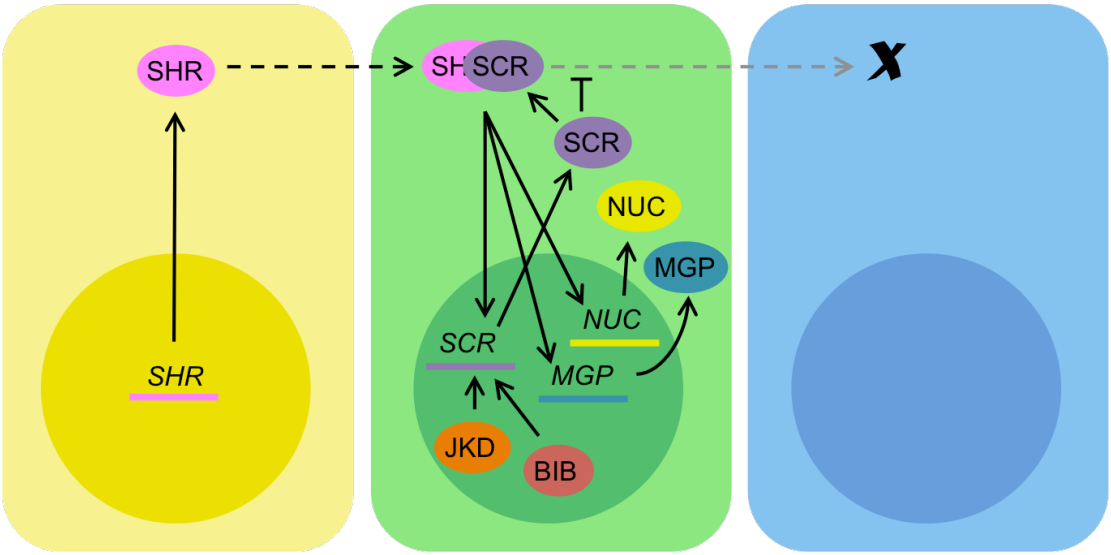


Figure 1.3 Cartoon depiction of the current model of endodermis specification in *Arabidopsis* roots. Yellow represents the stele, green represents the endodermis and blue represents the cortex. Circles represent the nucleus. Genes are represented by straight lines and italicised gene names above. Coloured ovals represent proteins. Solid black arrows within the nucleus or entering the nucleus indicate the promotion of gene expression. Solid black arrows leaving the nucleus indicate transcription and subsequent translation of a protein. Black dashed arrows represent protein movement, grey dashed arrows indicate inhibited protein movement. Lines with flat ends represent the inhibition of protein movement.

sequestered into the nucleus, thus preventing its movement more than one cell layer away from its site of translation (Levesque et al. 2006; Cui et al. 2007). Once in the nucleus, AtSHR further promotes *AtSCR* expression, providing a positive feedback loop that ensures there is sufficient AtSCR to prevent further movement of any AtSHR (Cui et al. 2007). In this model AtSHR is both necessary and sufficient to specify endodermal cell identity, and by limiting AtSHR movement, AtSCR prevents the formation of additional endodermis cell-layers (**Figure 1.3**) (Cui et al. 2007).

More recently, this model has been refined to take into account new experimental evidence. One of the primary pieces of evidence for the SCR/SHR model in its original form, came from the fact that in *AtSCR* RNAi lines AtSHR movement was no longer restricted to the endodermis, and increases in the AtSHR domain lead to the formation of extra endodermal cell layers (Cui et al. 2007). These cells were identified as endodermal based on *AtSCR* expression and suberin deposition. However, recent work has questioned whether suberin is always an appropriate marker of the endodermis (Naseer et al. 2012). To further investigate the identity of the extra-cell layers formed Wu *et al.* (2014) used PIN2, which is absent from the endodermis, but accumulates on the apical side of the cortex, as a marker of cortical cell identity. PIN2 accumulated on the apical side of the SHR induced extra cell layers in the *AtSCR* RNAi lines, indicating that they were cortex rather than endodermis (Wu et al. 2014). This finding was supported by a propidium iodide (PI) assay (in which PI is blocked from entering the endodermis by the lignified casparian strip), in that only the cell layer immediately surrounding the stele prevented PI accumulation (Wu et al. 2014). These findings have allowed a new model to be proposed whereby AtSHR is necessary, but not sufficient, to specify the endodermis. Instead, an as yet unidentified stele-derived signal is proposed to trigger endodermal cell-identity in the presence of AtSHR.

A number of additional transcription factors have been characterised that play a role in regulating the interaction between AtSHR and AtSCR. These include a number of C2H2 transcription factors, including *JACKDAW*, *MAGPIE* and *BALDIBIS* (see section 1.9), and an additional GRAS transcription factor named SCARECROW-LIKE23 (AtSCL23) (Welch et al. 2007; Long et al. 2015a, 2015b). AtSCL23 is the closest *Arabidopsis* homolog of AtSCR, and was previously found by yeast two-hybrid to interact with AtSHR (Lee et al. 2008). Single *Atscl23* mutants show no discernible difference to WT (Lee et al. 2008; Cui et al. 2014). However, the *Atscr Atscl23* double mutant possesses a single layer of ground-tissue that, unlike *Atscr,* lacks casparian strip markers, and thus does not appear to have any endodermal cell identity (Long et al. 2015a). Therefore, the available evidence points toward AtSCL23 acting partially redundantly with AtSCR in order to restrict AtSHR movement and promote correct endodermal specification.

Although identified and characterised in a dicot, a number of pieces of evidence point to the SCR/SHR pathway being conserved in monocots. Orthologs of both AtSHR and AtSCR were identified in rice and maize, with similar ground-tissue expression patterns and abilities to bind to each other (Lim et al. 2000; Kamiya et al. 2003; Cui et al. 2007; Ni et al. 2014). Furthermore, as well as interacting with OsSCR1, OsSHR1 was found to bind AtSCR in yeast (Cui et al. 2007), and expression of the maize SCR ortholog (*ZmSCR1*) in the *Arabidopsis Atscr* background was found to complement the impaired ground-tissue formation phenotype (Lim et al. 2005). This evidence suggests that there is some degree of functional conservation between the SCR/SHR pathway in monocots and dicots. However, it is known that some monocots, such as rice and *Brachypodium distachyon* (hereafter referred to as *Brachypodium*), form multiple cortex layers rather than the single layer formed in *Arabidopsis* (Clark and Harris 1981; Watt et al. 2009). The revised SCR/SHR model provides a mechanism for how these extra layers of cortex are specified, and

experiments in rice are supportive of this model (Kim Gallagher, personal communication). Expression of rice and *Brachypodium* SHR homologs in the *Arabidopsis Atshr* background enabled a partial rescue of the mutant phenotype: root growth was almost fully restored to WT levels, and meristem size was restored to an intermediate size between *Atshr* and WT (Wu et al. 2014). In addition, the endodermal cell layer (absent in *Atshr*) was restored, indicating that the monocot SHR homologs are capable of specifying endodermal cell identity in *Arabidopsis*. However, in contrast to WT *Arabidopsis*, these lines formed multiple cortical cell layers, reminiscent of both the roots of WT rice and *Brachypodium,* and also those of the *Arabidopsis AtSCR* RNAi lines where the domain of AtSHR movement was extended beyond the endodermis (Cui et al. 2007; Wu et al. 2014). It was subsequently demonstrated that when expressed in *Arabidopsis,* the monocot SHR proteins were able to move more than one cell layer from the stele, in marked contrast to AtSHR which normally moves only one cell-layer out (Wu et al. 2014). These results therefore support a model whereby the extent of SHR movement regulates the number of cortex cell layers formed, although this has yet to be tested in a monocot species, and it is not yet understood how the monocot SHR homologs are able to move further than AtSHR (and whether in fact they do so in their endogenous context). In summary, there appears to be a high degree of functional conservation regarding the SCR/SHR pathway between monocots and dicots.

It is now known that AtSCR and AtSHR function is not restricted to the root. For example, it has been demonstrated that reduced root growth observed in *Atscr* and *Atshr* plants is due to defective specification and maintenance of the stem cells which make up the quiescent centre (QC) (Sabatini et al. 2003). Furthermore a role for AtSCR and AtSHR in leaf growth has been established, with *Atscr* and *Atshr* mutants exhibiting a lower cell division rate in the leaf (Dhondt et al. 2010). This lead to the hypothesis that AtSCR and AtSHR act as general regulators of cell proliferation in

leaves, in contrast to the more specific role in ground-tissue formation and patterning in roots. More recently, however, a patterning role for both AtSCR and AtSHR in the shoot has been inferred due to impaired starch sheath and BS cell formation in *Atscr* and *Atshr* hypocotyls and leaves, respectively (Cui et al. 2014; Yoon et al. 2016) (see section 1.10). Interestingly, AtSCR and AtSHR roles are not always linked, in that AtSCR has an AtSHR-independent role in regulating the response to sugar (Cui et al. 2012). Similarly, AtSHR has an AtSCR-independent function, in that mutations in AtSHR rescue the cell death phenotype associated with *catalase2-2* deficiencies (Waszczak et al. 2016). *cat2-2* mutants display enhanced photorespiratory stress because the H_2O_2 produced during photorespiration is not broken down. Given that H_2O_2 is a component of stress signalling, the interaction with AtSHR points toward a point of integration between stress and developmental responses (Waszczak et al. 2016).

1.9. The role of INDETERMINATE DOMAIN (IDD) transcription factors in plant development

Zinc finger (ZnF) transcription factors have well elucidated roles in a diverse range of cellular processes across eukaryotes. A number of different ZnF motifs exist, varying in terms of the number of Cys (C) and His (H) residues that bind the zinc ion. Of these, the C2H2 ZnF motif was the first to be characterised and is the most common motif found in eukaryotic transcription factors (Miller et al. 1985). The full complement of genes encoding C2H2 transcription factors has been identified in *Arabidopsis thaliana,* rice and maize (Englbrecht et al. 2004; Agarwal et al. 2007; Ciftci-Yilmaz and Mittler 2008; Wei et al. 2015), revealing a number that are plant-specific. Functional characterisation of plant C2H2 transcription factors has indicated key roles in development, metabolism and responses to abiotic and biotic stress (Ciftci-Yilmaz and Mittler 2008).

C2H2 transcription factors from the INDETERMINATE DOMAIN (IDD) subfamily have emerged as important regulators of plant development. The IDD genes were originally identified based on their homology with the maize *INDETERMINATE1* (ID1) gene, which regulates the floral transition (Kozaki et al. 2004; Colasanti et al. 2006). There are similar numbers of IDD proteins encoded in the *Arabidopsis,* rice and maize genomes (16, 15 and 17 respectively), with monocot and dicot specific clades (Colasanti et al. 2006; Yi et al. 2015). Characterisation of a number of IDD mutants has provided insight into gene function in *Arabidopsis*. These roles, which reflect the diversity of plant C2H2 transcription factors more broadly, include gravitropism (Morita et al. 2006; Tanimoto et al. 2008; Wu et al. 2013), sugar metabolism (Seo et al. 2011), auxin biosynthesis and transport (Cui et al. 2013), and germination (Feurtado et al. 2011).

Interestingly, a number of *Arabidopsis* IDD proteins have well-established roles in cellular patterning, particularly in the context of root development and adaxial/abaxial cell fate in leaves (Welch et al. 2007; Reinhart et al. 2013). In the root, four IDD genes, *JACKDAW (AtJKD), BALDIBIS (AtBIB), MAGPIE (AtMGP)* and *NUTCRACKER (AtNUC)*, encode proteins that play a role in spatially regulating the SCR/SHR pathway (Welch et al. 2007; Ogasawara et al. 2011). *AtJKD* was initially identified in a mutant screen in which ~15,000 T-DNA insertion lines were screened for mutations affecting ground-tissue differentiation in the root (Welch et al. 2007). One mutant, named *Atjkd-1*, exhibited ectopic periclinal divisions in the cortex, with the responsible insertion mapped 3.5kb upstream of the *AtJKD* transcription start site. Subsequent expression analysis, the identification of independent *AtJKD* insertion lines, and construction of an RNAi line confirmed a role for JKD in ground tissue formation. AtBIB was identified as a potential redundant partner of AtJKD due to its similar expression profile in the ground-tissue and close homology (Long et al. 2015b). RNAi lines (*Atbib-i*) exhibited similar radial patterning to WT, however, *Atjkd;*

Atbib-i lines showed a more severe phenotype than *Atjkd* single mutants, with increased cell layers and cells per layer (Long et al. 2015b). AtBIB and AtJKD were subsequently shown to physically interact with both AtSCR and AtSHR, and to be capable of promoting AtSHR sequestration into the nucleus (Welch et al. 2007; Long et al. 2015b). As such, AtJKD and AtBIB are proposed to act redundantly alongside AtSCR in order to restrict AtSHR movement.

Similarly to *AtBIB, AtMGP* and *AtNUC* were identified based on their homology with *AtJKD* and their known ground-tissue expression profiles (Welch et al. 2007; Long et al. 2015b). Also like *Atbib-i,* the single *Atmgp* mutant and *Atnuc-i* RNAi line both had wild-type phenotypes. Interestingly, however, *Atmgp-1* complements the *Atjkd-1* ectopic cortex phenotype resulting in a return to wild-type appearance, indicating that *AtMGP* promotes cell division in the ground-tissue (Welch et al. 2007). The *Atjkd; Atmgp; Atnuc* triple mutant and *Atjkd; Atmgp; Atnuc; Atbib* quadruple mutant both exhibited patches of undivided ground-tissue, which retained casparian strips and thus endodermal cell identity (Long et al. 2015b). In addition, both *AtMGP* and *AtNUC* were previously identified as transcriptional targets of both AtSCR and AtSHR (Levesque et al. 2006; Cui et al. 2007), and protein-protein interaction assays have confirmed that AtMGP can physically interact with AtSCR and AtSHR (Welch et al. 2007). Taken together, these observations indicate that at least four IDD genes operate redundantly in order to regulate the SCR/SHR pathway, by both acting as transcriptional activators of *AtSCR* and regulators of AtSHR movement (Figure 1.3).

The roles of IDD genes in maize are less clear than in *Arabidopsis*. Despite the *ID1* gene being the founding member of the family, only one additional maize IDD mutant has been reported (Yi et al. 2015). The *NAKED ENDOSPERM (Zmnkd)* mutant was identified based on a distinct shrivelled kernel phenotype, and was subsequently shown to be caused by transposon insertions in the maize-specific duplicate genes *ZmNKD1* and *ZmNKD2* (Yi et al. 2015). Histological analysis revealed that *Zmnkd1;*

Zmnkd2 kernels had impaired aleurone cell-layer formation; in WT kernels, the aleurone forms a single cell layer surrounding the endosperm, however, in *Zmnkd1; Zmnkd2* kernels multiple ectopic aleurone cell-layers were observed. These observations indicate a role for *ZmNKD1* and *ZmNKD2* in cell patterning, the first described example of IDD gene involvement in cellular patterning in maize. It is, as yet, unclear whether other maize IDD genes may function analogously to their *Arabidopsis* orthologs in root and leaf cellular patterning.

1.10. A proposed model for the development of Kranz anatomy

It has previously been hypothesised that the formation of Kranz anatomy may be analogous to radial patterning in the root ground-tissue (Esau 1943; Nelson 2011). It is easy to see how this hypothesis arose: Kranz anatomy exhibits a single (often suberised) cell layer (BS) surrounding the vasculature, encircled by a distinct second cell layer (M). In the root, a single layer of suberised endodermal cells surround the vasculature (stele), encircled by a distinct layer of cortex cells. Furthermore, the BS and endodermis have long been considered to be analogous, despite arising from different stem cell populations and at distinct developmental points (Esau 1965). If correct, these observations raise the intriguing possibility that regulatory mechanisms underpinning radial patterning of the root may also underpin the development of Kranz anatomy.

Recent studies have suggested that the SCR/SHR pathway is active in the leaves of both C_3 and C_4 species. Notably, the leaves of *Atscr, Atshr and Atscl23 Arabidopsis* mutants exhibit impaired BS formation (Cui et al. 2014), suggesting involvement of these genes in BS cell formation in C_3 plants. The situation is less clear in monocots. OsSCR transcripts were not detected in the vasculature or BS cells of young rice leaves, however, only one of two OsSCR homologs was assessed (Kamiya et al.

2003; Cui et al. 2007). In contrast, cell-specific RNA-seq analysis in maize indicated that both *ZmSCR1* and *ZmSHR1* are preferentially expressed in BS cells (Wysocka-Diller et al. 2000; Li et al. 2010; Pick et al. 2011; Chang et al. 2012). In addition, *ZmSCR1*, *ZmSHR1* and *ZmSHR2* emerged from maize foliar and husk RNA-seq comparisons as high priority candidate regulators of Kranz anatomy in maize (Fouracre et al. 2014; Wang et al. 2013b).

Functional evidence has recently emerged that validates the proposed Kranz patterning roles of *ZmSCR1* and *ZmSHR1*, in that transposon insertions in both genes resulted in phenotypic perturbations to BS and M cell patterning (Slewinski et al. 2012, 2014). In the *Zmscr1* line, both extra and starchless BS cells were formed (Slewinski et al. 2012), whereas in the *Zmshr1* line extra M cells formed between veins (Slewinski et al. 2014). In addition, a number of vascular patterning defects were reported, including minor alterations to vein density and the number of intermediate veins separating laterals. Unfortunately, in the case of *Zmscr1* the reported phenotype was extremely subtle, and whilst quantification provides some validation, the majority of vascular bundles appeared normal. In the case of *Zmshr1*, the phenotype was complicated by the presence of additional transposon insertions in the line, and no quantification of the reported defects was provided. Thus, despite evidence pointing to recruitment of the SCR/SHR pathway into leaves for Kranz patterning, there are still a number of questions that remain unanswered.

In addition to *ZmSCR1, ZmSHR1 and ZmSHR2*, nine additional high priority candidate Kranz regulators were identified from the Wang et al. 2013 dataset (Fouracre et al. 2014) (Figure 1.4). These nine include two C2H2 IDD transcription factors that are expressed early in foliar leaf development, *ZmNKD1* and *ZmRVN1*. *ZmRVN1* is expressed preferentially in the BS, whereas *ZmNKD1* is expressed preferentially in the M (Li et al. 2010; Chang et al. 2012). *ZmNKD1* and *ZmRVN1* are strong candidates for the cell-specific regulators that were proposed in an early

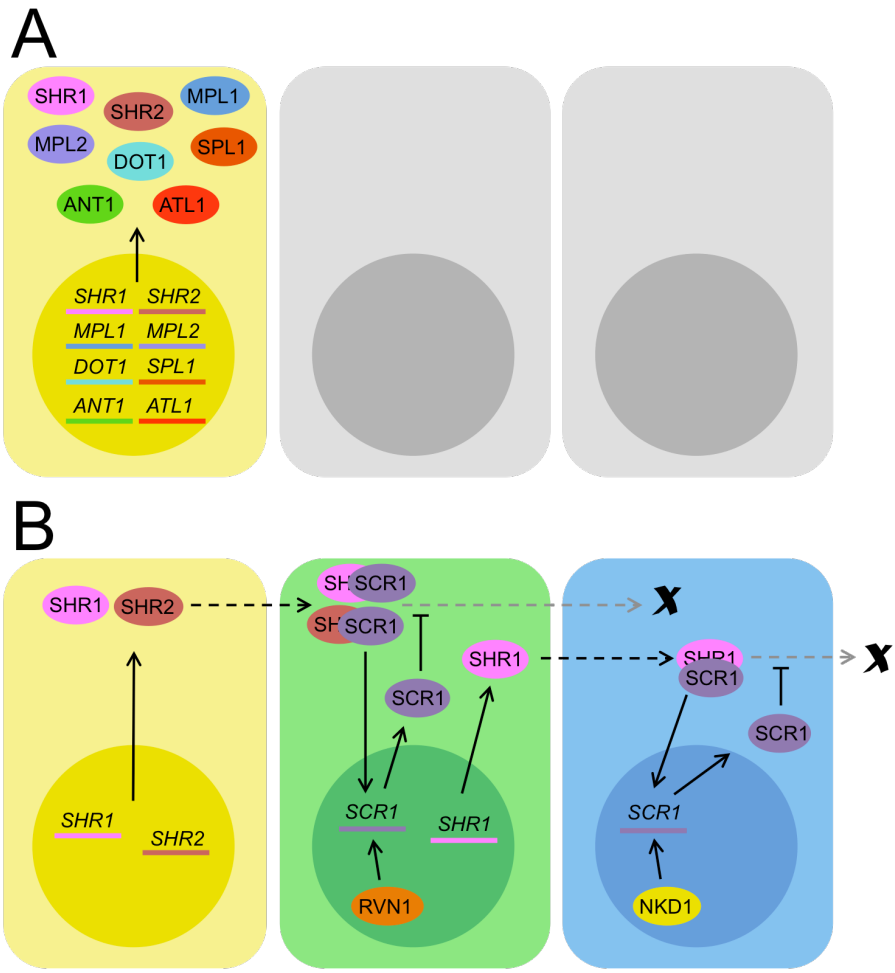


Figure 1.4 Cartoon depictions of Kranz development model in maize as proposed by Fouracre *et al.* (2014). A) Transcription factors hypothesised to play a role in procambium specification and vascular differentiation. **B)** Proposed model of Kranz BS and M patterning based on the *Arabidopsis* SCR/SHR pathway. Yellow represents the vein, grey represents unspecified ground meristem, green represents BS and blue represents C_4 M. Circles represent the nucleus. Genes are represented by straight lines and italicised gene names above. Coloured ovals represent proteins. Solid black arrows within the nucleus or entering the nucleus indicate the promotion of gene expression. Solid black arrows leaving the nucleus indicate transcription and subsequent translation of a protein. Black dashed arrows represent protein movement, grey dashed arrows indicate inhibited protein movement. Lines with flat ends represent the inhibition of protein movement.

model of Kranz development, in which a signal moves out of the vein before interacting with cell specific transcription factors to promote BS and M differentiation (Langdale and Nelson 1991).

1.11. Model monocot species for studying Kranz anatomy development

Given the increasing emergence of candidate genes from the analysis of transcriptome datasets, there is a renewed focus on establishing high-throughput systems in order to provide functional characterisation of these genes. In the context of C_4 research, this has been an on-going and often problematic quest. Maize has several advantages when it comes to C_4 study. There are extensive genetic resources and mutant libraries categorised, it is transformable, phylogenetically close to rice, and much of the previous work on C_4 has been carried out in maize. However, its large stature, large and highly duplicated genome, and long generation time mean that it is not an ideal species for high throughput characterisation of candidate genes, particularly when plants have to be grown in a greenhouse. Furthermore, whilst transformation is possible, it is often prohibitively expensive and logistically challenging. Other species such as *Setaria viridis* and *Setaria italica* have recently been proposed as potential alternatives to maize as model C_4 grasses. They have many potential advantages, including small, sequenced genomes, small stature and short generation times (Brutnell et al. 2010; Li and Brutnell 2011). Moreover, a number of successful *Agrobacterium tumefaciens* transformation protocols have recently been reported (Brutnell et al. 2010; Martins et al. 2015). Unfortunately, these have not been easy to replicate in Oxford (Daniela Vlad, personal communication), and therefore, despite the inherent disadvantages, maize currently remains the best resource for obtaining loss-of-function data for potential Kranz regulators.

1.12. The contribution of homeolog gene-pairs to maize leaf development

One factor that must be considered when utilising maize for functional validation of candidate genes, is the complexity of the maize genome. Polyploidy is often caused by whole genome duplication (WGD) or even triplication events, and is extremely common in plants, particularly in the angiosperms where at least 18 WGD events have been identified (Blanc and Wolfe 2004b; Adams and Wendel 2005). In the monocots, a WGD occurred at the base of the clade, following by two subsequent duplications at the base of the grasses (Paterson et al. 2004; Tang et al. 2010; D'Hont et al. 2012; Jiao et al. 2014; McKain et al. 2016). These grass-specific WGD events have been hypothesised to act as a pre-adaptation to the evolution of C_4, by enabling functional diversification (Monson 2003; Wang et al. 2009). Whilst these WGD events preceded the divergence of grasses, maize underwent a far more recent whole genome duplication event between 5 and 12 million years ago (Swigonova et al. 2004a). This WGD is not shared by its close relative sorghum, which shares a C_4 origin. Since the maize duplication, rather than remain tetraploid, multiple chromosomal breakage and fusion events have combined the duplicated chromosomes into a diploid genome containing 10 novel mosaic chromosomes (Wei et al. 2007; Schnable et al. 2011). In contrast, the sorghum genome has retained the ancestral non-duplicated state (Swigonova et al. 2004a). Analysis of the maize genome has identified 3228 gene-pairs that have been retained post-duplication, adding a layer of complexity to the maize genome that is not shared by other grasses (Schnable et al. 2011). Given that maize has ~32,000 genes (Schnable et al. 2009), homeolog gene-pairs thus account for over 20% of gene content. Despite constituting a large fraction of the maize genome, it is not yet known to what extent homeolog gene-pairs have functionally diverged since the whole genome duplication event. This has particular relevance in the context of testing candidate regulators of

Kranz anatomy, as many of the high priority candidate genes identified by Wang *et al.* 2013 have homeolog partners.

1.13. Thesis aims

Understanding how the development of Kranz anatomy is regulated will provide insight into one of the most remarkable examples of convergent evolution. Furthermore, this understanding will be of critical importance if attempts to engineer C_4 rice are to succeed. Recent RNA-seq analyses have revealed a number of promising candidate genes that may play a role in Kranz patterning. Generating loss-of-function data in C_4 species, and gain-of-function data in C_3 species is now key to determining whether these candidate genes play a role in Kranz development. Given the current necessity of utilising maize as a C_4 model species, it is important to consider the impact of the recent maize whole genome duplication, both broadly in terms of maize genome evolution, and also in relation to candidate Kranz regulators. The work in this thesis, therefore, first aimed to analyse the extent to which homeolog gene-pairs exhibited distinct gene expression profiles in early leaf development, both on a whole genome scale and with specific reference to candidate Kranz regulators (Chapter 3). The next aim was to test the hypothesis that a number of candidate genes function as Kranz regulators, specifically two GRAS transcription factors, *ZmSCR1* and *ZmSHR2* (Chapter 4) and two C2H2 transcription factors, *ZmNKD1* and *ZmRVN1* (Chapter 5). Finally, the broader implications of this work in relation to untangling the complexity of Kranz regulation, and C_4 rice prospects, were considered (Chapter 6).

Chapter 2: Materials and methods

2.1. Plant material and growth conditions

2.1.1. Rice

Rice (*Oryza sativa* L.) cv. Kitaake seed were used in all rice experiments. Relative to other rice cultivars, Kitaake has a faster generation time (producing seed 11-12 weeks after planting) and grows well in greenhouse conditions (Kim et al. 2013). Rice was grown in either 6cm or 8.5cm square pots containing water saturated John Innes compost number 3 (www.johninnes.info), in glasshouses supplemented (when necessary) with 100µmol photon m^{-2} s^{-1} and a 16hr/8hr light regime. Young seedling plants were covered with plastic bottles or contained within a clear autoclave bag to increase humidity. Daytime and night-time temperatures were maintained at 28°C and 20°C respectively.

2.1.2. Maize

Maize (*Zea mays* L.) inbred line B73 was used as a standard WT line for molecular experiments. Both B73 and W22 inbred lines were used for phenotypic analysis. Maize was either grown in the greenhouse in Oxford or in the field in Iowa, USA. Plants were grown in the greenhouse with 16hr/8hr light regime. Daytime temperature was 28°C and night-time temperature was 20°C. Supplemental light was provided when natural light dropped below 120µmol photon m^{-2} s^{-1}. Seed were germinated in warm, damp vermiculite before being transferred to 12cm diameter pots containing John Innes compost number 3 mixed with vermiculite. Plants were maintained in these pots on raised benches for 3-4 weeks, with tray watering every 2-3 days. Plants were then transferred into 28cm diameter pots containing the same soil mix supplemented with Osmocote slow release fertiliser (www.osmocote.co.za), and moved to lower benches for growth to full maturity and pollination. In the field,

seed were planted in May, pollinated in late July/ early August and harvested in late September.

2.2. Molecular cloning

2.2.1. Primer design

Primers for cloning, genotyping and expression analysis were designed using Primer3Plus (Untergasser et al. 2012). Most primers were designed with Tm 57-63°C, and length 18-24nt. Suggested primers were checked for self- and hetero-dimerisation using OligoAnalyzer® (IDT, www.idtdna.com/Scitools), and where necessary BLAST searched using Phytozome to confirm that the primer would not bind to off-target sequences. A full list of primers and their use in this study can be found in Appendix 1.

2.2.2. Gateway cloning

Constructs for constitutive expression in rice were assembled using the Gateway cloning system (ThermoFisher, www.thermofisher.com). Maize transcripts were amplified from leaf primordia (B73) enriched cDNA using primers flanked by the Gateway recombination sequences. Transcript amplicons were gel extracted using a kit according to the manufacturers instructions (Sigma-Aldrich, www.sigmaaldrich.com), and cloned into the entry vector pDONR207 using the Gateway enzyme BP Clonase® (ThermoFisher, www.thermofisher.com). For constitutive expression in rice using the maize ubiquitin promoter, the insert was recombined from pDONR207 into the destination pVEC8 vector using the Gateway enzyme LR Clonase® (ThermoFisher, www.thermofisher.com) (Wang et al. 1998). For preferential expression in the M or BS of rice, modified Gateway expression vectors were used with either the maize PEPC promoter (M) or *Zoysia japonica* phosphoenolpyruvate carboxykinase (PCK) promoter (BS) used to drive transgene expression (Matsuoka et al. 1994; Nomura et al. 2005).

2.2.3. Golden gate cloning

Constructs for inducible expression or for expression driven by maize endogenous promoters were cloned using the Golden Gate cloning system (Weber et al. 2011). Sequences to be cloned were first checked for four different type-IIS restriction enzyme sites (*Bpi*I, *Bsa*I, *Esp*3I and *Dra*III). If present in a coding sequence, the site was modified to remove the restriction site with reference to the rice codon use table, whilst ensuring the same protein sequence was encoded. If present in a sequence not to be translated (e.g. a promoter), the site was altered with reference to maize duplicate genes (if present), and/or the sorghum ortholog. The final sequences were then synthesised (GenScript, www.genscript.com) and inserted into a level-0 vector. Individual transcription units (promoter-gene-terminator) were assembled into a level-1 module in a ligation reaction with *Bsa*I. Final multi-gene level-2 constructs, consisting of two or more level-1 modules, were assembled using *Bpi*I. Both level-1 and level-2 ligation reactions were carried out by adding the following components to a 50µl micro-centrifuge tube: 100ng of donor modules, 200ng of acceptor plasmid, 2µl of ATP (10mM), 2µl of either NEB (level-1 ligation) or ThermoFisher (level-2 ligation) buffer, 400 units of T4 DNA ligase and 10 units of either *BsaI* (level-1 ligation) or *Bpi*I (level-2 ligation). Tube contents were collected by brief micro-centrifugation and incubated as follows: 3 cycles of 10 minutes 37°C, 10 minutes 16°C followed by 10 minutes 37°C and 20 minutes 65°C.

2.2.4. Restriction digests

Level-1 and -2 Golden Gate constructs were initially checked by restriction digest. A range of enzymes were utilised, depending on the sequence of the construct being validated, however, *Hin*cII (New England Biolabs, www.neb.com) proved to be useful for checking most constructs. 25µl reactions were carried out with 10 units of enzyme, 2.5µl of 10x NEB buffer and 0.5µg of plasmid DNA. Reactions were

incubated at 37°C for 3 hours before being visualised by gel electrophoresis (see section 2.5.4).

2.2.5. DNA sequencing

Insertions into Gateway vectors were sequenced using BigDye® Terminator v3.1 Cycle Sequencing kit, according to the manufacturers instructions (Applied Biosystems, www.appliedbiosystems.com). Sequencing reactions were precipitated by adding 2.5 volumes of 100% v/v ethanol and centrifuging at 4600rpm for 40 minutes. Supernatant was removed and 40µl of 70% v/v ethanol added, before centrifuging at 4600rpm for a further 15 minutes. The resultant DNA pellet was air dried in the dark for at least 1 hour, and then sequenced by the Department of Zoology Sequencing Service (Department of Zoology, University of Oxford, www.zoo.ox.ac.uk) using an ABI prism 3100 capillary sequencer (Applied Biosystems, www.appliedbiosystems.com). Final Golden Gate level-2 constructs were sent for Illumina sequencing by IMGM (www.imgm.com).

2.2.6. Bacterial transformation, growth and plasmid preparation

For replication, plasmids were transformed by heat-shock into the *Escherichia coli* strains DH5α or DH10β. DH5α competent cells were prepared according to the protocol of Inoue *et al.* (1990) by Zara Lewis, whereas DH10β competent cells were acquired from New England Biolabs (www.neb.com). For heat-shock transformation, competent cells (50µl for DH5α, 8µl for DH10β) were thawed on ice before 1-5µl of binary vector was added and left on ice for 20 minutes. Cell and plasmid mixtures were then heated to 42°C for 50s, before being returned to ice for at least 1 minute. Cells were then added to 250µl Lysogeny Broth (LB) media and incubated at 37°C for 45 minutes with gentle shaking, before being spread on LB agar media with the appropriate antibiotic selective marker for growth overnight. Single colonies were used to inoculate 5ml of LB media containing the appropriate antibiotic marker (either

Gentinomycin (12.5 µg/ml), Spectinomycin (100 µg/ml), Kanomycin (50 µg/ml) or Ampicillin (50 µg/ml)), with inoculated cultures grown for a further 24-hours at 37°C with shaking (200rpm). Glycerol stocks were prepared from liquid cultures by slowly adding 500µl of culture into 200µl of 100% v/v glycerol. Glycerol stocks were stored at -80°C. Plasmid DNA was prepared using a QIAprep Spin Miniprep Kit according to the manufacturers instructions (Qiagen, www.qiagen.com).

For rice transformation, plasmids were transformed into the *Agrobacterium tumefaciens* strain Eha105 by electroporation. Eha015 competent cells were prepared by Zara Lewis using the same protocol described previously (Inoue et al. 1990). Competent cells (50µl) were mixed with binary vector as described for heat-shock transformation previously, before being added to an electroporation cuvette. A Gene Pulser II (Bio-Rad, www.bio-rad.com) was used to apply an electric pulse of 1.8kV to cells (capacitance- 25µF, resistance- 600Ω). 1ml LB media was immediately added and the mixture left for 3 hours with gentle shaking at 28°C. Cells were spread on LB agar media containing Rifampicin (50 µg/ml) and the appropriate antibiotic selective marker (see previous concentrations). Bacteria were grown for 2-3 days at 28°C before single colonies were used to inoculate a 5ml LB media culture containing Rifampicin (50 µg/ml) and the appropriate antibiotic selective marker (see previous concentrations). Cultures were grown for a further 2 days at 28°C with shaking (200rpm). Glycerol stocks were prepared as described previously.

2.3. Plant transformation and tissue culture

2.3.1. Media

During the course of plant transformation and tissue culture a range of media were used. AAM media was used for *Agrobacterium* suspension prior to transformation, R1 plates were used to initiate callus formation from de-husked rice seed, R2 plates were used for co-cultivation of callus and *Agrobacterium* following transformation, R3

plates were used to select hygromycin resistant callus following transformation, R4 plates were used to regenerate plants from transformed callus and 50ml falcon tubes containing R5 media were used to encourage rooting of young plants. Media recipes were as follows:

- **AAM:** 4.3g/L MS salts and vitamins, 68.5g/L sucrose, 0.5g/L MES, 36g/L glucose, 500mg/L casamino acid, 0.876g/L L-glutamine, 0.266g/L L-aspartic acid, 0.174g/L L-arginine, 0.75mg/L glycine. pH 5.8 with KOH.

- **R1:** 4.3g/L MS salts and vitamins, 30g/L Sucrose, 0.5g/L MES, 300mg/L Casamino acid, 2.8g/L L-Proline, 2mg/L 2,4-D, 4g/L Phytagel. pH 5.8 with KOH.

- **R2:** 4.3g/L MS salts and vitamins, 30g/L Sucrose, 10g/L Glucose, 0.5g/L MES, 300mg/L Casamino acid, 2mg/L 2,4-D, 4g/L Phytagel. pH 5.2 with KOH. 1ml 20mg/ml acetosyringone after autoclaving.

- **R3:** 4.3g/L MS salts and vitamins, 30g/L Sucrose, 0.5g/L MES, 300mg/L Casamino acid, 2.8g/L L-Proline, 2mg/L 2,4-D, 4g/L Phytagel. pH 5.8 with KOH. Add 200mg/L Timentin and 30mg/L Hygromycin after autoclaving.

- **R4:** 4.3g/L MS salts and vitamins, 30g/L Sucrose, 0.5g/L MES, 2g/L Casamino acid, 30g/L Sorbitol, 2mg/L Kinetin, 1mg/L NAA, 4g/L Phytagel. pH 5.8 with KOH. Add 200mg/L Timentin and 20mg/L Hygromycin after autoclaving.

- **R5:** 2.15/L MS salts and vitamins, 15g/L Sucrose, 0.5g/L MES, 2g/L Phytagel. pH 5.8 with KOH. Add 200mg/L Timentin and 20mg/L Hygromycin after autoclaving.

2.3.2. Rice callus induction

Rice seed were de-husked and then sterilised for 2 min in 70% v/v ethanol, followed by 15 min in 25% v/v bleach. Sterilised seed were washed at least five times with sterile water before being placed gently onto the surface of an R1 plate (10-15 seed per plate). Plates were cultured for 2-3 weeks in a growth chamber set to 32°C with 24-hour light (100-150µmol photon m^{-2} s^{-1}). Secondary calli that had detached from the seed were then transferred to a fresh R1 plate and grown for another 2 weeks. At this stage callus was ready for transformation; however, callus was routinely sub-cultured once more to a fresh R1 plate, and grown for 1 more week prior to transformation.

2.3.3. *Agrobacterium* mediated transformation of rice

A. tumefaciens was either grown from a glycerol stock on LB agar plates for 2 days, with 4 loopfuls of bacteria harvested for transformation, or a 10ml LB culture was started directly from a glycerol stock and grown for 1 day before 4ml was spun down and the pellet used for transformation. In either case, cells were re-suspended in 4ml AAM media and 8µl acetosyringone (20mg/ml), and incubated at room temperature for 45 min. Rice calli were then added to a final volume of ~6ml, with this mix inverted before incubation for 10 min at room temperature. Excess media was then removed and calli blotted dry on sterile filter paper, before being transferred to an R2 plate and cultured in the dark at 25°C for 3 days.

2.3.4. Callus selection and regeneration

Following co-cultivation, calli were transferred to R3 plates containing timentin (200 µg/ml) (to prevent *A. tumefaciens* overgrowth) and hygromycin (30 µg/ml) (to select for transformed calli). Calli were cultured in the same growth chamber used for callus induction for 4 weeks total, with living calli transferred to fresh plates after 2 weeks. Following selection, calli were transferred onto R4 plates containing kinetin

and the synthetic auxin NAA, both of which promote plant regeneration. Calli were transferred to fresh R4 plates every 2 weeks.

2.3.5. Plant rooting and transfer to soil

Plants regenerating on R4 plates were transferred to R5 rooting media in 50ml closed plastic tubes, and grown in a growth chamber set at 28°C with 16hr/8hr light regime (100-150µmol photon m^{-2} s^{-1}). After 1-2 weeks plants, producing hygromycin resistant roots were removed and washed, and transferred into open tubes containing only water. Plants were grown for a further 1-2 weeks in the same growth chamber, before 3-4 plants arising from each independent callus (with the potential to have originated clonally rather than as independent transformation events) were transplanted into soil in the greenhouse.

2.4. Maize mutant identification

A list of high priority candidate genes were searched for available transposon insertions using MaizeGDB (Lawrence et al. 2008). Lines containing desired insertions were ordered from the UniformMu, IlluminaMu and Ac/Ds projects (Settles et al. 2007; Vollbrecht et al. 2010; Williams-Carrier et al. 2010). The majority of insertions were predicted to reside within exons, however, four lines were identified with insertions residing in the 5' UTR (Appendix 2).

2.5. DNA extraction and plant genotyping

2.5.1. CTAB DNA extraction

Where higher-quality DNA was desired, or the number of samples for extraction was low, genomic DNA was extracted using a 1.5x CTAB buffer (1.5% w/v CTAB, 75mM Tris-HCl pH 8.0, 15mM EDTA pH 8.0, 1.05M NaCl). Leaf tissue was harvested and added to a 2ml eppendorf tube alongside 2 ball bearings and 600µl extraction buffer. Tissue was homogenised for 2 min by shaking at 30Hz in a Retsch Mixer Mill 300

(Retsch, www.retsch.com). The homogenised mixture was then heated to 65°C for 20 min, followed by the addition of an equal volume of 24:1 chloroform: isoamyl alcohol (IAA). Mixtures were briefly vortexed before being centrifuged at 12,000rpm for 5 min. 500µl supernatant was transferred to a fresh tube and 2.5 volume of 100% v/v ethanol added and gently mixed by pipetting. DNA precipitate was collected by centrifugation and washed with 70% v/v ethanol, before being air-dried and re-suspended in 100µl Tris-EDTA (TE) buffer. DNA yield was checked using a nanodrop (Thermoscientific, www.nanodrop.com).

2.5.2. SDS 96-well plate extraction

For large numbers of samples a high-throughput SDS DNA extraction protocol was used. Tissue was harvested into 1.2ml collection tubes arranged in a 96-well rack. Ball bearings and 500µl SDS extraction buffer (0.5% w/v SDS, 200mM Tris pH 7.5, 25mM EDTA pH 8.0, 250mM NaCl) were added, and tissue homogenised in the same way as described for CTAB extractions. Plates were spun at 4000rpm for 15 min and 200µl supernatant transferred to a 96-well polypropylene plate. DNA was precipitated with an equal volume of 100% v/v isopropanol, and collected by centrifugation, before being washed with 70% v/v ethanol, air-dried and finally re-suspended in TE buffer.

2.5.3. Polymerase chain reaction

Validation of transgenic rice plants was carried out by PCR. Primers amplifying either the hygromycin resistance gene, or the expression module within a construct were used to confirm individual plant genotypes. In some cases betaine was used as an additive to aid strand separation. For maize lines in which *Mutator* transposons were predicted to be present in specific genes, two separate PCRs were undertaken (Figure 2.1). First, primers flanking the insertion site were used in order to amplify a WT amplicon. Second, one of the WT primers was used in conjunction with a 1:1 mix

of two primers that anneal to the *Mutator* transposon at either the 5' or 3' end. In combination, these two PCR reactions allowed WT (WT amplicon present but no transposon amplicon), heterozygous (both WT and transposon amplicon present), and homozygous mutant (transposon amplicon present but no WT amplicon) plants to be distinguished (Figure 2.1). If the transposon was present, it was theoretically possible for the flanking WT primers to amplify across the transposon to give a larger amplicon than observed for the WT allele, however this was not routinely observed. This may be because the extension time used was insufficient to amplify the larger product size. For *Ds* transposons an equivalent genotyping strategy was employed. In general, either DMSO or betaine was used to aid strand separation.

PCR conditions for both rice and maize were:

95°C	5 min	
95°C	30s	35x
53°C-57°C	30s	
72°C	1 min/ kb	
72°C	10 min	
15°C	∞	

2.5.4. Gel electrophoresis

PCR reactions and restriction digests (see section 2.2.4) were visualised by gel electrophoresis. Agarose gels (1-2% w/v for PCR products depending on the size of the expected amplicon, 1% w/v for restriction digestion fragments) were prepared using 1x Tris-Acetate-EDTA (TAE) buffer (40mM Tris, pH 7.6, 20mM acetic acid, 1mM EDTA) with ethidium bromide added to a final concentration of 0.3µg/ml. Samples were loaded and gels run routinely at 100V for the required time. Either a 100bp or 1kb DNA ladder was used to determine DNA fragment sizes (ThermoFisher, www.thermofisher.com). Gels were imaged using a Benchtop 3UV Transilluminator (UVP, www.uvp.com) using a UV wavelength of 302nm. Gel images were acquired using VisionWorks®LS Analysis Software (UVP, www.uvp.com/visionworks).

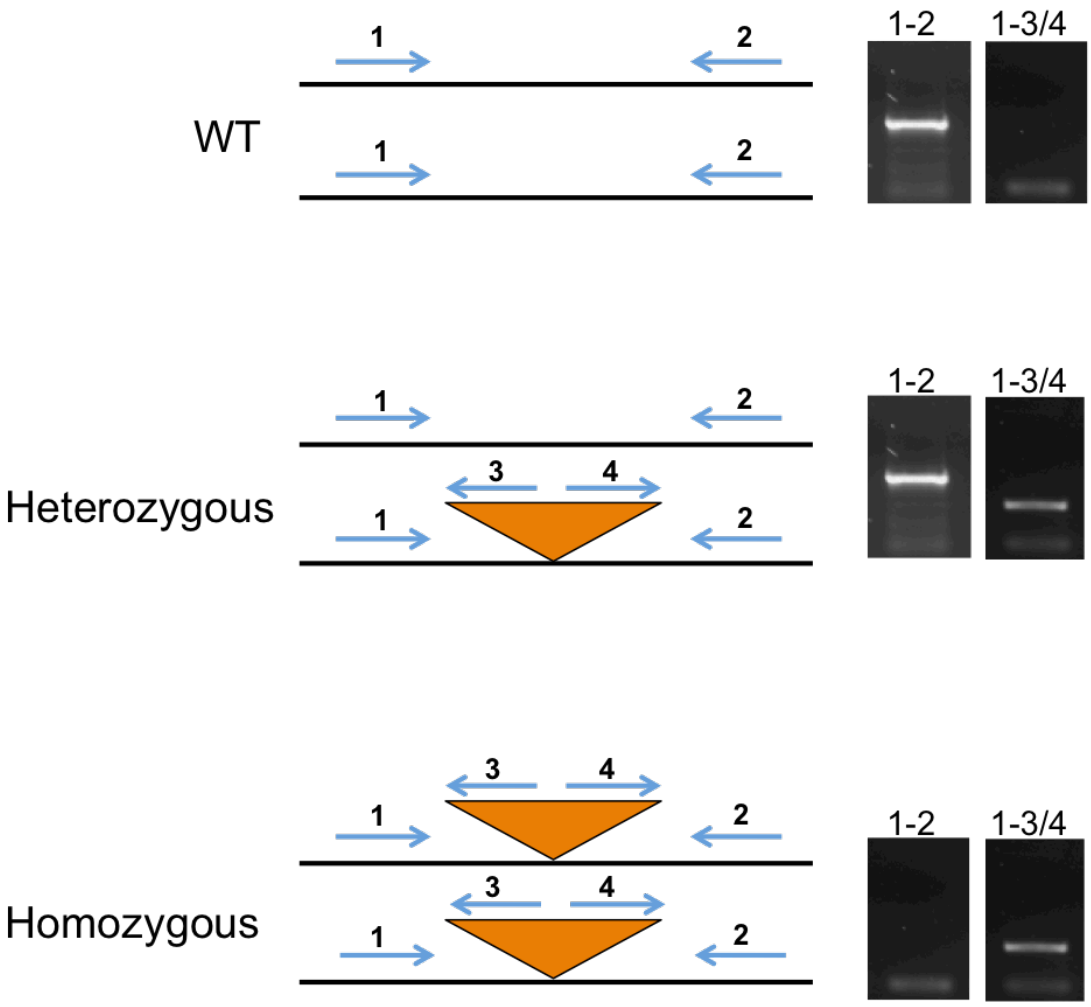


Figure 2.1 PCR genotyping strategy for maize mutants. Two combinations of primer pairs (represented by blue arrows) enabled the three possible genotypes to be distinguished. Primers 1 and 2 amplified from the WT allele. Primer 1 and a 1:1 mix of primers 3 and 4 amplified from the *Mutator* transposon (orange). WT plants produced an amplicon for primer combination 1-2, but not 1-3/4. Heterozygous plants produced an amplicon for primer combination 1-2 and 1-3/4. Homozygous plants produced no amplicon for primer combination 1-2 (likely because the PCR extension time was insufficient to amplify across the transposon), but did produce an amplicon for primer combination 1-3/4.

2.6. Expression analysis

2.6.1. RNA extraction

RNA for expression analysis in mature rice and maize leaves was extracted using the Qiagen RNeasy Plant Mini kit (Qiagen, www.qiagen.com), as per the manufacturer's instructions. Samples were homogenised as described for DNA extraction, except that the homogeniser adaptors were chilled in a -80°C freezer for at least 30 min beforehand, and samples kept on dry ice before and after homogenisation. RNA was extracted from maize leaf primordia using TRIzol® (Thermofisher, www.thermofisher.com). Primordia samples were dissected out on ice, and frozen immediately using liquid nitrogen. 1ml TRIzol was then added and the sample immediately vortexed, before being incubated at room temperature for 5 min and then centrifuged at 12,000rpm for 10 min at 4°C. The resulting supernatant was transferred to a fresh tube, and 200µl chloroform added before vortexing again. After a further incubation at room temperature for 3 min, samples were centrifuged at 12,000rpm for 10 min at 4°C. The aqueous phase was transferred to a fresh tube, and RNA precipitated by addition of 500µl 100% v/v isopropanol and incubation at room temperature for 10 min. RNA was then pelleted by centrifugation again at 12,000rpm for 10 min at 4°C. The supernatant was discarded and the pellet washed with 75% v/v ethanol, before a final centrifugation at 7,500rpm for 10 min at 4°C. The supernatant was again discarded and the pellet air-dried for ~10 min before being re-suspended in 30µl diethylpyrocarbonate (DEPC) water.

2.6.2. Reverse transcription-PCR

DNA contamination was removed from extracted RNA using TURBO DNase (Thermofisher, www.thermofisher.com), as per the manufacturer's instructions. RNA was then used for cDNA synthesis using Superscript III (Thermofisher, www.thermofisher.com). For cDNA synthesis, 1µl of oligo$(dT)_{20}$ primers (50µM), 1µg RNA, 1µl dNTPs (10mM) were made up to a total volume of 13µl in a nuclease free

microcentrifuge tube using DEPC water. This mix was then heated to 65°C for 5 min and then cooled on ice for at least 1 min. 4µl of 5X First Strand buffer, 1µl dithiothreitol (DTT) (0.1M), 1µl RNasin® Ribonuclease Inhibitor (Promega, www.promega.co.uk) and 1µl Superscript III reverse transcriptase (Thermofisher, www.thermofisher.com) were then added to the tube, and then incubated at 55°C for 60 min and then 70°C for 15 min. Ubiquitin (rice) or GADPH (maize) primers were used to check cDNA amplification, and gene-of-interest amplifications carried out with gene-specific primers. Reactions were normally run for 35 cycles.

2.7. Plant phenotyping

2.7.1. Plant photography

Plants were photographed individually against a black background using a Nikon D700 camera (www.imaging.nikon.com). Plants were photographed immediately before phenotypic analysis of leaf cross sections was undertaken. Representative plants from each mutant background and WT comparison were chosen.

2.7.2. Fresh sectioning

Fresh sections of both rice and maize leaves, and maize roots, were cut using a vibratome. 1cm segments of leaves were taken between 10 and 12cm from the tip of leaf 5-6 (rice) or halfway along the proximal-distal axis of leaf 4-6 (maize). Maize roots were harvested near to the tip. For rice it was generally possible to section across the whole width of the leaf, however, maize leaves were cut in the medial/lateral axis to include at least three lateral veins either side of the mid-vein. Segments were set in small blocks of 7% w/v agar, and left to cool at 4°C for at least 1 hour before sectioning. Blocks were then cut down and glued to a circular mount so that the sample stood vertically, with veins perpendicular to the mount. An initial cut of 100µm was made, followed by cuts decreasing by 10µm each time until the section consistently ripped. Depending on exact leaf age, it was generally possible to

take rice leaf sections of 30-40µm, whereas maize leaf and root sections were more often 50-60µm. Leaf sections were then cleared using 3:1 ethanol: acetic acid. Rice leaf sections were cleared for 24 hours, whereas maize leaf sections were cleared for 10 min before either imaging immediately or storing in 70% v/v ethanol overnight. Maize root sections were not cleared. Sections were floated on 70% v/v ethanol (leaves) or water (roots) under a coverslip, before imaging using a Leica DMRB microscope with QImaging MicroPublisher camera (QImaging, www.qimaging.com) and Image-Pro Insight software (MediaCybernetics, www.mediacy.com). Images were taken using both brightfield and UV illumination, and subsequent quantification of leaf anatomical traits was undertaken using the ImageJ software package (www.imagej.nih.gov).

2.7.3. Tissue fixation, embedding and sectioning

Foliar and husk leaf segments were harvested from mature maize leaves and fixed for 24 hours in FAA fixative (50% v/v ethanol, 3.75% v/v formaldehye, 5% v/v acetic acid), before being transferred to 50% v/v and then 70% v/v ethanol. Maize root segments were fixed in 4% paraformaldehyde fixative (4% w/v paraformaldehyde with 0.1% tween-20 and 0.1% triton-x-100) at 4°C for 3 days before being dehydrated in 10%, 30%, 50% and 70% v/v ethanol solutions. Segments were then embedded in paraffin wax using a Tissue-Tek VIP machine (Sakura, www.sakura.eu), set on the following program:

• 70% EtOH	2 hours	35°C	
• 80% EtOH	2 hours	35°C	
• 90% EtOH	2 hours	35°C	
• 95% EtOH + Eosin	2 hours	35°C	
• 100% EtOH	2 hours	35°C	x3
• Histoclear	2 hours	35°C	x3
• Paraffin	2 hours	60°C	x2
• Paraffin	3 hours	60°C	x2

Embedded segments were then set in wax blocks and left to solidify at 4°C overnight. Sections (10µm) were cut using a Leica RM2135 rotary microtome. The resultant ribbon was floated on water, before residual water was blotted off and the slide left to dry at 40°C overnight. Leaf sections were stained using Safranin-O (50% ethanol, 1% Safranin-O) and Fastgreen (95% ethanol, 0.5% Fastgreen). Slides were stained at room temperature using the following protocol:

- Histoclear 5 minutes x2
- 100% EtOH 2 minutes x2
- 70% EtOH 2 minutes x2
- Safranin-O 30 minutes
- 70% EtOH 2 minutes x2 with gentle shaking
- Fast-Green 30 seconds
- 100% EtOH 2 minutes x2 with gentle shaking
- Histoclear 5 minutes

Root sections were stained with toluidine blue (0.05% toluidine blue in 0.05M citrate buffer at pH 4.2) using the following protocol:

- Histoclear 5 minutes x2
- 100% EtOH 2 minutes x2
- 85% EtOH 2 minutes x2
- 70% EtOH 2 minutes x2
- 50% EtOH 2 minutes x2
- 30% EtOH 2 minutes x2
- 10% EtOH 2 minutes x2
- dH_2O 2 minutes x2
- Toluidine blue 1 minute x1
- dH_2O 2 minutes x2
- Histoclear 1 minute x1

Slides were then dried with gentle blotting of the slide edge, before a drop of DPX mounting media was added and a coverslip applied. Slides were visualised and imaged as described in section 2.7.2.

2.7.4. Leaf clearing

To assess vein density across the full width of maize leaves, a 3cm region at the midpoint of the leaf was harvested and initially cleared in 3:1 ethanol: acetic acid at room temperature for at least 24 hours. Following this, leaves were boiled in 85% v/v EtOH for an hour, before being transferred to pre-warmed lactic acid at 70°C overnight. Leaves were then stored in 70% v/v EtOH before being photographed using a Nikon D300 camera (www.imaging.nikon.com).

2.7.5. GUS staining

Plant tissue was incubated in 90% v/v ice-cold acetone for 1 hour at -20°C, rinsed in 100mM phosphate buffer (pH 7.6), and then transferred into staining solution (100mM phosophate buffer, 10mM EDTA, 2mM ferricyanide, 2mM ferrocyanide, 2mM X-GlcA) at 37°C overnight. Stained plants were then fixed for 1 hour in 3:1 ethanol: acetic acid, before being transferred to 70% v/v ethanol. Images were taken either of whole plants, or for finer-scale imaging of young leaf primordia, samples were embedded in the same way described in section 2.7.3, and sectioned at 25µm. Wax was removed by transfer to histoclear for 10 minutes, before tissue was hydrated and then dehydrated to remove some eosin stain, by transferring slides to 100% (x2), 85% (x2) and 70% v/v ethanol for 1 minute each.

2.8. Phylogenetic analysis

2.8.1. Tree construction

Relationships between gene sequences were elucidated through the construction of phylogenetic trees. Sequences from seven species were included: *Zea mays* (B73), *Sorghum bicolor*, *Setaria italica* and *Oryza sativa* from the grasses, *Arabidopsis thaliana* and *Solanum lycopersicum* from the eudicots, and *Physcomitrella patens* from the mosses (included as an outgroup). Primary transcript proteomes were downloaded for each of the seven species from Phytozome11, and gene of interest

protein sequences used as a query in a BLASTp search (evalue of 1e-3) against this proteome database. The retrieved sequences were aligned using MergeAlign (Collingridge and Kelly 2012), and the alignment was trimmed using trimAL to remove poorly aligned regions. Columns with less than 30% of the sequences represented were discounted from further analysis (Capella-Gutiérrez et al. 2009). Trimmed alignments were used as input for inference of a maximum likelihood tree using RAxML (Stamatakis 2014).

2.8.2. Enrichment analysis

Enrichment analyses were carried out for Gene Ontology (GO) (www.geneontology.org) terms, MaizeCyc pathways (www.maizecyc.maizegdb.org), MapMan terms (www.mapman.gabipd.org) and Pfam domains (www.pfam.sanger.ac.uk). Each category of neofunctionalised genes was tested relative to all annotated maize genes (Appendix 4), and all analysed homeolog genes (Appendix 5), using code written by Steve Kelly. This code calculated P-values by approximating Wallenius' noncentral hypergeometric distribution. GOseq was used to compensate for over-detection of differential expression for long and highly expressed transcripts (Young et al. 2010). The resulting p-values were subject to multiple hypothesis test correction to correct for Type I family-wise error using the Benjamini-Hochberg method. Significantly enriched annotation terms were identified as those that obtained a corrected p-value of ≤ 0.05.

2.8.3. Calculation of K_a/K_s ratios

Maize gene sequences were obtained from the B73 reference genome (Schnable et al. 2009), and sorghum genes from the published *Sorghum bicolor* genome (Paterson et al. 2009). For each homeolog gene-pair the longest representative gene models were selected and the amino acid sequences were aligned to the sorghum ortholog using MergeAlign (Collingridge and Kelly 2012). The amino acid sequences

in the three-sequence multiple sequence alignment were then replaced with their corresponding coding sequences in order to test for selection acting at the nucleotide level. Each aligned homeolog gene was compared independently to its sorghum ortholog and the ratio of non-synonymous (protein changed) substitution rate per nonsynonymous site (K_a) to synonymous (protein unchanged) substitution rate per synonymous site (K_s) was calculated using KaKs_Calculator Toolbox 2.0 (Wang et al. 2010). K_a/K_s values for homeolog genes in each category were averaged and standard errors calculated.

2.8.4. Monte Carlo resampling tests

To account for differences in sample sizes and avoid requirement for distributional assumptions, Monte Carlo resampling tests were employed to assess the significance of differences in K_a/K_s ratios between groups. To do this, the mean K_a/K_s ratio of the test group was compared to a reference group of the same size that was randomly selected with replacement from the whole pool of homeolog genes. For each test group the process was repeated 10,000 times and the proportion of times that the mean of the test group was greater than the reference group recorded. A proportion ≥ 0.95 indicates a significantly greater difference in the test group than the overall population of homeolog genes, whereas a proportion of ≤ 0.05 indicates a significantly smaller difference in the test group.

Chapter 3: The impact of widespread regulatory neofunctionalisation on homeolog gene evolution following whole genome duplication in maize

3.1. Introduction

3.1.1. Whole genome duplications provide a source of evolutionary innovation

Whole genome duplications are a dominant feature of plant genome evolution and have been detected in all angiosperm lineages (Adams and Wendel 2005; Soltis et al. 2009). Post-duplication, many duplicated genes are lost in a process known as fractionation, with this loss often showing a bias towards one of the subgenomes (Thomas et al. 2006; Woodhouse et al. 2010). However, some duplicated genes are retained in the genome as homeolog gene-pairs. If retained, the individual genes in a homeolog gene-pair may have identical functions, they may partition and share the original gene function (subfunctionalisation), or they may diverge and develop novel functions (neofunctionalisation) (Ohno 1970; Lynch and Conery 2000; Moore and Purugganan 2005; Freeling 2008; McGrath and Lynch 2012). Neofunctionalisation has previously been categorised as regulatory neofunctionalisaton (R-NF) or coding neofunctionalisation (C-NF) (Moore and Purugganan 2005). C-NF provides novel protein function via a gain-of-function mutation in the open reading frame of a gene. In contrast, R-NF results from expression divergence, which provides extant protein function in novel temporal or spatial environments. Thus both R-NF and C-NF of homeolog gene-pairs have long been proposed as a major source of evolutionary innovation (Ohno 1970; Moore and Purugganan 2005). Whilst neofunctionalisation has been studied in specific genes, or at a genome scale in *Arabidopsis thaliana* and

Glycine max (Blanc and Wolfe 2004a; Duarte et al. 2006; Liu et al. 2011; Erdmann et al. 2010; Guo et al. 2013; Roulin et al. 2013), it is unknown to what extent genome-wide neofunctionalisation is important in other species, or to what extent R-NF and C-NF interact and interdepend.

3.1.2. Maize underwent a recent whole genome duplication

The grass lineage provides an excellent system to study genome-wide functional divergence of duplicated genes because of the contrasting evolutionary trajectories of the closely related species maize (*Zea mays* L.) and sorghum (*Sorghum bicolor*) (Swigonova et al. 2004a). Maize underwent a recent whole genome duplication event between 5 and 12 million years ago (Swigonova et al. 2004a). Since that time, rather than remain tetraploid, multiple chromosomal breakage and fusion events have combined the duplicated chromosomes into a diploid genome containing 10 novel mosaic chromosomes (Wei et al. 2007; Schnable et al. 2011). In contrast, the sorghum genome has retained the ancestral non-duplicated state (Swigonova et al. 2004a). Comparison of the maize genome to that of sorghum has thus facilitated the reconstruction of the two duplicated subgenomes of maize, based on biased fractionation (Paterson et al. 2009; Schnable et al. 2011). This analysis identified 3228 homeolog gene-pairs that have been retained post-duplication (Schnable et al. 2011). Given that maize has ~32,000 genes (Schnable et al. 2009), homeolog gene-pairs thus account for over 20% of gene content. Despite constituting a large fraction of the maize genome, it is not yet known to what extent homeolog gene-pairs have diverged in expression since the whole genome duplication event.

In this chapter a transcriptional dataset from two maize leaf types (foliar and husk) was exploited to investigate R-NF of genes in maize leaves following whole genome duplication. Foliar leaves are the primary source tissue in maize and develop from the main shoot apical meristem. In contrast, husk leaves develop from axillary meristems and function to surround and protect the developing female inflorescence

(the ear) (Iltis 2000). Widespread R-NF of retained homeolog genes in maize was found, with significantly more genes undergoing R-NF in foliar leaves than husk leaves. In addition, Enrichment analysis revealed that the functional annotations of R-NF genes are consistent with the biological role of the leaves in which those genes are expressed. Moreover, R-NF genes are under strong purifying selection and biased subgenome expression dominance occurs only in the presence of R-NF. Finally, this chapter demonstrated that relaxed purifying selection is more likely to occur in the absence of R-NF, and is further enhanced by expression dominance. Taken together, this chapter represents the first assessment on a genome-scale of homeolog gene-pair divergence, and provides evidence that regulatory-neofunctionalisation has played an important role in maize leaf evolution. This has implications for our understanding of the transcriptional regulation underpinning Kranz anatomy development in maize, but also contributes more broadly to our understanding of genome evolution following whole genome duplication events.

3.2. Results

3.2.1. R-NF during maize early leaf development

3.2.1.1. Signatures of R-NF are apparent in over 400 homeolog gene-pairs that are expressed during leaf development

To assess the evolutionary trajectory of homeolog gene-pairs in maize, a recently published dataset was analysed, which took advantage of biased fractionation of one ancestral genome in order to reconstruct the maize subgenomes (Schnable et al. 2011). This dataset was employed alongside a transcriptome dataset from different stages of foliar and husk leaf development (Wang et al. 2013b). Transcripts were quantified using RSEM (using default parameters) and the predicted coding sequences from version 5b of the Maize genome. RSEM is specifically designed to enable accurate expression quantification between highly similar gene sequences (Li and Dewey 2011), and has been used extensively to successfully distinguish the

expression of duplicated genes, including in maize (Harikrishnan et al. 2015; Li et al. 2016). The mean K_s between maize homeolog genes is 0.38 thus on average a 90bp paired end read (180bp) will have ~20 distinguishing SNPs thus enabling RSEM to accurately discriminate between reads uniquely originating from each homeolog. Furthermore, no homeolog gene-pairs have identical sequences, with even the most similar pair (GRMZM2G067707-GRMZM2G401139, 98% sequence identity) having 4 SNPs. This was sufficient to distinguish the genes in the RNA-seq dataset. For differential gene expression analysis, expected transcript counts originating from the same gene locus were summed and all possible pairwise comparisons between biologically replicated samples were performed using DESeq (Anders and Huber 2010). From these data, 2607 of the 3228 homeolog gene-pairs that are present in the maize genome are expressed (i.e. gene expression of both genes > 0) during early development in both foliar and husk leaves.

To assess homeolog expression during early leaf development differential expression between developmental stages was calculated. Expression values were normalised by DESeq using the default method, and in all cases differentially expressed genes were identified as those genes with a Benjamini-Hochberg corrected p-value ≤ 0.05 (Benjamini and Hochberg 1995). Genes were assigned to three expression profiles based on expression across three developmental stages during early leaf development: ascending (A), neutral (N) and descending (D) (Figure 3.1A). These three stages were: 1) shoot apical meristem plus plastochron-1 (P1) and P2 primordia (FP1/2 and HP1/2), 2) P3 and P4 primordia (FP3/4 and HP3/4) and 3) P5 primordia (FP5 and HP5). Analysis of gene expression during these stages thus enables investigation into the transcriptional changes that underpin early leaf development. The A expression cluster included those genes that increased significantly from either P1/2 to P3/4 (but not necessarily P3/4 to P5), from P3/4 to P5 (but not necessarily from P1/2 to P3/4), and those genes that increased

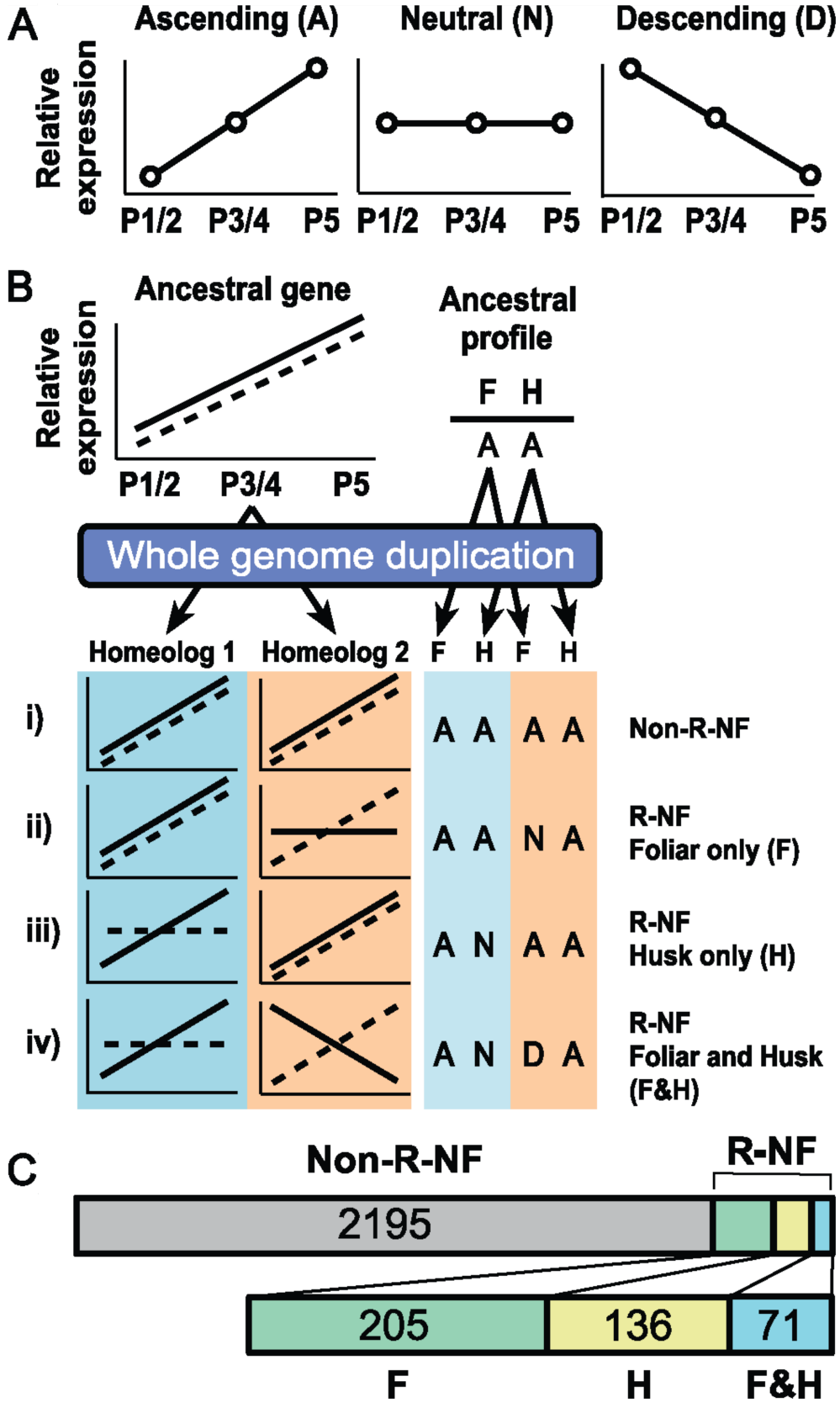


Figure 3.1 Identification and quantification of regulatory neofunctionalised genes. Figure legend is on next page.

Figure 3.1 Identification and quantification of regulatory neofunctionalised genes. A) Genes expressed in foliar (F) and husk (H) leaves were allocated to one of three expression profiles, ascending (A), neutral (N) and descending (D), based on relative expression levels at three stages of early leaf development: plastochron (P) 1/2, P3/4 and P5. **B)** Cartoon example of how an ancestral gene with expression allocated to profile A in both F and H could have evolved following whole-genome duplication. Blue backgrounds refer to the homeolog from subgenome-1, and orange to the homeolog from subgenome-2. Continuous lines refer to the expression profile (A, N or D) in F, whereas a dashed line refers to expression profile in H. 4 letter expression codes consist of [homeolog-1 F expression profile],[homeolog-2 F expression profile],[homeolog-1 H expression profile],[homeolog-2 H expression profile]. **i)** If expression profile of homeolog-1 matches expression profile of homeolog-2 in F, and expression profile of homeolog-1 matches expression profile of homeolog-2 in H, the gene pair is assumed to have undergone 0 expression changes since duplication, and is thus non-regulatory neofunctionalised (non-R-NF); **ii)** If expression profiles between homeologs match in H but not in F, the gene pair is said to have been regulatory neofunctionalised (R-NF) in F only; **iii)** If expression profiles between homeologs match in F but not in H, the gene pair is said to have been regulatory neofunctionalised (R-NF) in H only; **iv)** If expression profiles between homeologs do not match in F or H, the gene pair is said to have been regulatory neofunctionalised (R-NF) in F&H. **C)** Results of dividing homeolog gene-pairs up using the method described in (B). Numbers refer to homeolog gene-pairs that have expression profiles consistent with the labelled category. A total of 2607 homeolog gene-pairs are expressed in both foliar and husk leaf samples.

significantly across both steps ($p \leq 0.05$). In all of these cases the overall expression pattern is one of increasing expression. The same applies inversely to the descending category. In the neutral profile there could not be a significant difference between any of the samples. Genes that had significantly higher (or lower) expression in P3/4 than both P5 and P1/2 for both leaf types were assigned to 'peak' (or 'trough') profiles. The number of genes assigned to these profiles was low (Foliar "P3/4 Peak" profile n = 34, Foliar "P3/4 Trough" profile n = 10, Husk "P3/4 Peak" profile n = 27, Husk "P3/4 Trough" profile n = 8) and hence statistical analysis using these profiles was not possible (Wang et al. 2013b). Thus, they were excluded from further analysis.

Transcriptomes from two additional developmental stages were available for analysis: immature leaf and expanded leaf (FI, HI, FE, HE), however, for expression profile analysis only those stages which were directly equivalent between foliar and husk samples were used (Wang et al. 2013b). FP1/2 is equivalent to HP1/2, FP3/4 is equivalent to HP3/4 and FP5 is equivalent to HP5. However, HI is not the equivalent developmental stage to FI and HE is not the equivalent developmental stage to FE (Wang et al. 2013b). Thus, in order to provide a directly comparable developmental series, profiles did not include these later samples. A four-letter expression code was generated by taking the expression profile of the subgenome-1 homeolog in both foliar and husk leaves, followed by the expression profile of the subgenome-2 homeolog in both foliar and husk leaves. Together this forms a 4-letter expression code that describes the behaviour of a given gene-pair during early leaf development. For example, for a hypothetical homeolog gene-pair, if the subgenome-1 homeolog had an ascending (A) expression profile in both foliar and husk leaves, and the subgenome-2 homeolog had a descending (D) expression profile in foliar leaves and a neutral (N) expression profile in husk leaves then the expression code for this homeolog gene-pair would be AADN ([foliar profile

subgenome-1][Husk profile subgenome-1][Foliar profile subgenome-2][Husk profile subgenome-2]) (Figure 3.1B).

The maize whole genome duplication is thought have an allotetraploid origin, whereby two closely related diploid species were hybridised (Gaut and Doebley 1997; Swigonova et al. 2004a; Steige and Slotte 2016). Unfortunately, the parental genomes of maize are not available, and therefore it is not possible to assess how much divergence in homeolog gene expression predated the whole genome duplication event. However, this work considered gene expression profiles that are designed to include broad changes in expression across multiple stages of development, rather than expression levels at one particular stage. It is, therefore, most parsimonious to assume that expression profile allocation in the diploid parental species were the same, and thus that the resultant homeolog gene-pair had the same profile allocation in the foliar and the husk leaves. Therefore, expression profiles in which homeolog-1 and homeolog-2 are allocated to the same expression clusters indicate that the gene-pair has not diverged in expression (Figure 3.1Bi). Correspondingly, homeolog gene-pairs that have non-identical expression profiles represent divergence in expression (Figure 3.1B ii-iv). In the simplest case, where one gene of a homeolog gene-pair has changed expression profile in either foliar or husk leaves, then the most parsimonious explanation is that there has been one change in expression since genome duplication (Figure 3.1B ii-iii). If both genes in a homeolog gene-pair differ in expression profile in both leaf types then there have been at least two changes in expression since genome duplication (Figure 3.1B iv). Using the above criteria, 2195 (84%) of the expressed homeolog gene-pairs (n = 2607) have not changed expression profiles in leaf development since whole genome duplication, 341 (13%) have undergone 1 change, and 71 (3%) have undergone 2 or more changes (Figure 3.1C). Taken together these data indicate that

16% of expressed homeolog gene-pairs, and 13% of all homeolog-gene pairs, have diverged in expression in the context of maize leaf development (Figure 3.1C).

There is extensive literature concerning the fate of duplicated genes (Ohno 1970; Lynch and Conery 2000; Moore and Purugganan 2005; Innan and Kondrashov 2010; McGrath and Lynch 2012; Freeling et al. 2015), and there is frequently confusion regarding the terminology used to describe functional divergence of duplicate genes (Freeling et al. 2015). As this study was concerned with shifts in expression following gene duplication, functional divergence of duplicate genes was assumed to result from regulatory changes upstream of the gene. Regulatory divergence of gene pairs has previously been divided into regulatory subfunctionalisation (R-SF) and regulatory neofunctionalisation (R-NF) (Moore and Purugganan 2005). In R-SF, the ancestral expression of a gene is now divided between the two duplicated genes in different tissues. In R-NF, one gene member acquires a novel expression pattern not present in the ancestor. In practice, however, it is difficult to accurately distinguish R-SF and R-NF, and a given gene-pair may exhibit traits associated with either case in different contexts (Moore and Purugganan 2005). In this analysis the majority (97%), of gene-pairs identified as having diverged in expression have expression profiles that are asymmetrically distributed between the gene-pairs. In addition, S-NF likely involves the ancestral expression level of a gene being shared between the two duplicated genes across different tissues (Liu et al. 2011), however, in this analysis, changes to expression profile rather than expression levels were considered. In combination, this means it is difficult to infer R-SF. However, within the two-change category, 13 homeolog gene pairs do have palindromic expression profiles (DNND, ANNA, NAAN), which could be consistent with R-SF of the ancestral gene-function (Table 3.1). Furthermore, this analysis does not take into account dosage subfunctionalisation, whereby the total expression of an ancestral gene is partitioned across two duplicates such that the total expression remains the same after

		R-NF					
Non-R-NF		F		H		F&H	
Profile	Homeolog gene-pairs	Profile	Homeolog gene-pairs	Profile	Homeolog gene-pairs	Profile	Homeolog gene-pairs
DDDD	5	NNAN	47	DNDD	1	AAND	1
NNNN	2128	DNAN	4	DDDN	3	ANND	1
NANA	9	AANA	6	NNND	5	DNND	2
AAAA	14	ANNN	65	NNNA	36	NNDD	2
NDND	4	DNNN	39	NDNN	19	NNDA	2
ANAN	24	ANDN	3	ANAA	11	NDAN	3
DNDN	11	NAAA	4	AAAN	11	DNNA	2
DADA	0	NNDN	34	NANN	50	ANNA	7
ADAD	0	DDND	1			DNAA	1
		DAAA	1			ANDA	1
		NDDD	1			AANN	17
						NNAA	19
						NADD	1
						DDNN	6
						NAAN	4
						NADN	2
Total	2195	Total	205	Total	136	Total	71

Table 3.1 Frequency of each 4-letter expression profile across expressed homeolog gene-pairs. R-NF = regulatory neofunctionalised, F = Foliar, H = Husk. 4 letter expression codes consist of [homeolog-1 F expression profile],[homeolog-2 F expression profile],[homeolog-1 H expression profile],[homeolog-2 H expression profile]. Counts were of homeolog gene-pairs rather than individual homeolog genes.

duplication (Gout and Lynch 2015). Crucially, this would not necessarily lead to a change in expression pattern during early leaf development, and as such gene-pairs identified here as non-divergent could still have undergone dosage subfunctionalisation. Given the small number of gene-pairs for which S-NF is possible to infer, all gene-pairs that have diverged in expression are hereon referred to as R-NF, defined as a change in expression profile of at least one homeolog gene. These results are consistent with the previous finding in *Arabidopsis,* that divergent expression patterns were more frequently the result of R-NF than R-SF (Liu et al. 2011).

3.2.1.2. R-NF genes are more likely to have altered expression patterns in foliar leaves than in husk leaves

To determine whether R-NF is associated equally with altered foliar and husk leaf expression patterns, homeolog gene-pairs exhibiting one change in expression profile were categorised into those that provide evidence for R-NF in foliar leaves and those that provide evidence for R-NF in husk leaves. Assuming that the observed expression code is the result of one change in expression profile, it follows that the ancestral pre-whole genome duplication profile could have been that of either subgenome-1 or subgenome-2. For example, for the expression code AANA, the ancestral expression profile could have been either AA or NA. Post-duplication this would have become AAAA or NANA, and in both cases the single change to generate AANA must occur in the foliar leaf (Figure 3.1B ii-iii). Partitioning the homeolog gene-pairs from the one-change category in this way revealed that significantly more homeolog gene-pairs are R-NF in foliar leaves than in husk leaves ($n = 205$ and $n = 136$ respectively, $p \leq 0.001$) (Figure 3.1C). A full breakdown of the number of gene-pairs allocated to each category is provided in Table 3.1.

3.2.1.3. Functional annotations of R-NF homeolog genes support selection for distinct biological roles in foliar and husk leaves

The unequal distribution of R-NF genes between foliar and husk leaves might suggest that foliar leaves have been subject to greater selection since the recent whole genome duplication event. If this is the case, R-NF genes should enhance the role of the foliar leaf as the primary source tissue. To determine whether specific functional annotations are over-represented in the homeolog gene-pair dataset the occurrence of annotation terms derived from gene ontology (GO) terms, MaizeCyc pathways, MapMan terms and Pfam domains were independently assessed relative to both all annotated maize genes and all annotated homeolog genes (Appendices 4&5). In the 2195 non-R-NF genes there are a total of 133 significantly over-represented annotation terms relative to all annotated maize genes ($p \leq 0.05$), whilst in the 412 R-NF genes there are 56, 7 and 20 over-represented annotation terms in foliar, husk and both leaves respectively (Appendix 4). Given the use of all maize genes as a baseline, it might be expected that genes involved in photosynthesis would be over-represented in all categories, given that all of the genes analysed in this study were selected based on expression in the foliar and husk leaf. It is notable, therefore, that photosynthesis related terms only appear enriched in the R-NF-F category (Appendix 4). R-NF of photosynthesis-related genes in foliar leaves may indicate that selection has acted on expression patterns of photosynthetic genes in this tissue type. As foliar leaves are the primary photosynthetic tissue, this selection may have been for enhanced photosynthetic productivity and/or efficiency since the whole genome duplication event. Other enriched annotation terms in this category include cell wall precursor synthesis and hormone metabolism (Appendix 4).

Although fewer in number than those in foliar leaves, husk-specific R-NF genes also provide evidence of selection for specific functions with over-represented annotations including responses to stress and proline metabolism (Appendix 4). Proline is

accumulated by plants in response to osmotic stress, and has previously been found to be the major osmotic regulator accumulated in the husk leaves of maize (Xu et al. 2011). R-NF of stress response genes in husk leaves might indicate selection on these leaves for improved stress tolerance.

Of the 20 annotation terms over-represented in those homeolog gene-pairs R-NF in both foliar and husk leaves, there are eight concerning nitrogen metabolism, plus those related to cell wall biosynthesis and transport (Appendix 4). The over-representation of genes involved in nitrogen metabolism is consistent with previous findings showing that genes involved in amino acid biosynthesis and protein catabolism have been under artificial selection since the domestication of maize (Wright et al. 2005). Interestingly, the only annotation terms that are consistently enriched across all categories are those concerned with the regulation of transcription. Thus, although there is a distinct partitioning of R-NF genes in both foliar and husk leaves that is consistent with leaf function, in each category there are significantly more transcriptional regulators then would be expected given the number of such genes in the maize genome. This is consistent with the previous finding that transcription factors are preferentially retained following whole genome duplications (Blanc and Wolfe 2004a; Freeling 2009).

3.2.2. The interaction between the maize subgenomes and R-NF

3.2.2.1. Subgenome-1 homeologs are more likely to be R-NF than subgenome-2

The two ancestral subgenomes of maize have recently been reconstructed (Schnable et al. 2011). For those homeolog gene pairs that are R-NF in either the foliar or the husk (but not both) there are two possible ancestral profiles (assuming only one change in expression profile). Whilst the ancestral gene expression profile is unknown, assuming that the expression profiles of non-R-NF homeolog gene-pairs are an unbiased sample of the underlying distribution of all expression profiles, the most likely ancestral profile for R-NF homeolog gene-pairs can be inferred using a

Bayesian model comparison approach (Appendix 6). For example, in R-NF-Foliar profile ANDN, homeolog-1 has ascending expression in the foliar, whereas homeolog-2 has descending expression, and in the husk both have neutral expression. Therefore, in this example, the ancestral profile immediately after duplication is assumed to be either A or D in the foliar, and N in the husk (ANAN or DNDN). For those genes that are non-R-NF there are 24 gene pairs that have the profile ANAN (and by extension the ancestral profile AN), and 11 that have the profile DNDN. Therefore, it is assumed to be more likely that for the R-NF profile ANDN the more likely ancestral profile is ANAN. If this is true, then the change that resulted in the observed ANDN profile occurred in the subgenome-2 homeolog. Applying this approach across all gene-pairs that are R-NF in the foliar or husk, revealed that significantly more genes are likely to be R-NF from changes in the subgenome-1 homeolog than from the subgenome-2 homeolog ($n = 197$ and $n = 144$, $p \leq 0.01$). However, there was no association between subgenome and leaf type ($p > 0.05$, Chi-squared test for independence). Therefore, although there are more R-NF genes in foliar leaves and there are more R-NF genes from subgenome-1, there is no evidence that subgenome-1 genes are preferentially R-NF in foliar leaves or *vice versa*.

3.2.2.2. Biased subgenome expression dominance is only observed in R-NF homeolog gene-pairs

Subgenome dominance is a well-established phenomenon in allopolyploids, whereby the genes from one ancestral genome are more likely to be retained, and contribute higher expression levels than those genes in the other subgenome (Grover et al. 2012; Garsmeur et al. 2014). Expression bias towards the dominant subgenome has been reported in a number of plant allopolyploids, including maize (Wang et al. 2006; Buggs et al. 2010; Flagel and Wendel 2010; Schnable et al. 2011; Cheng et al. 2012; Li et al. 2014). In *Brassica rapa*, subgenome dominance has been linked to

epigenetic regulation and transposon distribution (Cheng et al. 2016). However, relatively little is known about subgenome dominance in maize, nor is it clear how it interacts with functional divergence of duplicates. As maize homeolog genes have previously been assigned to subgenomes based on fractionation patterns (Schnable et al. 2011), the extent to which R-NF impacted on subgenome expression dominance was investigated.

In previous analyses of this kind a homeolog gene was said to 'dominate' if its expression was at least double that of its homeolog pair (Schnable et al. 2011; Pophaly and Tellier 2015). To improve both the specificity and sensitivity of expression dominance classification, statistically significant ($p \leq 0.05$) differential expression of homeolog gene-pairs was used rather than the two-fold method employed previously. This analysis revealed that for non-R-NF genes there is no detectable bias in subgenome expression dominance, with both subgenomes having equal numbers of dominant homeologs at all stages of development in both foliar and husk leaves (Figure 3.2A). Interestingly, the proportion of genes that exhibit dominance in either subgenome is higher at later stages of development in both foliar and husk leaves (Figure 3.2A). This indicates that in non-R-NF homeolog gene-pairs, subgenome dominance is progressively acquired during development.

In contrast, R-NF genes exhibit a consistent subgenome dominance bias, and do not exhibit the same progressive increase in dominance during development as non-R-NF genes (Figure 3.2B-D). In virtually all cases, the proportion of homeolog gene-pairs that exhibit subgenome-1 dominance is larger than the proportion that exhibit subgenome-2 dominance (Figure 3.2B-D). Notably, the earlier primordia samples exhibit stronger subgenome-1 dominance bias than later stages of leaf development (Figure 3.2B-D). This indicates that in R-NF homeolog gene-pairs, biased subgenome dominance is already established in the earliest leaf primordia stages and becomes progressively less pronounced during leaf development (Figure 3.2B-

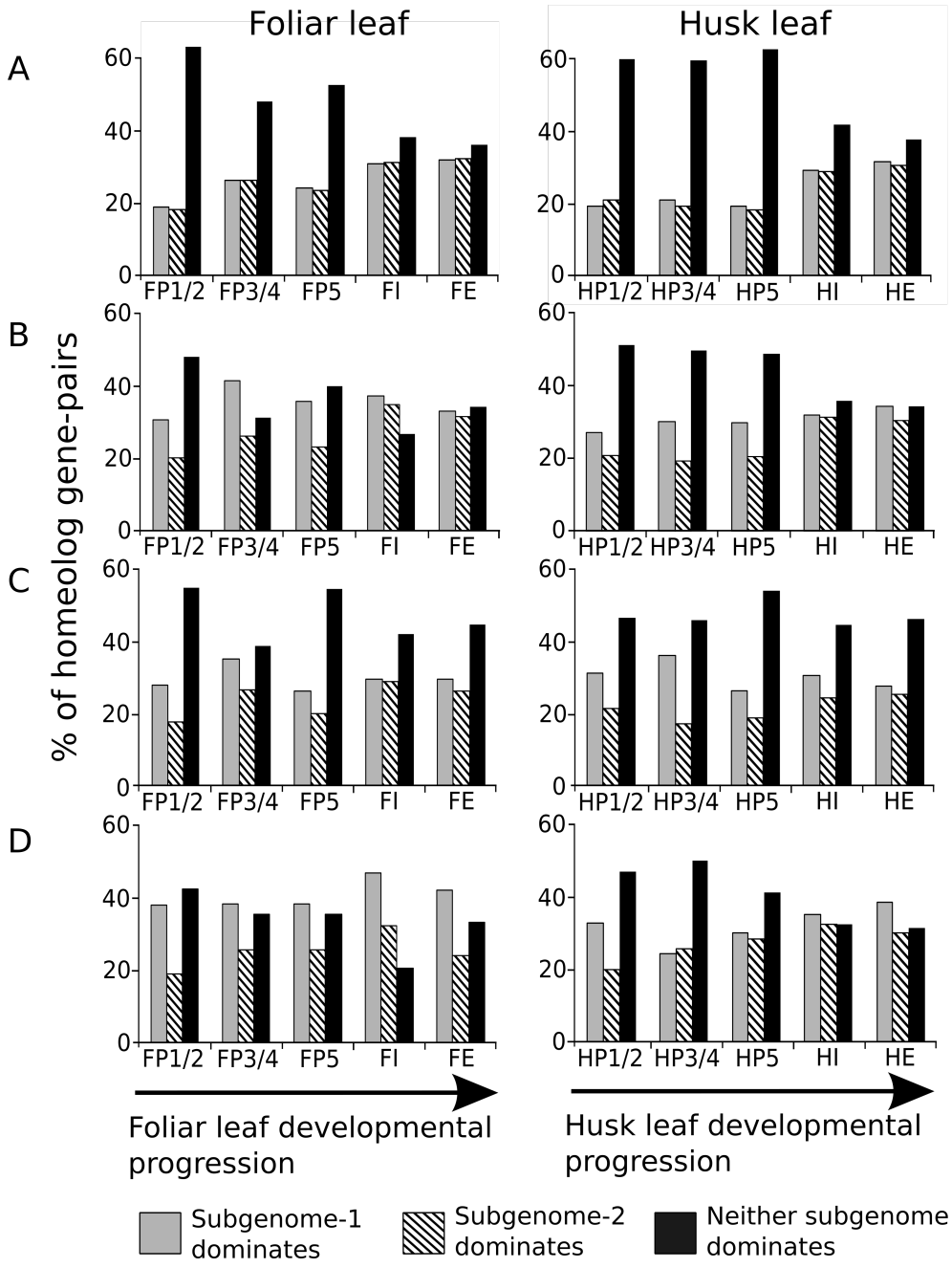


Figure 3.2 Barplots showing expression dominance of homeolog gene-pairs during leaf development. A) Non-R-NF homeologs. **B)** R-NF in just foliar leaves. **C)** R-NF in just husk leaves. **D)** R-NF in both foliar and husk leaves. Expression dominance was defined using differential expression testing. X-axis labels refer to distinct developmental stages in both foliar and husk leaf development. P- plastochron. I- immature leaf, E- fully expanded leaf (Wang et al. 2013b).

D). This is in contrast to non-R-NF genes where subgenome dominance is progressively acquired.

3.2.3. Selection acts differently on non-R-NF and R-NF homeolog gene-pairs

3.2.3.1. R-NF homeolog gene pairs are under stronger purifying selection and have more similar evolutionary rates than non-R-NF gene pairs

It is not clear whether genes that have diverged in expression profile may also diverge in terms of protein function. To determine whether R-NF genes exhibit differential rates of coding sequence evolution, each homeolog gene-pair was aligned to its single copy sorghum ortholog and the K_a/K_s ratios were calculated. A K_a/K_s ratio of less than 1 indicates negative (purifying) selection, whereas greater than 1 indicates positive selection. This analysis revealed that in general homeolog gene-pairs are under strong purifying selection (Figure 3.3A). Moreover, comparing R-NF homeolog gene-pairs to all homeolog gene-pairs revealed that R-NF homeolog gene-pairs are under significantly higher purifying selection than non-R-NF pairs (Figure 3.3A).

Although generally R-NF homeolog gene-pairs are on average under greater purifying selection than non-R-NF genes, it might be that one homeolog in a pair is under relaxed purifying selection relative to the other homeolog. To test this hypothesis the magnitude in the difference in K_a/K_s ratio of both homeologs in each gene-pair was calculated. This revealed that R-NF genes have, on average, a significantly smaller difference in the magnitude of K_a/K_s than non-R-NF homeolog gene-pairs, indicating that non-R-NF gene-pairs are more likely than R-NF gene-pairs to have more dissimilar selection acting on the coding region of each gene (Figure 3.3B). When analysed by category this difference is significant for those homeolog gene-pairs R-NF in the foliar and in both the foliar and husk, but not for

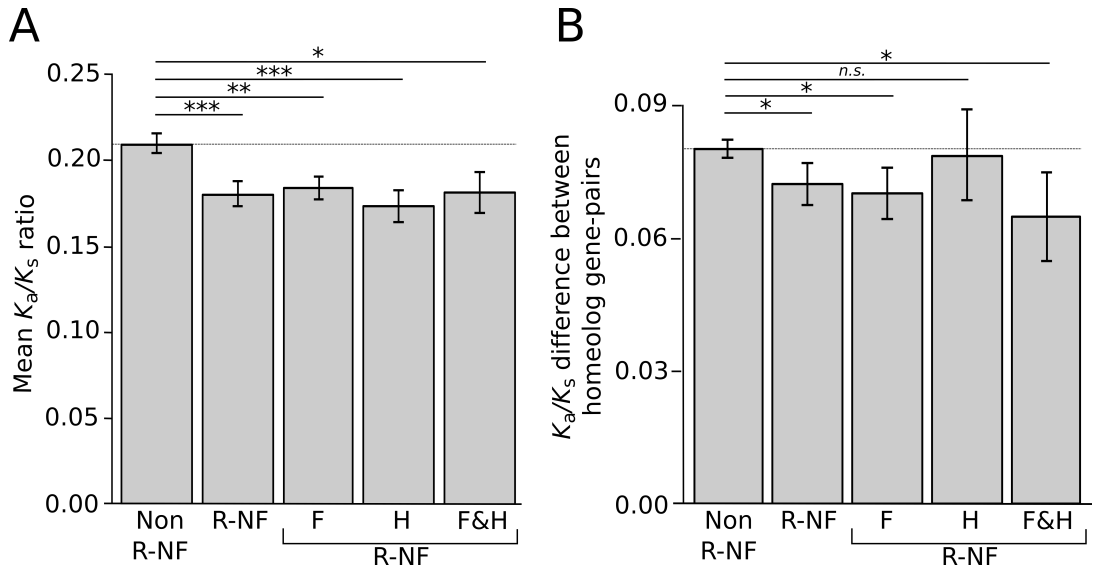


Figure 3.3 Barplots displaying *Ka/Ks* ratios for homeolog gene-pairs. A) Mean K_a/K_s ratio of each homeolog plotted for each R-NF category. A ratio of less than 1 indicates negative (purifying) selection, whereas greater than 1 indicates positive selection. P-values were calculated for all R-NF homeolog gene-pairs (n = 412), and each category of R-NF homeolog gene-pairs using Monte Carlo resampling tests. F- Foliar; H- Husk; F&H- Foliar and Husk. All R-NF categories show significantly greater purifying selection than non-R-NF homeolog gene-pairs. **B)** Mean magnitude of the difference in K_a/K_s ratios between each gene in a homeolog pair. The greater the difference in K_a/K_s the more selection has been relaxed relative to the other gene in the pair. The closer the ratio is to zero the more similar the selection acting upon both genes in a pair. P-values were calculated using Monte Carlo resampling tests. R-NF homeolog gene-pairs show a significantly smaller difference to non-R-NF homeolog gene-pairs. Significance levels are indicated by asterisks above bars. * indicates $p<0.05$; ** indicates $p<0.01$; *** indicates $p<0.001$. *n.s.*- not significant. Error bars are standard errors.

those R-NF in just the husk (Figure 3.3B). Taken together these data reveal that R-NF homeolog gene-pairs experience stronger purifying selection and have evolutionary rates that are more similar to each other than non-R-NF gene-pairs. Thus, in non-R-NF homeolog gene-pairs one member of the gene-pair is more likely to experience reduced purifying selection relative to its homeolog pair.

3.2.3.2. Dominant homeolog genes exhibit stronger purifying selection than their non-dominant gene-pair

To investigate whether subgenome expression dominance impacts on selection, the K_a/K_s ratio of the homeolog of each non-R-NF gene-pair that dominates expression was compared to its non-dominant partner. This revealed that non-dominant homeologs are on average under significantly reduced purifying selection relative to their dominant counterparts. This observation is consistent for dominant and non-dominant homeolog gene-pairs at all stages of leaf development in both foliar and husk leaves (Figure 3.4) (although there is significant overlap at all of these stages). Whilst it is known that, broadly speaking, low-expressed genes evolve more rapidly than high-expressed genes (Zhang and Li 2004; Wright et al. 2004; Ingvarsson 2006; Wang et al. 2011; Hoffmann et al. 2016), this has not previously been shown to apply to such recent duplicates. Non-dominant homeolog genes are therefore a potential source of protein sequence novelty, and are thus strong candidates for coding sequence neofunctionalisation.

3.2.4. Many high-priority candidate Kranz regulators have homeolog genes, which show evidence of expression divergence

This analysis reveals that expression divergence of homeolog gene-pairs is a significant phenomenon in maize in the context of early leaf development, with 16% of expressed homeolog gene-pairs exhibiting divergence in either foliar or husk leaves, or both. Given that this analysis considered only broad expression changes

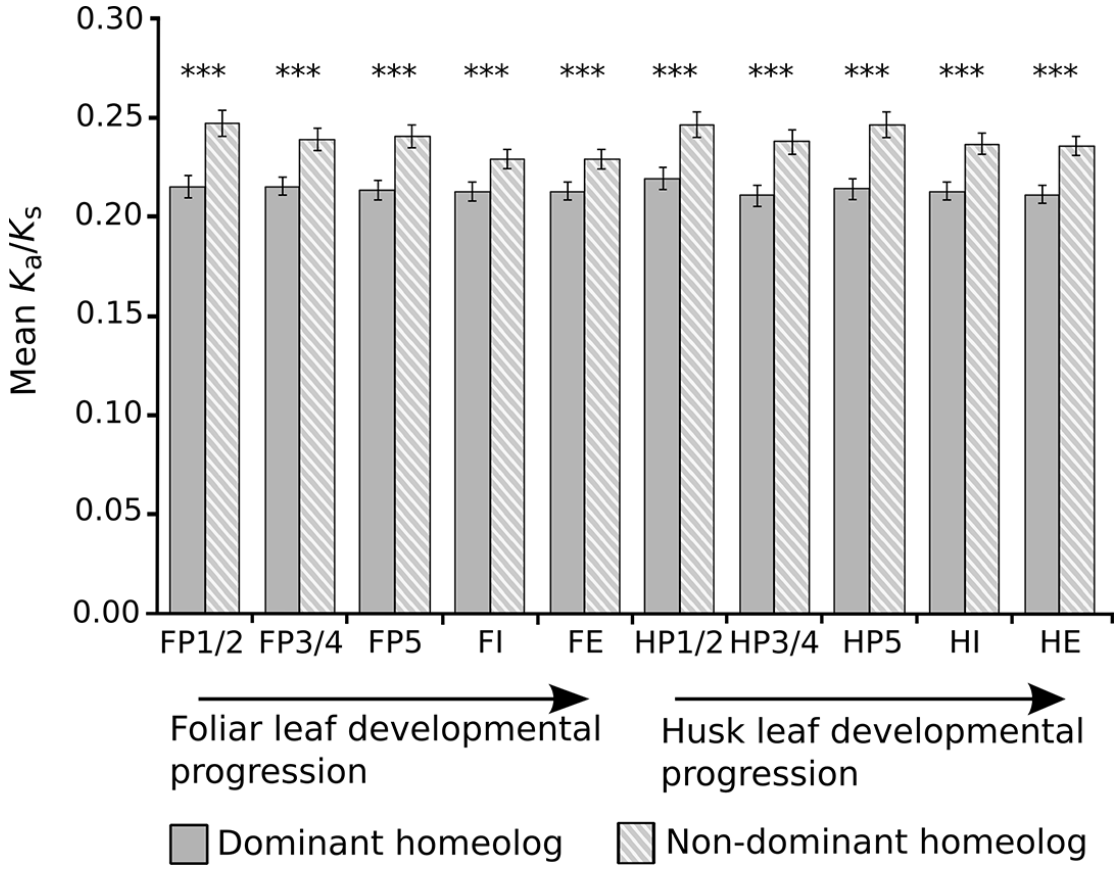


Figure 3.4 Average *Ka/Ks* ratios for non-R-NF genes where one gene dominates expression. K_a/K_s ratios were compared for those non-R-NF homeolog gene-pairs that exhibit subgenome dominance. The mean K_a/K_s ratio of non-dominant homeolog genes (those expressed at significantly lower levels than their homeolog pair) was significantly higher than the dominant homeolog (those expressed at significantly higher levels than their homeolog pair). P-values were calculated using Wilcoxon signed-rank tests. Significance levels are indicated by asterisks above bars, *** indicates $p<0.001$. Error bars are standard errors.

across three distinct developmental stages, and in only one particular context (early leaf development), it is likely that this represents a conservative underestimate of the significance of expression divergence in maize. Conversely, however, 84% of expressed homeolog gene-pairs have not diverged in expression in this context, and this has implications for our interpretation of the outputs of candidate gene approaches from maize RNA-seq datasets, which will form the basis of chapters 4 and 5 of this thesis.

The RNA-seq dataset on which this work is based, was used previously to identify 283 candidate positive regulators of Kranz anatomy (Wang et al. 2013b). Manual annotation of these genes, including analysis of known *Arabidopsis* orthologs and published cell-specific M and BS transcriptome datasets, revealed 12 high-priority model genes (Fouracre et al. 2014). However, neither the bioinformatic filtration process nor the manual annotation used to identify these candidate genes took into account possible homeolog gene-pairs, and treated all genes separately. Given that the majority of homeolog gene-pairs have not diverged in expression in the context of foliar and husk-leaf development, and on average homeolog gene-pairs are under high purifying selection to maintain the coding sequence of the encoded protein, understanding the behaviour of any duplicated pairs of these model genes is of critical importance when compiling a testable model for Kranz anatomy development in maize.

Of the 12 model genes identified previously by Fouracre *et al.* (2014), a number exist as homeolog gene-pairs in the Schnable *et al* (2011) dataset (Table 3.2). Two of the genes (*ZmDLK1* and *ZmDLK1h*) were in fact a homeolog gene-pair where both genes were included in the list. Of the remaining ten genes, five have homeolog gene-pairs included in the dataset generated by Schnable *et al.* (2011) (*ZmSCR1, ZmSHR2, ZmRVN1, ZmDOT1* and *ZmSPL1*). In addition, *ZmNKD1* has a duplicated pair (previously identified as *ZmNKD2* (Yi et al. 2015)), not included in the list of

Gene name	Gene accession	Subgenome	Mean expression FP1/2-FP5	Mean expression HP1/2-HP5	Foliar expression cluster	Husk expression cluster	*Ka/Ks* ratio relative to sorghum ortholog	R-NF category
ZmSCR1	GRMZM2G131516	1	67.01	41.01	FN	HN	0.04711	Non-R-NF
ZmSCR1h	GRMZM2G015080	2	16.10	11.60	FN	HN	0.06749	Non-R-NF
ZmDLK1	GRMZM2G114998	2	13.78	8.29	FA2	HN	0.17448	Non-R-NF
ZmDLK1h	GRMZM2G171852	1	14.00	8.82	FA2	HN	0.08882	Non-R-NF
ZmSHR2	GRMZM2G172657	1	27.32	12.94	FA2	HN	0.05367	R-NF-F
ZmSHR2h	GRMZM2G019060	2	1.41	0.27	FN	HN	0.08587	R-NF-F
ZmRVN1	GRMZM2G143723	1	13.76	3.59	FA2	HN	0.06855	R-NF-F
ZmRVN1h	GRMZM2G046290	2	1.04	1.17	FN	HN	0.33893	R-NF-F
ZmDOT1	GRMZM2G150011	1	34.02	6.35	FA2	HN	0.11570	R-NF-F
ZmDOT1h	GRMZM2G134998	2	4.37	2.36	FN	HN	0.18408	R-NF-F
ZmNKD1	GRMZM2G129261	n/a	44.24	21.32	FA2	HN	0.08870	R-NF-F
ZmNKD2	GRMZM5G884137	n/a	8.44	2.41	FN	HN	0.15099	R-NF-F
ZmSPL1	GRMZM2G045883	1	6.17	2.14	FA2	HN	0.15270	R-NF F&H
ZmSPL1h	GRMZM2G008691	2	2.77	10.87	FN	HD2	0.27810	R-NF F&H

Table 3.2 Subset of candidate Kranz regulators with duplicate genes. Genes from Fouracre *et al.* (2014) included in the Schnable *et al.* (2011) homeolog dataset, or, in the case of *ZmNKD1*, found through analysis here (see Chapter 5). Mean relative expression was averaged across all three early leaf development stages in foliar and husk (P1/2, P3/4, P5).

homeolog gene-pairs identified by Schnable *et al.* (2011). Therefore, only 4 of the 12 model genes identified by Fouracre *et al.* (2014) exist in a non-duplicated state in the maize genome.

In most cases it appears that the homeolog duplicates did not emerge as candidate Kranz regulators in their own right due to either significantly lower expression levels, or because the expression profiles did not match the expectations of how a Kranz regulator should behave. For example, *ZmNKD1, ZmSHR2, ZmRVN1, ZmDOT1* and *ZmSPL1* were all identified in the FA2 cluster (a sub-cluster of genes with ascending expression during early foliar leaf development), whereas in all cases their homeolog pairs were found in the FN cluster and were expressed at lower levels (Table 3.2).

This indicates that there has been divergence of these gene-pairs in terms of expression level and expression profile. However, given that all are under strong purifying selection it is likely that there remains redundancy in terms of protein function (Table 3.2).

3.3. Discussion

3.3.1. Expression divergence is a prominent feature of maize homeolog gene-pairs

Duplication followed by functional divergence has long been proposed as a source of evolutionary novelty in genome evolution (Ohno 1970). Moreover, whole genome duplications are a principal theme of plant evolution and have occurred in the evolution of all angiosperm lineages (Adams and Wendel 2005; Soltis et al. 2009). In maize, a recent whole genome duplication (5-12 MYA) followed by multiple gene losses has sculpted the contemporary maize genome such that duplicated homeolog gene-pairs comprise >20% of all genes (Swigonova et al. 2004a; Schnable et al. 2011). In this work maize was used as a model system to analyse the evolutionary paths taken by homeolog gene-pairs following whole genome duplication. This work

used RNA-sequencing to investigate R-NF on a whole genome scale in maize for the first time. This revealed that R-NF is an important feature of maize evolution at the gene, genome, tissue and developmental levels. RNA sequencing allowed characterisation of R-NF at a genome-wide scale, revealing that ~13% of homeolog gene-pairs in the maize genome have undergone R-NF. As this estimate was generated only in the context of leaf development, and considered broad changes in expression profile across three developmental stages, it likely represents an underestimation of the extent of R-NF in maize genome evolution. Presumably if more tissues and developmental contexts were considered, the percentage of R-NF genes would increase. However, it would be interesting to see whether the same cohort of genes that have been identified as R-NF here are also R-NF in other contexts.

Genes that have diverged in expression profile were identified as R-NF rather than regulatory subfunctionalised (R-SF). This was because the majority of divergent expression profiles were asymmetric, meaning that the profiles between foliar and husk did not mirror each other. Furthermore, R-SF is more like to involve changes in expression level such that the ancestral expression of a gene is now compartmentalised in different tissues between the duplicate genes. This analysis focused on changes in expression profile rather than changes in absolute expression level, and as such it is difficult to detect examples of R-SF. For the same reason, dosage subfunctionalisation, which has been proposed to play a large role in the maintenance of new duplicated genes, was not considered (Hoffmann et al. 2016). In practice, the likely overlap between R-NF and S-NF, means it is often difficult to distinguish these two evolutionary fates. Furthermore, the terms are often defined and used differently in different studies, which further complicates efforts to make inter-species comparisons (Freeling et al. 2015).

Given that the maize whole genome duplication is most likely to have an allotetraploid origin (Gaut and Doebley 1997; Swigonova et al. 2004a, 2004b), it is possible that some of the divergence in expression classified as regulatory neofunctionalisation (R-NF), preceded the whole genome duplication in the two progenitors of the contemporary maize genome. In this context it is pertinent to note that differential expression of some genes has been found between diploid cotton genomes (pre-allotetraploidisation) (Rapp et al. 2009; Flagel and Wendel 2010; Yoo et al. 2013). However, it is more parsimonious to assume that the majority of genes had the same expression profile at the point of genome duplication, particularly when considering broad changes in expression across multiple developmental stages rather than absolute expression levels at a specific point. In support of this assumption, 84% of gene-pairs in this dataset are non-R-NF, and whilst some of these pairs may represent convergence, it is more likely that 16% have changed expression than 84% have converged. Moreover, regardless of when divergence occurred, distinct selection signatures and patterns of subgenome dominance are associated with homeolog-gene-pairs that have divergent expression profiles.

3.3.2. Expression dominance is a dynamic feature of homeolog gene-pairs in maize

Expression dominance bias towards the less fractionated subgenome has previously been observed in both maize and *Brassica rapa* (Schnable et al. 2011; Cheng et al. 2012) but is not observed in more ancient plant tetraploids (Blanc and Wolfe 2004a; Li et al. 2006). Similarly, recent work in cotton has demonstrated that expression dominance between homeologs occurs immediately following tetraploidisation and increases over time (Yoo et al. 2013; Adams and Wendel 2013). Many previous studies concerning expression dominance have employed a 2-fold cut-off to define expression dominance (Schnable et al. 2011; Pophaly and Tellier 2015), however, in this analysis differential expression testing was used to avoid mis-identifying low-

expressed genes as having expression dominance toward one subgenome. Interestingly, when defined using statistically significant differential expression testing, subgenome expression dominance is a dynamic feature of homeolog gene-pairs in maize. Those gene-pairs that have not undergone R-NF show no consistent expression bias toward either subgenome, and the proportion of gene-pairs exhibiting either subgenome-1 (less fractionated) or subgenome-2 (more fractionated) dominance increases over the course of leaf development. In contrast, subgenome-1 expression dominance bias is consistently observed in R-NF homeolog gene-pairs, with no apparent developmental progression in dominance. This indicates that in maize, the previously reported expression dominance bias towards the less fractionated subgenome occurs alongside R-NF of homeolog gene-pairs (Schnable et al. 2011).

3.3.3. Expression divergence makes protein divergence less likely

In this analysis non-R-NF homeolog genes are under significantly reduced purifying selection, and of those non-R-NF homeolog gene-pairs that exhibit subgenome dominance, the non-dominant homeolog is consistently found to be under reduced purifying selection relative to the dominant homeolog. In light of this, it is interesting to note that only 10 homeolog genes show evidence of being under positive selection, and 9 of these are non-R-NF (Appendix 7). The fact that so few genes are under positive selection supports the previous finding that positive selection on coding regions does not play a significant role in post-genome duplication C-NF (Duarte et al. 2006). Whilst positive selection has been shown to drive neofunctionalisation in some duplicates (Zhang et al. 1998; Shiu et al. 2006; Conant and Wolfe 2008), this analysis suggests that following whole genome duplication, reduced purifying selection is a more likely evolutionary mechanism for altering protein function in plants. In support of this finding, relaxed purifying selection has previously been associated with functional divergence of duplicate genes within both

plant and non-plant systems (Kondrashov et al. 2002; Zhang 2003; Flagel and Wendel 2009; Shan et al. 2009). In addition, it has previously been shown that reduced purifying selection is more important than positive selection in the maintenance of duplicated genes in the human genome (Nguyen et al. 2008). As expression levels were not considered directly in this analysis, it is possible that the reduction in purifying selection is associated with homeolog gene-pairs where one gene has significantly reduced expression. An inverse correlation between expression level and protein evolution rate is well established (Zhang and Li 2004; Wright et al. 2004; Ingvarsson 2006; Wang et al. 2011; Hoffmann et al. 2016), and in this study the non-dominant partners of non-R-NF gene-pairs where one gene dominates expression were found to be under reduced purifying selection. However, this does not account for the significant difference in purifying selection found between non-R-NF and R-NF genes.

Collectively, these results point to a plausible model for gene evolution following whole-genome duplication in maize whereby the emergence of novel protein function occurs through the relaxation of purifying selection, and is more likely to occur in genes that have not undergone R-NF. It remains possible that C-NF could follow R-NF, but these data indicate that C-NF is more likely to occur in the absence of R-NF. Of course, to fully understand both regulatory and protein function divergence of duplicated genes, in-depth functional analysis is necessary on a case-by-case basis. However, for the first time this analysis gives an overview of the extent of regulatory divergence of homeolog gene-pairs on a whole genome scale in maize.

3.3.4. Understanding homeolog divergence is critical for understanding transcriptional regulation of development in maize

Maize has long been established as a critical model organism for a range of research areas, including plant domestication, genome evolution, developmental physiology,

epigenetics, and C_4 photosynthesis (Strable and Scanlon 2009). Understanding the genetic pathways that underpin maize development has been enhanced by the advent of RNA-sequencing approaches, which allow gene expression in different contexts to be determined at a whole genome scale. The grass lineage is known to have undergone at least two whole genome duplications at the base of the clade, in addition to a further duplication at the base of the monocots (Paterson et al. 2004; Tang et al. 2010; D'Hont et al. 2012; Jiao et al. 2014; McKain et al. 2016). This has resulted in extensive potential genetic redundancy within individual grass species. In maize, the situation is further exacerbated by the more recent whole duplication that occurred subsequent to the divergence from the close relative sorghum (Swigonova et al. 2004a). As such, there is the potential for high levels of genetic redundancy in maize, and this must be considered when trying to understand gene-function.

In the case of C_4 photosynthesis, a number of studies have recently taken a candidate gene approach to identify putative regulators underpinning C_4 leaf development (Li et al. 2010; Wang et al. 2013b; Liu et al. 2013; Tausta et al. 2014). One of these studies has identified a subset of twelve high-priority candidate regulators of Kranz anatomy development, based on expression profiles during early foliar and husk leaf development (Wang et al. 2013b; Fouracre et al. 2014). However, in this analysis it was apparent that considering these twelve genes in isolation is likely to be problematic because only four of the twelve exist as single copy in the maize genome. The remaining 8 have maize-specific duplicates. There was only one case (*ZmDLK1* and *ZmDLK1h*) where both genes were independently identified as candidate Kranz regulators, meaning that homeologs of the other 6 need to be added to the candidate list. Of these 7 gene-pairs, there is evidence to support R-NF of 5 of the 7 gene-pairs in the foliar leaf. However, all 14 genes were under high purifying selection indicating that it is most parsimonious to assume that

the proteins function redundantly. Thus, considering these gene-pairs in combination is necessary for loss-of-function analysis in maize.

Chapter 4: The role of ZmSCARECROW1 (ZmSCR1) and ZmSHORTROOT2 (ZmSHR2) in the development of Kranz anatomy

4.1. Introduction

The *Arabidopsis* GRAS transcription factors SCARECROW (AtSCR) and SHORTROOT (AtSHR) are two of the best-studied developmental regulators in plants. Both genes were identified due to their characteristic mutant phenotypes, which result in impaired ground-tissue formation in the root (Benfey et al. 1993; Scheres et al. 1995; Laurenzio et al. 1996). Since then, a number of studies have elucidated an elegant mechanism underpinning the transcriptional regulation of cellular patterning. *AtSHR* is transcribed and translated in the stele (the root vasculature), and then moves into the adjacent cell layer where it regulates the expression of a suite of target genes (Nakajima et al. 2001; Cui et al. 2007). Importantly, AtSHR movement beyond the cell layer adjacent to the site of expression is prevented by AtSCR, which binds to AtSHR and sequesters it to the nucleus (Levesque et al. 2006; Cui et al. 2007). The transcriptional changes induced by AtSHR and AtSCR in the cell layer adjacent to the stele promote the differentiation of the endodermis, whereas the absence of AtSCR and AtSHR in the adjacent cell layer leads to the formation of cortex. Although this mechanism emerged from work in *Arabidopsis*, the available evidence points to SCR and SHR function being conserved in monocots; OsSHR was found to interact with AtSCR *in vitro* (Cui et al. 2007), and *ZmSCR1* complemented the *Atscr* mutant phenotype (Lim et al. 2005).

More recently, both AtSHR and AtSCR have been implicated in the control of leaf development. The leaves of *Atscr* and *Atshr* mutants exhibit impaired BS formation

(Cui et al. 2014), and transcripts of both genes have been detected at high levels in developing maize leaves (Wang et al. 2013b). Furthermore, an RNA-seq dataset designed to provide insight into the developmental regulation of Kranz anatomy in maize, identified one *AtSCR* ortholog (*ZmSCR1),* and two *AtSHR* orthologs (*ZmSHR1* and *ZmSHR2)* as candidate Kranz regulators (Wang et al. 2013b; Fouracre et al. 2014). A hypothesised role for these genes in Kranz development has been supported by expression analysis. *ZmSCR1* is preferentially expressed in the M cells of developing maize primordia (Olga Sedelnikova, personal communication), although RNA-seq datasets from separated BS and M cell populations isolated from fully expanded maize leaves, indicate that it may be BS specific at later stages (Li et al. 2010; Chang et al. 2012). In contrast, *ZmSHR1* appears to be expressed in both the developing vasculature and BS cells (Li et al. 2010; Chang et al. 2012), whereas *ZmSHR2* is expressed only in the early vasculature (Olga Sedelnikova, personal communication). Interestingly, both *Zmscr1* and *Zmshr1* mutants were found to have impaired Kranz formation, with both exhibiting a number of defects in vein patterning (Slewinski et al. 2012, 2014). In addition, *Zmscr1* mutant leaves form with significantly higher occurrences of extra BS cell formation and merged intermediate veins (Slewinski et al. 2012). In contrast, *Zmshr1* plants displayed increased occurrences of extra M cells in between BS cells (Slewinski et al. 2014). *ZmSCR1* and *ZmSHR1* are thus the first established developmental regulators of Kranz cell patterning, and are of considerable interest to attempts to engineer C_4 rice.

There remain a number of issues regarding the role of ZmSCR1 and ZmSHR1/2 in the development of Kranz anatomy. For example, the *Zmscr1* leaf phenotype appears relatively subtle, and the majority of vascular bundles appear normal (Slewinski et al. 2012). Although the *Zmshr1* phenotype appears more severe, no quantification was provided to support this observation (Slewinski et al. 2014). These factors make it difficult to assess the importance of *ZmSCR1* and *ZmSHR1* in the

regulation of Kranz development, and there has not been, as yet, any functional validation of the proposed role for *ZmSHR2*. Furthermore, the finding that *Atscr* and *Atshr* mutants have impaired BS differentiation, raises the possibility that SCR and SHR may act generally across C_3 and C_4 species to regulate leaf patterning, rather than acting as Kranz-specific regulators *per se*. Finally, both *ZmSCR1* and *ZmSHR2* have homeolog gene-pairs in maize (see Chapter 3), although the relationship between these genes and their grass and *Arabidopsis* orthologs is not clear. These duplicated *ZmSCR* and *ZmSHR* genes may either function redundantly to regulate Kranz development, or have functionally diversified.

The first aim of this chapter was to validate the role of *ZmSCR1* and its homeolog pair *ZmSCR1h* in the development of Kranz anatomy. This was achieved through the generation and phenotypic characterisation of a *Zmscr1-1; Zmscr1h-1* double mutant, followed by analysis of a number of rice lines that constitutively overexpress *ZmSCR1*. The second aim was to determine whether *ZmSHR2* functions in Kranz development. This was achieved through analysis of a loss-of-function *Zmshr2-1* mutant, and through generation and analysis of a *Zmscr1-1; Zmshr2-1* double mutant.

4.2. Results

4.2.1. Both *ZmSCR1* and *ZmSHR2* have duplicated gene-pairs

To determine the phylogenetic relationships between *ZmSCR1*, *ZmSHR2* and their grass and *Arabidopsis* orthologs, a maximum likelihood phylogenetic tree was constructed of the GRAS transcription factors. Both *ZmSCR1* and *ZmSHR2* were previously found to have homeolog gene-pairs (Schnable et al. 2011), and this relationship was confirmed here (Figure 4.1A, Appendix 8). The *SHR* clade contains one *Arabidopsis* (*AtSHR*) gene, however, a duplication occurred at the base of the grasses resulting in two distinct clades of monocot *SHR* genes, all of which are

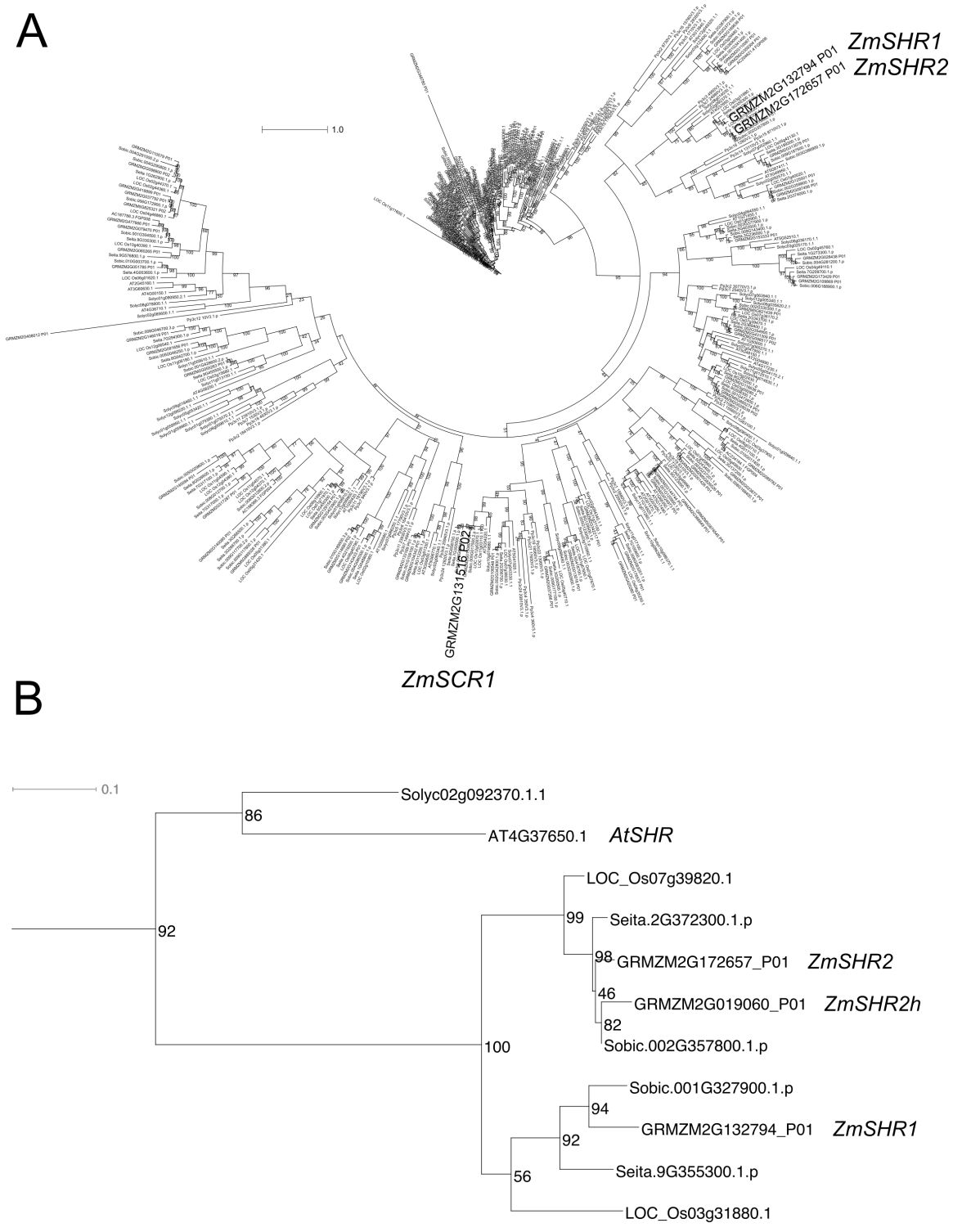


Figure 4.1 Maximum likelihood phylogenetic tree of the GRAS transcription factors (A and B). Part (C) and the figure legend are on the next page.

C

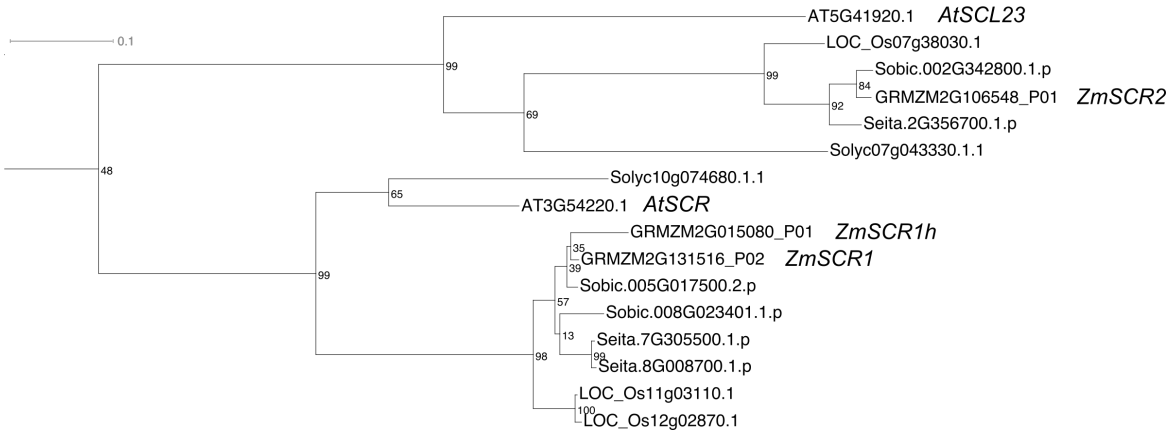


Figure 4.1 Maximum likelihood phylogenetic tree of the GRAS transcription factors. A) Complete phylogenetic tree, with the position of *ZmSCR1, ZmSHR1* and *ZmSHR2* highlighted. **B)** The *SHR* clade of the tree. **C)** The *SCR* and *SCL23* clades of the tree. GRAS sequences were acquired using the *ZmSCR1* and *ZmSHR2* protein sequences as a query for a BLASTp search (evalue = 1e-3) against the proteomes of all seven species. The retrieved sequences were aligned using MergeAlign, and the alignment subsequently trimmed using trimAL. A maximum likelihood tree was constructed using RaxML. Scale bars indicate the number of substitutions per site. Bootstrap values are displayed on each branch. Maize, sorghum, *Setaria italica,* rice, *Arabidopsis*, tomato and *Physcomitrella patens* gene accessions begin GRM, Sobic, Seita, LOC_Os, AT, Solyc and Ppat respectively. An electronic version of this tree is included in Appendix 8.

orthologous to *AtSHR* (Figure 4.1B). One of these clades contains *ZmSHR1* and three single-copy orthologs in sorghum, *Setaria italica,* and rice (Figure 4.1B). The *ZmSHR2* clade contains single copy *SHR2* genes for sorghum, *Setaria italica,* and rice, but two distinct genes in maize (*ZmSHR2* and *ZmSHR2h*) (Figure 4.1B). In the *SCR* clade, there is also only one *Arabidopsis SCR* gene (*AtSCR*), but two *SCR* orthologs in each grass species (Figure 4.1C). However, in contrast to *SHR*, there are not two distinct clades of *SCR* genes in the grass species included in this analysis. Instead, in both *Setaria italica* and rice there have been independent *SCR* duplications subsequent to the divergence of each species. The sorghum genome also encodes two *SCR* genes, however, the precise orthology of these genes is less clear, and a more precise manual alignment of the sequences prior to phylogenetic tree construction might be necessary to resolve these relationships. The maize *SCR* duplicates are believed (based on synteny) to have resulted from the recent maize whole genome duplication (Schnable et al. 2011), and hence were named *ZmSCR1* and *ZmSCR1h*. The origins of the independent duplications of *SCR* in the other grass species are less clear. Regardless of the precise origins, however, the presence of these monocot-specific *SCR* and *SHR* duplicates means that the SCR/SHR pathway is likely to be more complex in monocots than in *Arabidopsis*. The impact of this added complexity, has not yet been considered in relation to Kranz patterning. In particular, previous work that identified the role of *ZmSCR1* in Kranz development did not consider potential redundancy or functional diversification of *ZmSCR1h* (Slewinski et al. 2012). Furthermore, the impact of two additional maize *SHR* genes was not considered in the previous analysis of *ZmSHR1* function (Slewinski et al. 2014).

4.2.2. Transposon insertion mutants were identified for three of the four maize *ZmSCR1/ZmSHR2* genes

To determine the role of *ZmSCR1* and *ZmSHR2* function in Kranz patterning, MaizeGDB was searched for available transposon insertion mutants (Lawrence et al. 2004, 2008). *Mutator* transposon insertions from the UniformMu collection were identified for *ZmSCR1*, *ZmSCR1h* and *ZmSHR2*, with 3 additional *Mutator* transposons from the Mu-Illumina collection identified in *ZmSCR1* (Appendix 2). No insertions were identified in or near the *ZmSHR2h* sequence. Where more than one insertion was available for a line, priority for phenotypic analysis and crossing was determined by the following criteria: 1) the position of the insertion, with insertions into the coding sequence (CDS) given higher priority; 2) the genetic background, with UniformMu insertions preferred given reported difficulties associated with the Mu-Illumina background; and 3) the number of additional mapped transposon insertions in the line to be analysed, with priority given to lines with fewer additional insertions. Using these criteria one insertion line for each gene was selected and named for characterisation in this thesis (Figure 4.2) (Appendix 2). Seed for each insertion were acquired and the presence of the insertion validated by PCR, with transposon specific primers enabling detection of the insertion and flanking primers enabling detection of the WT fragment (Figure 4.3A-D,F,G). The transposon insertion in the *Zmscr1-1* allele has previously been shown to be retained in the transcribed *ZmSCR1* mRNA (Slewinski et al. 2012). To confirm whether this was also true for the *Zmscr1h-1* and *Zmshr2-1* insertions, RNA was extracted from immature leaves or primordia (depending on expression in RNA-seq datasets) and PCR used to amplify the allele from the resultant cDNA (Figure 4.3E-G).

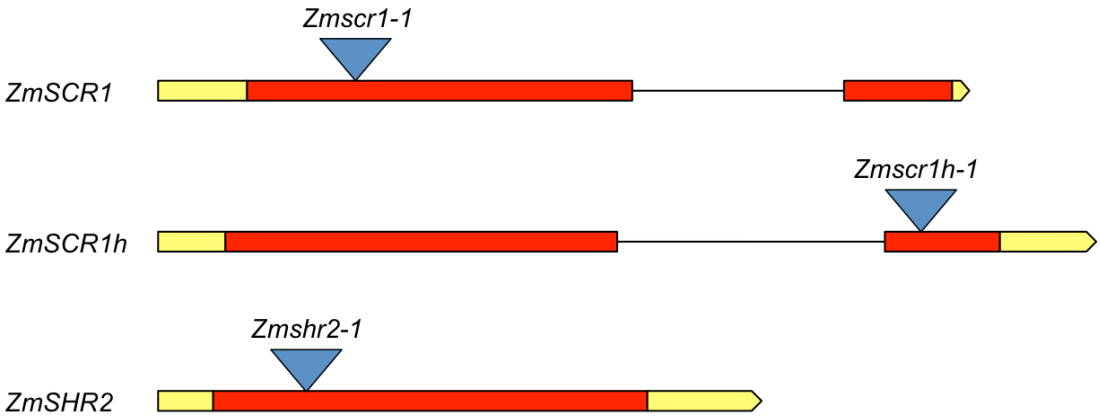


Figure 4.2 Schematic of transposon insertions into three GRAS transcription factors. Gene models are displayed for each transcription factor: in all cases the 5' end is displayed on the left, yellow regions refer to the UTRs, red to the CDS and black lines to introns. Blue triangles represent *Mutator* transposons (not to scale) with insertion sites acquired from the UniformMu project. Allele names are displayed above the transposon.

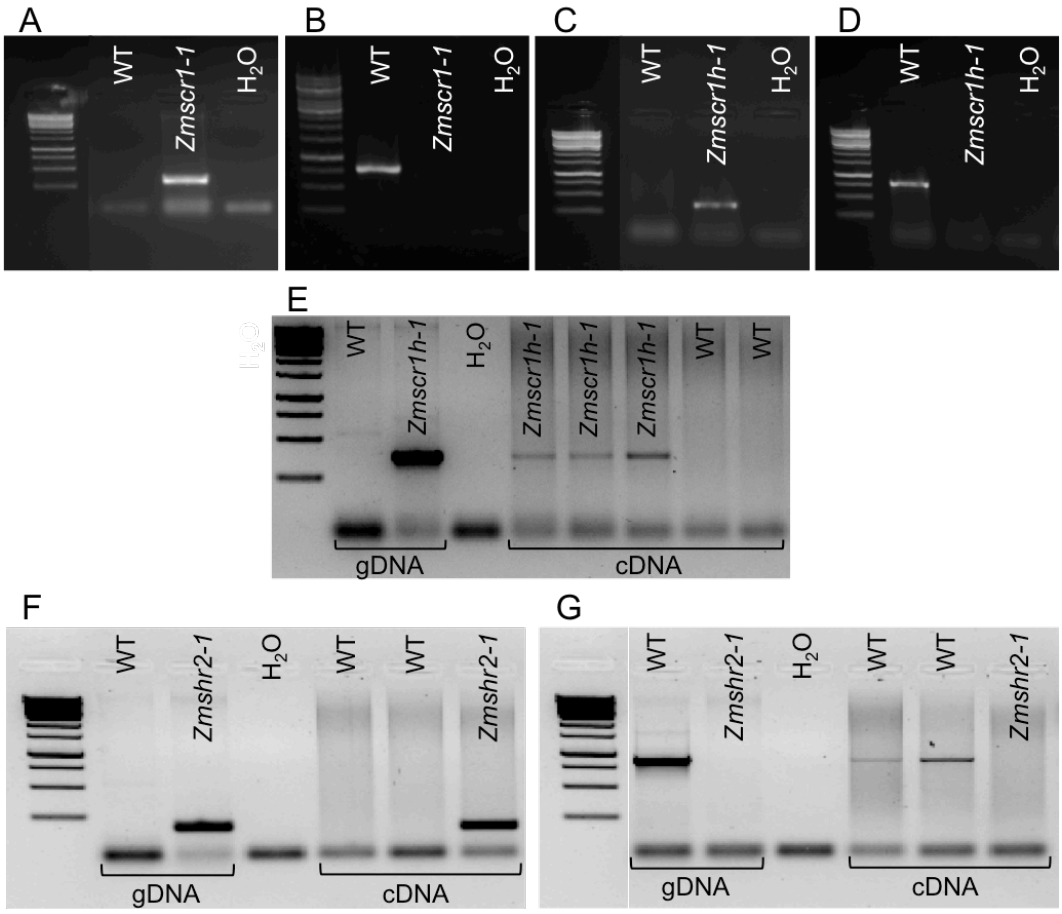


Figure 4.3 Gel images of genotype and expression analysis of GRAS *Mutator* transposon insertion alleles. Figure legend is on the next page.

Figure 4.3 Gel images of genotype and expression analysis of GRAS *Mutator* transposon insertion alleles. A,B) PCR genotyping of *Zmscr1-1* gDNA. Primers amplifying from the *Mutator* transposon to the flanking *ZmSCR1* sequence (A) and flanking the *Zmscr1-1* insertion site (B). **C,D)** PCR genotyping of *Zmscr1h-1* gDNA. Primers amplifying from the *Mutator* transposon to the flanking *ZmSCR1h* sequence (C) and flanking the *Zmscr1h-1* insertion site (D). In (B) and (D) extension times were insufficient to amplify across the transposon in the mutant. **E)** PCR validation of *Mutator* retention in *Zmscr1h-1* transcripts. Primers as in (C). gDNA controls are shown. cDNA was amplified from RNA isolated from immature foliar leaves. Three different *Zmscr1h-1* individuals and two different WT individuals are shown. **F,G)** PCR genotyping of *Zmshr2-1* and validation that the *Zmshr2-1* insertion is maintained in the resultant transcript. Primers amplifying from the *Mutator* transposon to the flanking *ZmSHR2* sequence (F) and flanking the *Zmshr2-1* insertion site (G). gDNA genotyping samples are shown. cDNA was amplified from RNA isolated from pooled leaf primordia samples. In (A-G) the smallest fragment of the DNA ladder is 250bp. The fourth smallest is 1kb.

4.2.3. The role of *ZmSCR1* and *ZmSCR1h* in maize

4.2.3.1. The maize *Zmscr1-1; Zmscr1h-1* double mutant exhibits a reduced growth phenotype not seen in either of the single mutants

To test the hypothesis that ZmSCR1h functions redundantly with ZmSCR1 in Kranz patterning a *Zmscr1-1; Zmscr1h-1* double mutant was generated. A line segregating for both insertions was used for phenotypic analysis, with both the single and double mutants compared to segregating UniformMu W22 WT plants from the *Zmscr1-1* line. The single *Zmscr1-1* mutant exhibited no obvious perturbations in overall plant morphology or growth, which is consistent with previous reports (Slewinski et al. 2012). This was also true of the *Zmscr1h-1* single mutant. However, the double mutant exhibited a severe reduction in growth that was obvious from the early stages of development (Figure 4.4). This reduction in growth was maintained throughout plant development, with mature *Zmscr1-1; Zmscr1h-1* plants significantly shorter and with reduced stem circumference than both WT and single mutant plants (Figure 4.5).

To determine whether the reduced growth in double mutants was due to a developmental delay, leaf counts were taken 35 days after planting, with the newest emerged leaf recorded for all plants. Across the WT and single mutant plants, in only 2 out of 13 individuals was leaf 7 rather than leaf 8 the most recent leaf to have emerged. In contrast, in 3 out of 5 of the double mutant plants, leaf 8 had not yet emerged, implying a very slight developmental delay. This slight delay was reflected in the flowering time of the double mutant, which in the greenhouse occurred roughly a week after the other lines. Attempts to self-pollinate the double mutant in the greenhouse were largely unsuccessful, with very few seed obtained. It is not clear if a better seed set could be achieved in the field.

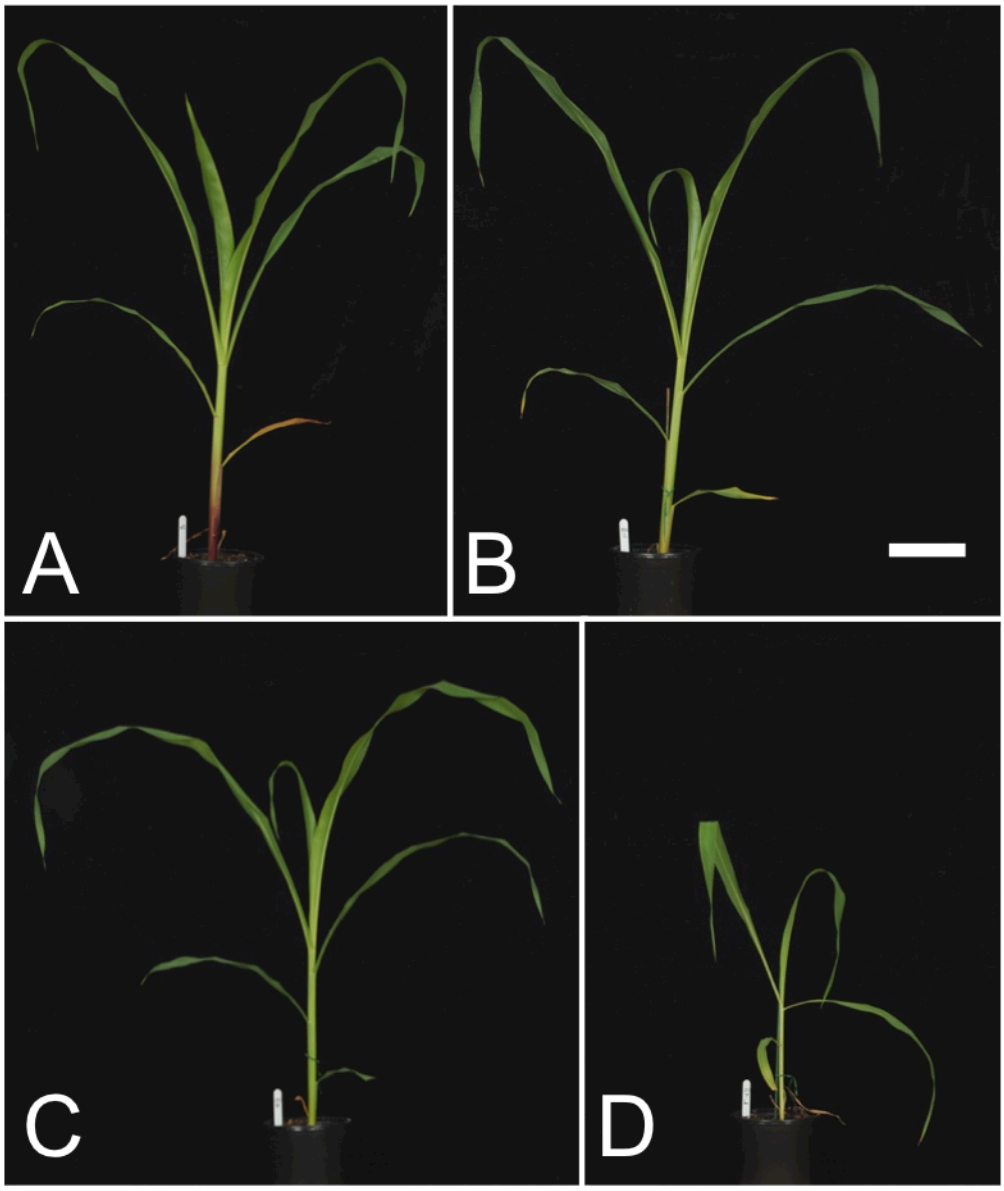


Figure 4.4 Whole plant phenotype of *Zmscr* mutants and WT plants. **A)** W22 WT, **B)** *Zmscr1-1*, **C)** *Zmscr1h-1* **D)** *Zmscr1-1; Zmscr1h-1*. White scale bar = 10cm. Scale is the same in (A-D). Images were taken 31 days after planting.

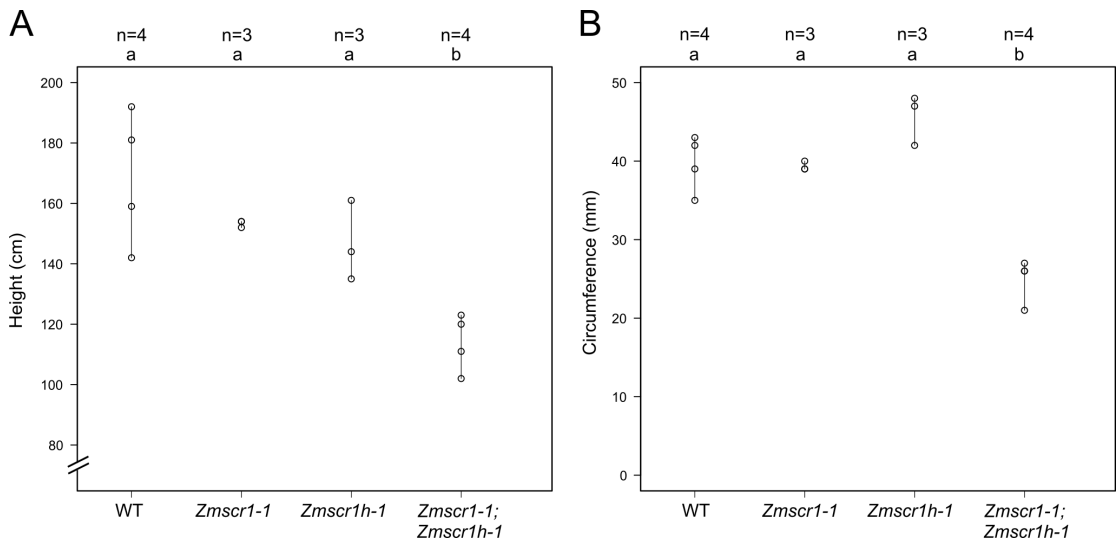


Figure 4.5 Lineplots showing morphological phenotypic perturbations of mature *Zmscr* mutants and WT plants. A) Height **B)** Base stem circumference. Measurements were taken at plant maturity after flowering. Height was measured from base to the tip of the tassel. Base stem circumference was measured above the first buttress root. Statistical groups are indicated by letters above the plot, with different letters indicating statistically different ($p<0.05$) groups, as indicated by a one-way ANOVA and Tukey HSD (honest significant difference) *post-hoc* test**.** Sample sizes are indicated by n= above significance groupings. Breaks in y-axis scale are indicated by a double diagonal line.

4.2.3.2. Endodermal cell specification and casparian strip formation in the root are disrupted in the double homozygous *Zmscr1-1; Zmscr1h-1* mutant

AtSCR is essential for correct ground-tissue patterning in the root (Laurenzio et al. 1996), and *Zmscr1* is expressed in the endodermis of maize roots (Lim et al. 2000). Furthermore, *ZmSCR1* is able to complement the *Atscr* root phenotype, indicating that *SCR* function in root patterning is likely conserved between monocots and dicots (Lim et al. 2005). It was previously reported that the endodermal cell layer is present in *Zmscr1* mutants, however, casparian strip formation was impaired (Slewinski et al. 2012). To test whether root development is further impaired in the *Zmscr1-1; Zmscr1h-1* double mutant, cross sections were taken of mature roots. Like WT roots, *Zmscr1-1* and *Zmscr1h-1* single mutants plants both exhibited an organised layer of endodermal cells, although reduced autofluorescence of these cells when imaged using UV illumination may indicate reduced casparian strip formation (Figure 4.6B and C). In contrast, roots of the *Zmscr1-1; Zmscr1h-1* double mutant did not form a obvious single endodermal layer (Figure 4.6D). To investigate this further, WT and double mutant root segments were fixed in 4% paraformaldehyde, wax embedded and sectioned. In WT maize roots stained with toluidine blue, the endodermal casparian strip surrounding the stele formed a distinctively stained band (Figure 4.6E,F). However, in the *Zmscr1-1; Zmscr1h-1* double mutant this band was absent, and no consistent ring of organised endodermal cells was identified (Figure 4.6G,H). It is unclear whether any of the remaining cell layers retain endodermal identity, however, the phenotype appears distinct to that exhibited by *Atscr* plants, where only a single layer of ground-tissue was formed with features of both the endodermis and cortex (Laurenzio et al. 1996).

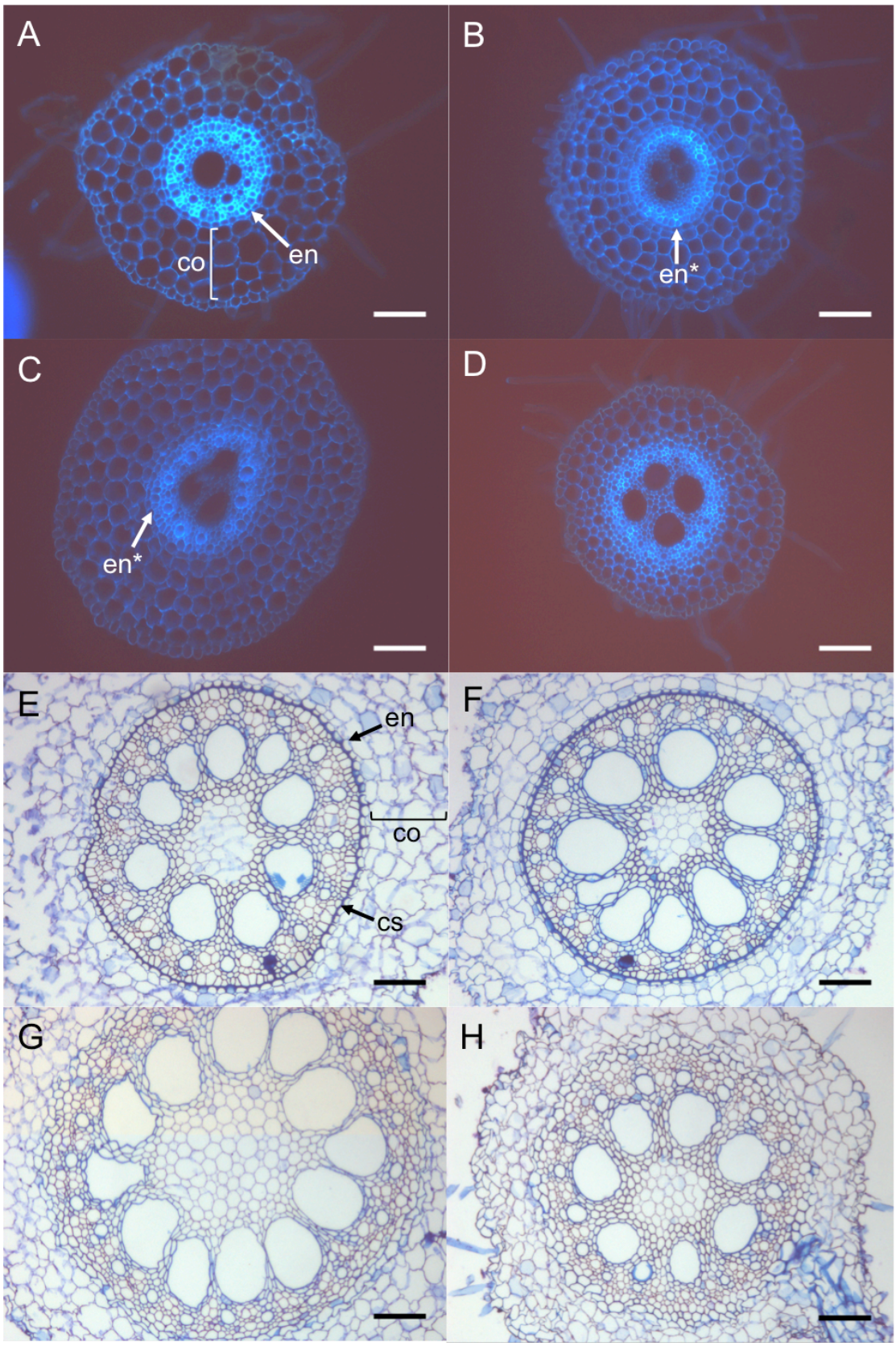


Figure 4.6 Transverse sections of *Zmscr* mutants and WT roots. Figure legend is on next page.

Figure 4.6 Transverse sections of *Zmscr* mutants and WT roots. A-D) Fresh sections imaged under UV illumination. Segregating WT in the *Zmscr1-1* line (A), *Zmscr1-1* (B), *Zmscr1h-1* (C), *Zmscr1-1; Zmscr1h-1* (D). **E-H**) Fixed 10µm sections stained with toluidine blue and imaged with brightfield microscopy. Segregating WT in the *Zmscr1-1* line (different individuals) (E,F), *Zmscr1-1; Zmscr1h-1* (different individuals) (G,H). Scale bars in (A-G) are 100µm. In (A) and (E) cortex is indicated by 'co' and endodermis by 'en'. In (B) and (C) 'en*' indicates the presumed endodermis with reduced autofluorescence. In (E) the casparian strip is indicated by 'cs'.

4.2.3.3. Delayed mid-rib formation causes a droopy leaf phenotype in *Zmscr1-1; Zmscr1h-1* double mutant plants

Zmscr1-1; Zmscr1h-1 double mutants exhibited a droopy leaf phenotype, particularly in emerging leaves (Figure 4.4). In the rice mutant *drooping leaf 1 (dl1*) drooped leaves resulted from the lack of a midrib, which normally forms acropetally along the proximal/distal axis of monocot leaves (Nagasawa et al. 2003; Yamaguchi et al. 2004). The *Zmscr1-1; Zmscr1h-1* double mutant still formed a visible mid-rib, however, it was hypothesised that the drooping leaf phenotype may be caused by delayed elongation of the midrib towards the tip of the leaf. To test this, samples of fully expanded leaf 11 were examined 71 days after planting, and the point along the leaf length that the midrib terminated was recorded. This revealed that the midrib occupied a significantly reduced proportion of the leaf length in the double mutant than in the WT or single mutants (Figure 4.7). As the midrib provides structural support for the leaf, it is likely that this reduced elongation in the double mutant caused the drooping phenotype. It also points to a possible role for both *ZmSCR1* and *ZmSCR1h* in vascular development.

4.2.3.4. Vascular patterning is perturbed in the *Zmscr1-1; Zmscr1h-1* double mutant

Slewinski *et al.* (2012) previously reported that mutations in *ZmSCR1* resulted in slight reductions in vein density and in the lateral vein number across the width of the leaf. Given that the *Zmscr1-1; Zmscr1h-1* double mutant displays a more severe morphological phenotype than either single *Zmscr* mutant, it was hypothesised that the double mutant might exhibit an even greater reduction in lateral vein number and vein density. However, the double mutant had significantly higher overall vein density and lateral vein density than either WT or single mutant plants (Figure 4.8, Figure 4.9A,B). In contrast to previous results, no difference in either overall vein density or lateral vein density was found between the WT and the single mutants. Given that

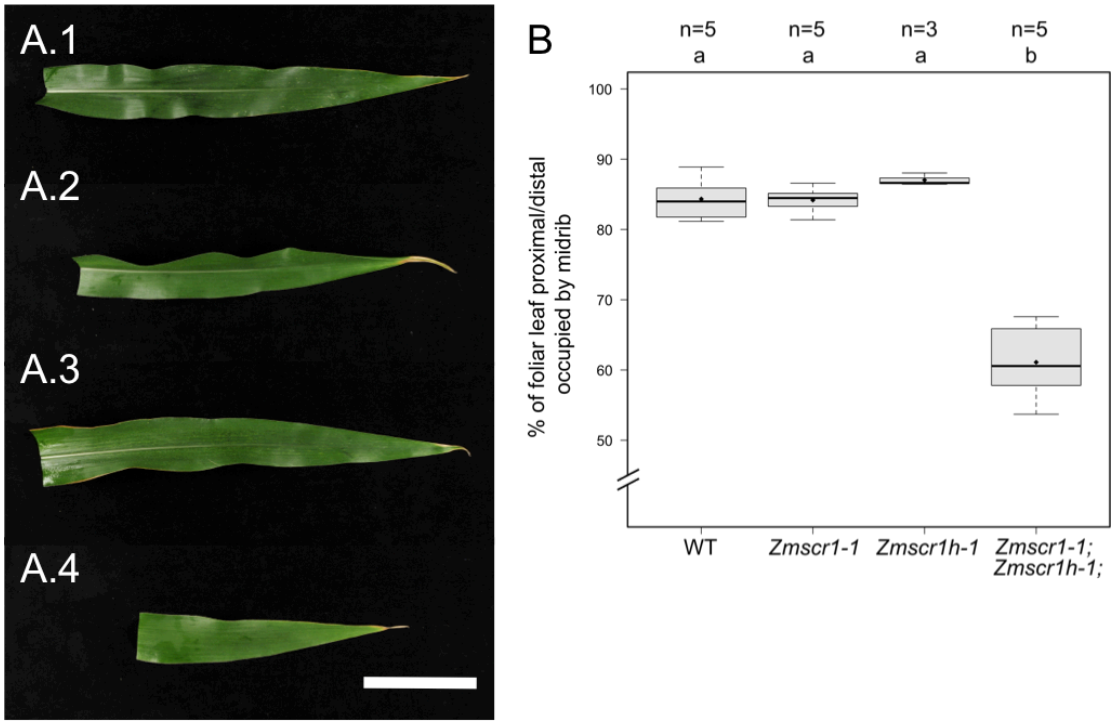


Figure 4.7 Midrib extension along the proximal/distal axis of *Zmscr* mutants and WT foliar leaves. A) Images of maize foliar fully expanded leaf 11 taken from 4cm beyond the midpoint of the proximal/distal axis to the tip of the leaf blade. Photos were taken 71 days after planting. Segregating WT in the *Zmscr1-1* line (A.1), *Zmscr1-1* (A.2), *Zmscr1h-1* (A.3) and *Zmscr1-1; Zmscr1h-1* (A.4). Scale bar = 10cm. **B)** Boxplots displaying length of the midrib as a % of the total length of the leaf, measured from ligule to tip of the blade along the proximal/distal axis. Statistical groups are indicated by letters above the plot, with different letters indicating statistically different ($p<0.05$) groups, as indicated by a one-way ANOVA and Tukey HSD *post-hoc* test**.** Means are indicated by black diamonds for each genotype. Sample sizes are indicated by n= above significance groupings. Breaks in y-axis scale are indicated by a double diagonal line.

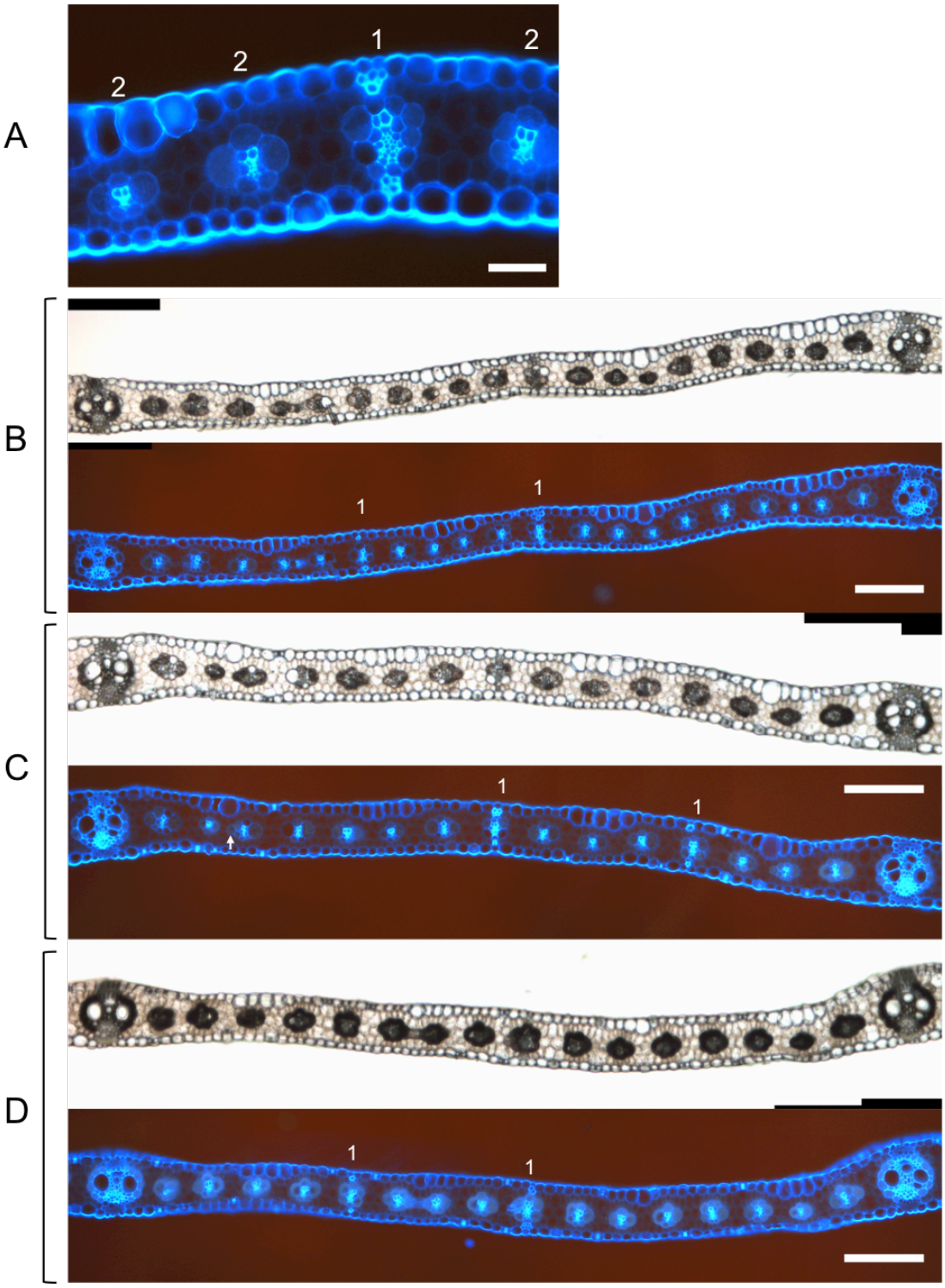


Figure 4.8 Fresh transverse sections of foliar leaves from *Zmscr* mutant and WT plants (A-D). Parts (E-G) are included on the next page. Figure legend is included on page 108.

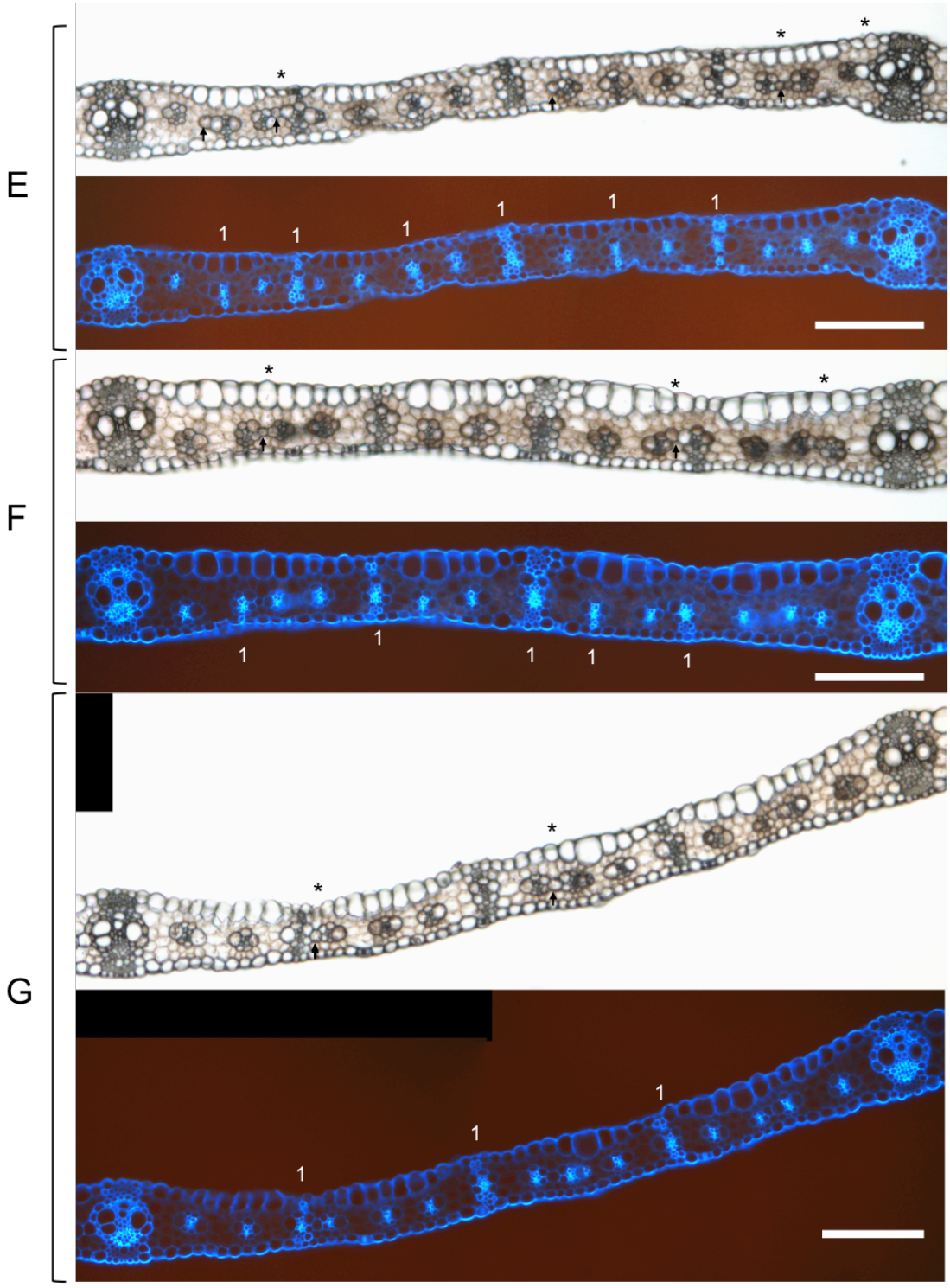


Figure 4.8 Fresh transverse sections of foliar leaves from *Zmscr* mutant and WT plants (E-G). Parts (A-D) are on the previous page. Figure legend is on the next page.

Figure 4.8 Fresh transverse sections of foliar leaves from *Zmscr* mutant and WT plants. A) Examples of rank-1 and rank-2 intermediate veins (indicated by number above vein) in WT leaves visualised using UV illumination. Rank-1 intermediates are defined by the presence of sclerenchyma (which fluoresce bright blue under UV) abaxially and/or adaxially to the vein. Scale bar = 50μm. **B-G)** Fresh transverse sections of fully expanded maize foliar leaf 5. Segregating WT in the *Zmscr1-1* line (B), *Zmscr1-1* (C), *Zmscr1h-1* (D)*, Zmscr1-1; Zmscr1h-1* (E-G). Each pair of images in (B-G) (brightfield and UV) are from a different plant, with images tiled to include all intermediate veins between two laterals. Images were taken (where possible) of the second set of laterals to the side of the midrib. Scale bars = 200μm. In (B-G) rank-1 intermediate veins are indicated by '1' above or below the vein on the UV image. In (C) and (E-G) arrows (white or black) indicate occurrences of extra BS cells outside of the normal ring of BS cells surrounding the vein. In (E-G), asterisks indicate the presence of a pair of veins with no adjoining M cells (V-BS_n-V).

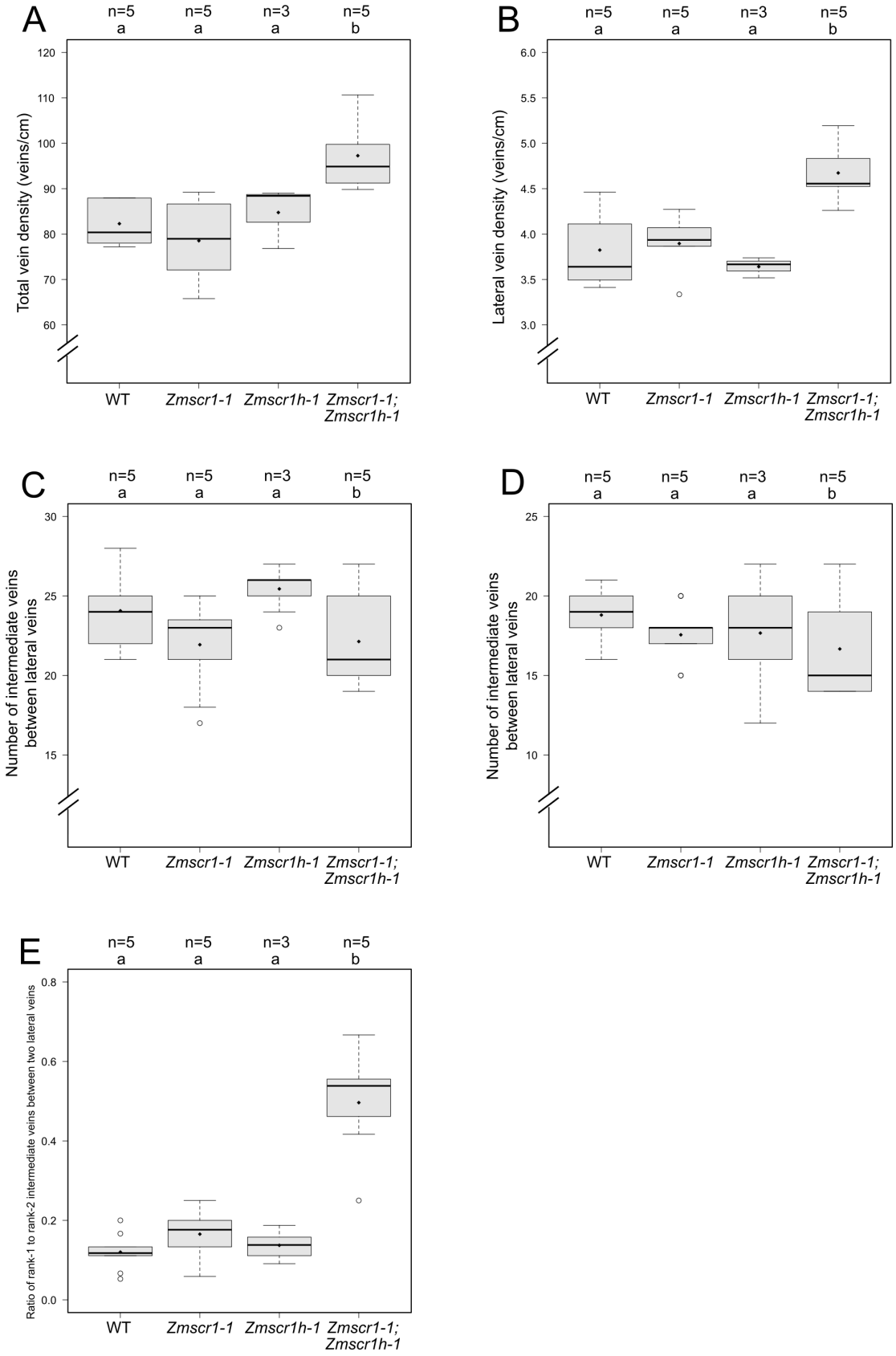


Figure 4.9 Quantification of vein-patterning traits in *Zmscr* and WT foliar leaves. Figure legend is included on the next page.

Figure 4.9 Quantification of vein-patterning traits in *Zmscr* and WT foliar leaves. Boxplots of **A)** Total vein density (veins/cm) **B)** Lateral vein density (lateral veins/cm) **C)** Number of intermediate veins between two lateral veins in leaf 11 **D)** Number of intermediate veins between two lateral veins in leaf 5 **E)** The ratio of rank-1 to rank-2 intermediate veins between two laterals. Measures taken from the midpoint of the proximal/distal axis of fully expanded leaf 11 (A-C) or leaf 5 (D,E). Counts were taken across the full width of the leaf (A,B), or between three (C), or two (D,E), pairs of lateral veins adjacent to the midvein. Statistical groups are indicated by letters above the plot, with different letters indicating statistically different ($p<0.05$) groups, as calculated by a Kruskal-Wallis rank sum test (A), one-way ANOVA and Tukey HSD *post-hoc* test (B,E) or a mixed effect poisson regression model and subsequent likelihood ratio calculation by comparison to a null model (C,D). WT labels refer to segregating WT plants in the *Zmscr1*-1 line. Biological replicates are indicated by n= above significance groupings. Means are indicated by black diamonds for each genotype, outliers are indicated by empty circles. Breaks in y-axis scale are indicated by a double diagonal line.

the previously reported difference between WT and *Zmscr1-1* in regard to lateral veins and vein density was small, it is possible that this discrepancy is the result of differing sampling points (Slewinski *et al.* (2012) sampled leaf 7, 9 and 11 and pooled the measurements). Despite the increased number of lateral veins per cm of leaf width found in the double mutant, and corresponding reduction in the distance between lateral veins, no significant difference in the number of intermediate veins was found between lateral veins in either leaf 5 or leaf 11 (Figure 4.9C,D). The increased vein density found in the double mutant is therefore likely a result of intermediate veins being placed closer together between lateral veins. This could either occur due to a reduction in vascular, BS and/or M cell size, or due to a reduction in the number of BS or M cells separating intermediate veins.

Interestingly, there is an obvious alteration to the ratio of rank-1 to rank-2 intermediate veins found in the region separating two lateral veins. A rank-1 intermediate vein was defined by the presence of sclerenchyma located adaxially and/or abaxially to the vein (Figure 4.8A). In all of the WT and single mutant leaves examined there were never more than 3 rank-1 intermediate veins between laterals (Figure 4.8B-D), which results in the ratio of rank-1: rank-2 intermediate veins being less than 0.2 in most cases. In contrast, in the double mutant this ratio is increased to above 0.4 in virtually all cases, by the increased frequency of rank-1 intermediate veins and corresponding reduction in rank-2 intermediates (Figure 4.8E-G, Figure 4.9E). In combination, these data indicate that *ZmSCR1* and *ZmSCR1h* act redundantly to regulate correct vein spacing and intermediate vein patterning.

4.2.3.5. BS and M cell patterning is perturbed in the *Zmscr1-1; Zmscr1h-1* double mutant

The *Zmscr1-1; Zmscr1h-1* double mutant exhibited significantly increased formation of extra BS cells (resulting in a V-BS-BS-M-M-BS-V pattern) and veins with no separating M cells (resulting in a V-BS_n-V pattern), relative to either of the single

mutants or WT (Figure 4.10, Figure 4.11). It was previously reported that the *Zmscr1* single mutant exhibited occurrence of both of these phenotypes (Slewinski et al. 2012), and these results are corroborated by the analysis presented here (Figure 4.8C, Figure 4.10, Figure 4.11). However, both of these reported phenotypes are extremely subtle, with the vast majority of vascular bundles displaying normal Kranz anatomy (Figure 4.8C). The *Zmscr1h-1* mutant shows a similar, but even subtler phenotype, with slightly increased occurrences of extra BS cells, but no recorded occurrences of veins with no separating M cells (Figure 4.10, Figure 4.11). Taken together, these results indicate that ZmSCR1 and ZmSCR1h act redundantly in Kranz cellular patterning.

4.2.3.6. Alterations in foliar leaf patterning are not reflected in the husk leaves of *Zmscr1-1; Zmscr1h-1* double mutants

To determine whether ZmSCR1 and ZmSCR1h act generally in leaf patterning, rather than as specific Kranz anatomy regulators, leaf tissue from maize husk leaves was examined. Whilst not a true C_3 leaf, maize husk leaves have far more C_3-like anatomy than their foliar counterparts, with more widely spaced veins and smaller BS cell size. If ZmSCR1 and ZmSCR1h act generally in leaf development, it would be expected that similar phenotypic perturbations to those observed in foliar leaves would also be found in husk leaves. To test this, leaf tissue from the second outermost husk leaf was sampled from at least three plants for each genotype. Leaf segments were fixed in FAA overnight and then paraffin embedded for sectioning. Examination of the resulting sections revealed two types of vein were formed across all lines. These are termed here ‘major’ and ‘minor’ veins, as it is not clear how they relate to the previously defined lateral and intermediate (rank-1 and rank-2) terminology used for foliar veins. Major veins in the husk have well developed vascular cores and form sclerenchyma on the abaxial side of the leaf, resulting in a

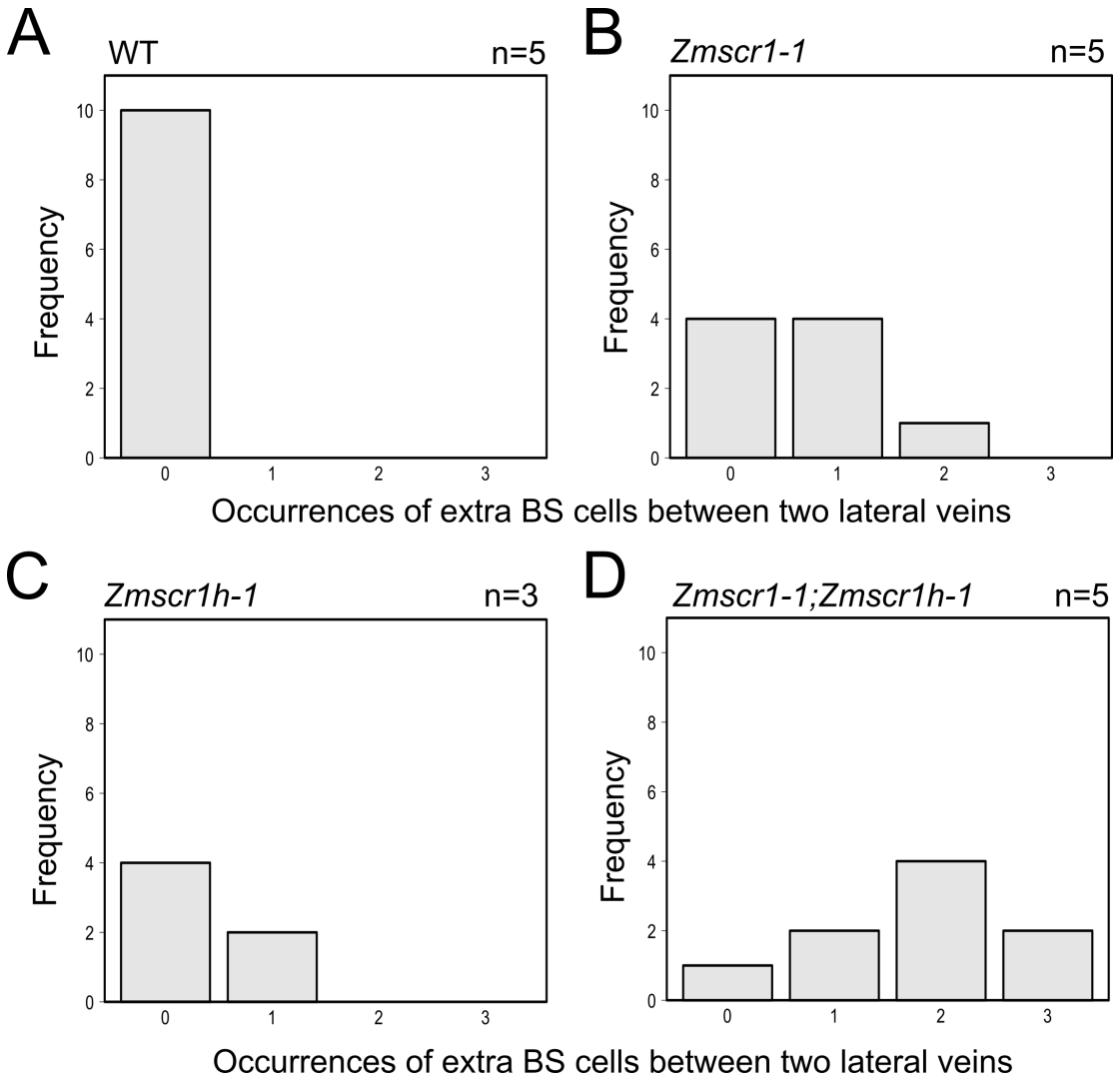


Figure 4.10 Barplots displaying the frequency of extra BS cell occurrences in *Zmscr* and WT foliar leaves. A) WT segregating from the *Zmscr1-1* line, **B)** *Zmscr1-1*, **C)** *Zmscr1h-1*, **D)** *Zmscr1-1; Zmscr1h-1.* In (A-D) the frequency that BS cells outside of the normal BS cell layer occurred was assessed using transverse fresh sections taken from the mid-point of the proximal/distal axis of fully expanded leaf 5. Extra-BS occurrences were counted between two adjacent lateral veins next to the midvein, with two pairs of lateral veins used per leaf. The total number of biological replicates is indicated by n= at the top right of each plot. Genotypes are indicated at the top left of each plot. The statistical significance ($p<0.05$) of the effect of genotype was calculated using the Pearson's Chi-squared Test, and subsequent post-hoc testing using an FDR correction.

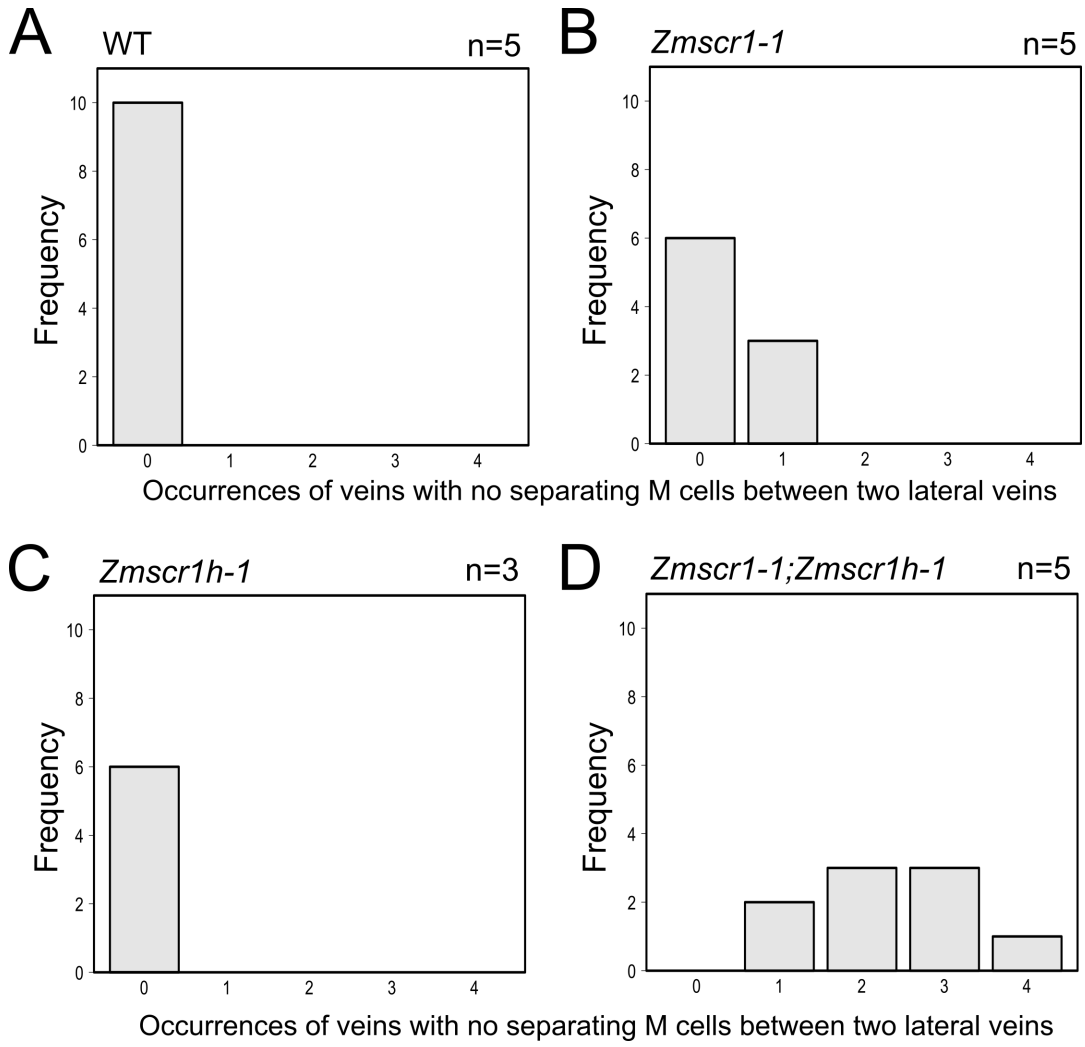


Figure 4.11 Barplots displaying the frequency of occurrences of veins with no separating M cells in *Zmscr* and WT foliar leaves. A) WT segregating from the *Zmscr1-1* line, **B)** *Zmscr1-1*, **C)** *Zmscr1h-1*, **D)** *Zmscr1-1; Zmscr1h-1.* In (A-D) the frequency that veins occurred with no separating M cells was assessed using transverse fresh sections taken from the mid-point of the proximal/distal axis of fully expanded leaf 5. Occurrences of veins with no separating M cells were counted between two adjacent lateral veins next to the midvein, with two pairs of lateral veins used per leaf. The total number of biological replicates is indicated by n= at the top right of each plot. Genotypes are indicated at the top left of each plot. The statistical significance ($p<0.05$) of the effect of genotype was calculated using the Pearson's Chi-squared Test, and subsequent post-hoc testing using an FDR correction.

curved abaxial surface (Figure 4.12). In addition, they appear to be regularly spaced across the leaf. In contrast, minor veins form very small vascular cores with only one xylem element visible, and are not associated with sclerenchyma (Figure 4.12). Interestingly, they appear to occur somewhat randomly between major veins, showing no clear spacing pattern. These veins have not previously been reported in the few studies carried out on husk leaf anatomy (Langdale et al. 1988; Pengelly et al. 2011). However, those studies were exclusively focused on the B73 inbred background, whereas the lines analysed here all came from the UniformMu W22 inbred background. It is possible, therefore, that the formation of minor veins in the husk reflects a difference between these backgrounds.

Neither the single nor double mutant exhibited any clear difference from WT in regard to vein patterning (Figure 4.13). Vein densities (veins/cm) were calculated across the width of the section, counting both major and minor veins, with no significant differences observed (Figure 4.13A). This contrasts to the foliar leaf where there is a clear increase in vein density in the double mutant relative to both WT and the single mutants (Figure 4.9A). In addition, whilst in the foliar leaves there were clear alterations in vein order, in the husk no change in the ratio of major to minor veins was observed in either the single or double mutants (Figure 4.13B). The presence of minor veins and the associated irregular spacing means that there is a wide spread in the number of M cells that separate veins (Figure 4.13C). However, this remains unaltered in the single and double mutants, which supports the lack of difference found in overall vein density in husk leaves. Finally, in the foliar leaves there was a significant increase in occurrences of extra BS cells and veins with no separating M cells in the double mutant (Figure 4.10, Figure 4.11). No occurrences of the latter are observed in any of the WT, single or double mutant husk leaves. Quantifying the former is challenging, as husk BS cells are much smaller than their foliar counterparts, and chloroplast differentiation is uneven across the sections

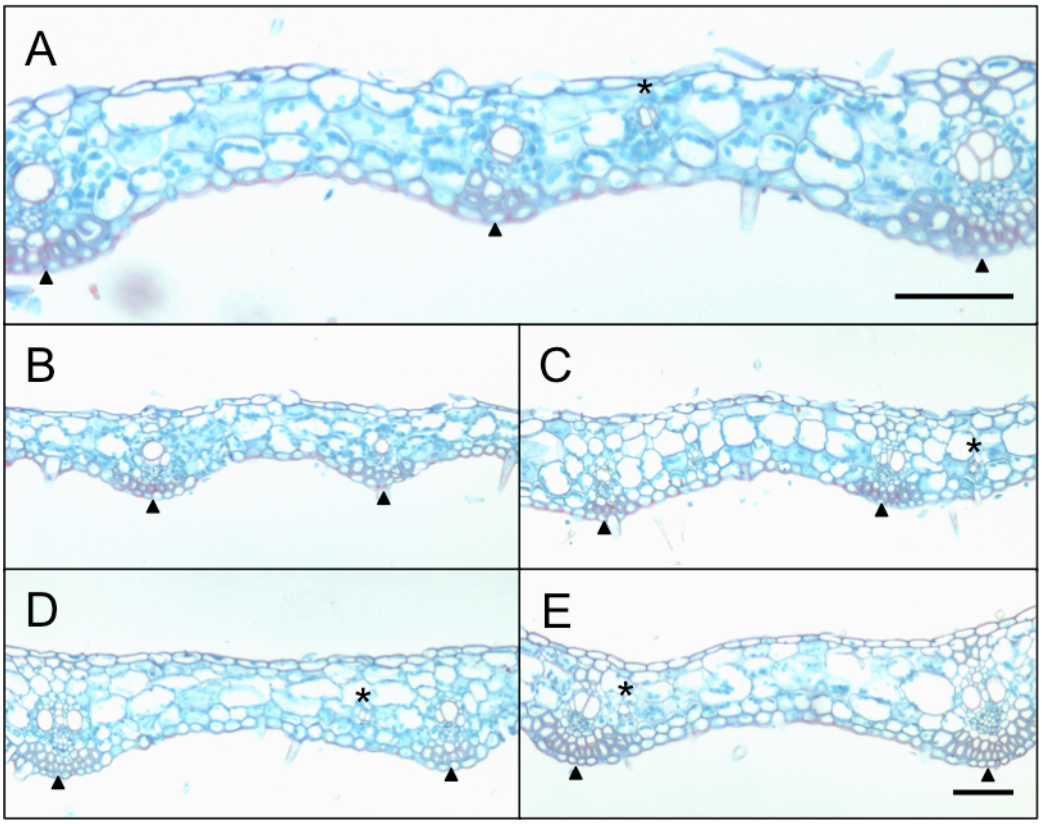


Figure 4.12 Transverse sections of husk leaves from *Zmscr* and WT plants. A,B) Segregating WT from the *Zmscr1-1* line **C)** *Zmscr1-1* **D)** *Zmscr1h-1* **E)** *Zmscr1-1; Zmscr1h-1*. In (A-E) leaf tissue from the second outermost husk leaf was fixed, sectioned (10µm) and stained with Safranin-O/ Fast Green. Arrows in (A-E) indicate sclerenchyma used to define major veins, asterisks indicates a minor vein without accompanying sclerenchyma. Scale bars = 10µm. Scale bar in (E) is the same in (B-D).

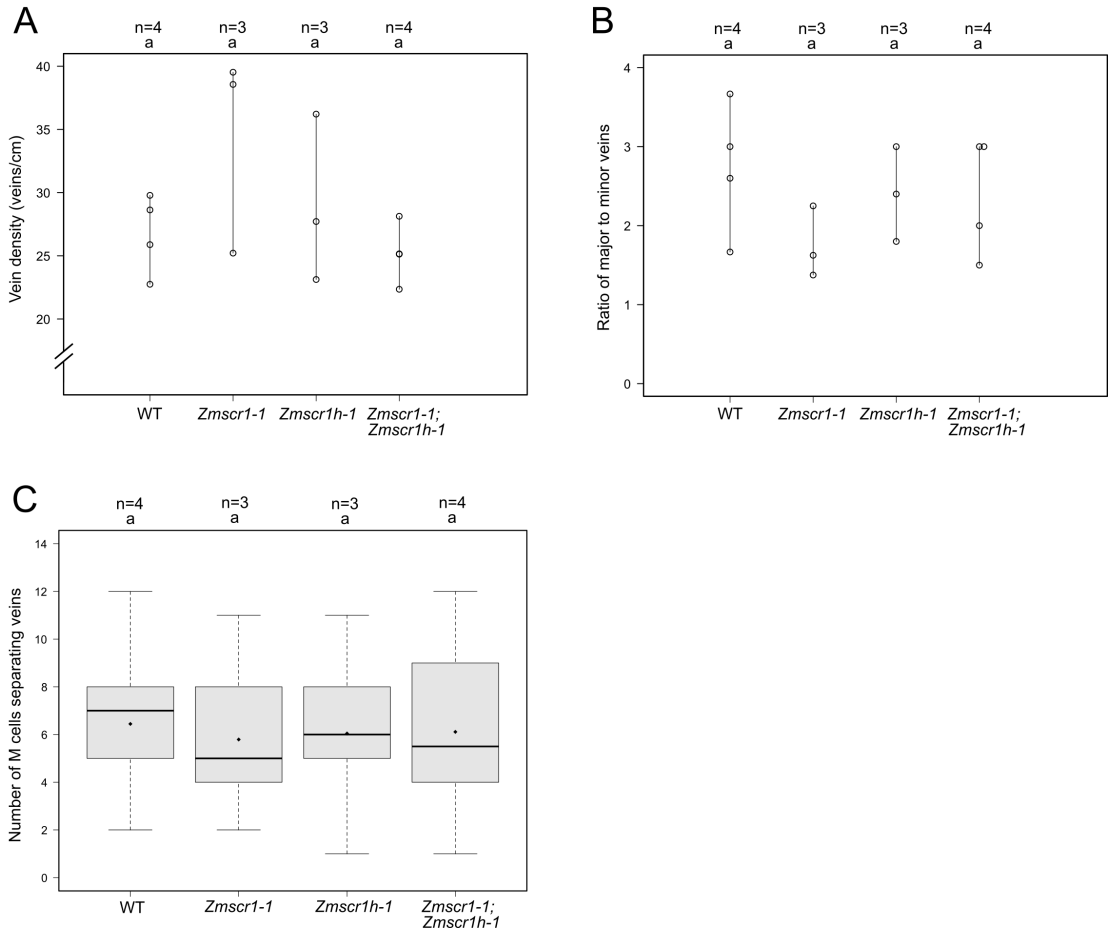


Figure 4.13 Quantification of leaf anatomy traits for *Zmscr* and WT husk leaves. A,B) Lineplots displaying total vein density (A) and the ratio of major to minor veins (B). **C)** Boxplots of the number of M cells separating two veins. Analysis was undertaken using transverse sections from the second outermost husk leaf. In (A) and (B) quantification was undertaken across the full section width. In (C) at least three M cell counts were taken per section. Statistical groups are indicated by letters above the plot, with different letters indicating statistically different ($p<0.05$) groups, as indicated by a one-way ANOVA (A,B) or mixed effect poisson regression model and likelihood ratio calculation (C). Sample sizes are indicated by n= above plots. In (C) black diamonds indicate the sample mean. Breaks in y-axis scale are indicated by a double diagonal line.

analysed, meaning it is often difficult to distinguish BS cells. However, where BS cells could be distinguished, there were no clear examples of extra BS cell formation. In summary, these data indicate that ZmSCR1 and ZmSCR1h act redundantly to regulate the development of Kranz anatomy in maize foliar leaves, but do not obviously regulate the development of more C_3-like leaf structures such as the husk leaves.

4.2.3.7. *ZmSCR* genes contain an N-terminal domain that shows some homology to that of *AtSCR*

Recent *in-situ* hybridisation work has revealed that both *ZmSCR1* and *ZmSCR1h* are expressed preferentially in the M cells of maize leaf primordia (Olga Sedelnikova, personal communication). In light of this, it is intriguing that the double mutant phenotype indicates a role for both ZmSCR genes in vascular organisation and BS specification. One possibility is that, unlike AtSCR, ZmSCR1 and/or ZmSCR1h act non-cell autonomously. A number of GRAS transcription factors have been found to be mobile in *Arabidopsis*, including AtSHR and AtSCL23 (Nakajima et al. 2001; Cui et al. 2007; Rim et al. 2011; Long et al. 2015a). The mobility of AtSCL23 has been partially attributed to the lack of an N-terminal domain upstream of the GRAS domain, which in AtSCR has been demonstrated to prevent protein movement (Gallagher and Benfey 2009; Long et al. 2015a). Therefore, if either ZmSCR1 or ZmSCR1h lack an N-terminal domain it might suggest that they might be mobile, and thus able to act cell non-autonomously. To test this hypothesis, both ZmSCR1 and ZmSCR1h protein sequences were aligned to the AtSCR and AtSCL23 sequences using Clustal Omega (Figure 4.14). This alignment revealed that, unlike AtSCL23, both ZmSCR1 and ZmSCR1h have an N-terminal domain of similar length to that of AtSCR. However, whilst the GRAS domain is highly conserved between AtSCR and both ZmSCR1 and ZmSCR1h (78.61% and 75% identity respectively), the N-terminal domains are far less well conserved (32.19% and 29.24% identity respectively).

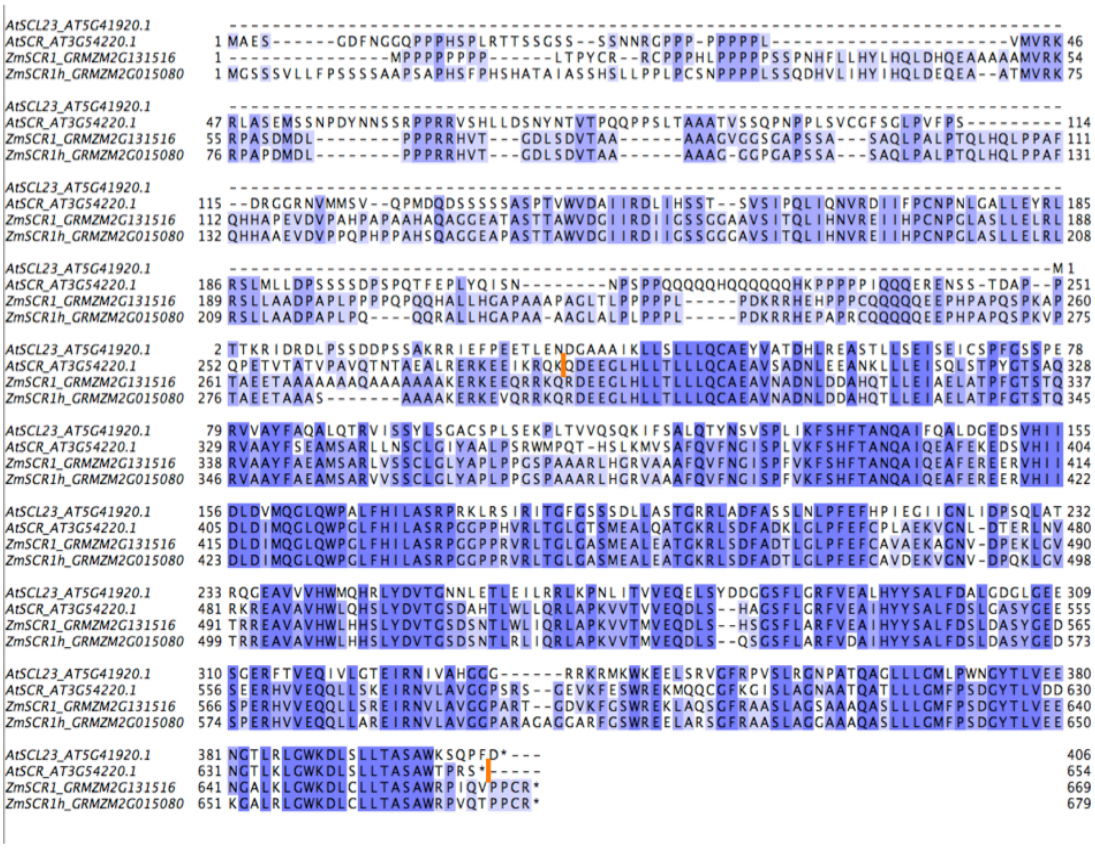


Figure 4.14 Alignment of *Arabidopsis* and maize SCR protein sequences. Primary transcript protein sequences for AtSCL23, AtSCR, ZmSCR1 and ZmSCR1h were acquired from Phytozome 11 and aligned using Clustal Omega. The region between the two orange lines indicates the AtSCR GRAS domain, as defined by Gallagher and Benfey (2009). Upstream of this region is the N-terminal domain (missing in AtSCL23), that is essential for preventing movement of AtSCR (Gallagher and Benfey 2009). Darker blue highlighting indicates greater sequence conservation between proteins.

Despite this low sequence conservation, there are some small areas of sequence homology between the species. As the precise regions of the N domain that prevent AtSCR movement are unknown, it is possible that these regions are conserved in maize. In summary, both ZmSCR1 and ZmSCR1h contain an N-terminal domain, which although exhibiting low sequence similarity with the AtSCR N-terminal domain, might prevent *ZmSCR* movement beyond its site of expression.

4.2.4. The effect of *ZmSCR1* overexpression in rice

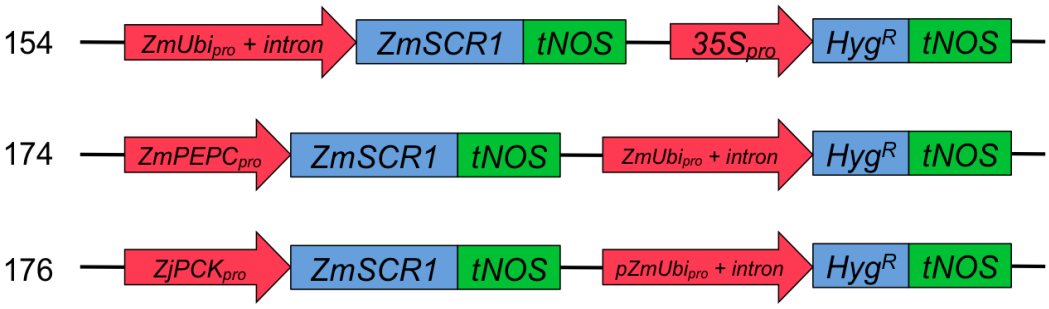
4.2.4.1. *ZmSCR1* overexpression constructs were cloned for rice transformation

As the *ZmSCR* genes appear to be key regulators of Kranz anatomy development, rather than leaf development more generally, it was hypothesised that expressing them in rice might induce more C_4-like leaf anatomy. To test this hypothesis, three constructs were designed and cloned that placed *ZmSCR1* under the control of the maize ubiquitin promoter (constitutive overexpression, construct 154), the maize *PEPC* promoter (M cell specific expression, construct 174) and the *Zoysia japonica PCK* promoter (BS cell specific expression, construct 176) (Figure 4.15). *ZmSCR1,* rather than *ZmSCR1h,* was used in these experiments as *ZmSCR1* and *ZmSCR1h* function redundantly in maize foliar leaf development, and thus overexpression of either should be sufficient to reveal potential leaf perturbations in rice. To ensure high expression levels in rice, the maize gene sequence was first codon-optimised for expression in rice.

4.2.4.2. Rice plants transformed with *ZmSCR1* overexpression constructs exhibited transgene-expression

The rice codon-optimised *ZmSCR1* sequence was synthesised and cloned into the relevant Gateway vectors using by Mara Schuler. Transformation and copy-number analysis was also carried out by Mara Schuler. Constructs were transformed into

A



B

Construct ID	Transgenic line ID
154	1B, 5A, 7C
174	1D, 2I, 3D
176	18, 6O, 9D

Figure 4.15 Cartoon depictions and transgenic line IDs for *ZmSCR1* overexpression constructs. A) Cartoon depictions of three constructs designed and cloned by Mara Schuler for *ZmSCR1* overexpression in rice. In cartoon depictions of each construct, red refers to a promoter (pro), blue to a gene and green to a terminator (in all cases tNOS). **B)** Three independent transgenic lines were analysed for each construct, with IDs for these included in the table. IDs in which a number is followed by a letter indicate T2 seed, whereas IDs made up only of a number refer to T1 seed.

Kitaake rice callus and regenerated on hygromycin, with the resultant T0 seedlings allowed to set T1 seed. T1 plants were then grown and DNA samples sent for copy-number analysis (iDNAgenetics, www.idnagenetics.com), allowing single insertion events to be identified, and T2 seed were then collected from these plants.

Phenotypic analysis was carried out for most lines on T2 plants. Where possible, segregating single insertion hemizygous lines were used in order to allow comparisons to segregating WT sibling plants. In two cases (MS154.1B and MS176.18) homozygous lines had to be used (due to seed availability), and in these cases overexpressing (OE) plants were compared to Kitaake WT. In all cases, three independent transgenic lines were used for analysis. Plants were genotyped and three WT and OE plants selected for each line for further analysis (data not shown). RNA was extracted from the 4th fully expanded leaf and DNase treated to remove DNA contamination. This RNA was used as template for first strand cDNA synthesis. Expression of the maize *ZmSCR1* gene was checked by RT-PCR for all plants to be analysed (Figure 4.16). For the ZmUbi (construct 154) and ZmPEPC promoter (construct 174) lines, a strong amplicon was observed (Figure 4.16). For the ZjPCK promoter (construct 176) lines a far weaker amplicon was observed, consistent with the predicted BS-specific expression and relatively weaker activity of the *ZjPCK* promoter (Figure 4.16).

4.2.4.3. No obvious differences in leaf cellular patterning were observed in rice plants overexpressing *ZmSCR1*

To determine whether expression of *ZmSCR1* in rice perturbs development, phenotypic analysis was undertaken 42 days after planting. No obvious changes in plant morphology were observed in any of the lines, with OE plants of similar height and following a similar developmental trajectory to WT. To determine whether *ZmSCR1* overexpression caused more subtle changes in cellular patterning, fully expanded leaf 5 was sampled 12cm from the leaf tip, and sections taken over the

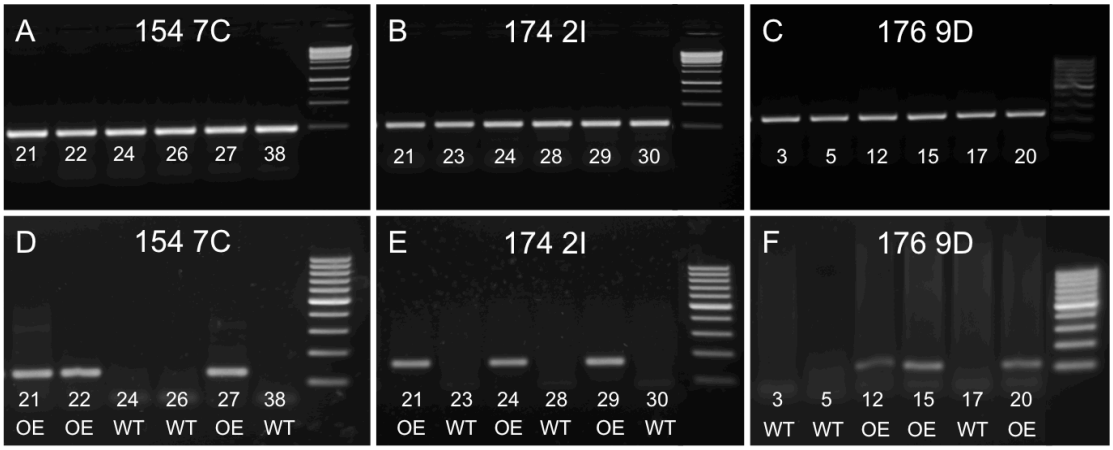


Figure 4.16 Gel images showing expression of *ZmSCR1* in transgenic rice lines. A-F) RT-PCR of the rice ubiquitin gene (A-C) and *ZmSCR1* (D-F). One representative T2 line is displayed for each construct: line 7C for construct 154 (*ZmUbi$_{pro}$::ZmSCR1*) (A,D), line 2I for construct 174 (*ZmPEPC$_{pro}$::ZmSCR1*) (B,E) and line 9D for construct 176 (*ZjPCK$_{pro}$::ZmSCR1*) (C,F). cDNA was amplified from RNA isolated from mature leaf tissue. In (A-C) the smallest band of the DNA ladder = 250bp. In (D-F) the smallest band of the DNA ladder = 100bp. WT and overexpression (OE) plants are indicated in (D-F).

whole leaf width using a vibratome. No consistent alterations in leaf patterning were observed (Figure 4.17, Figure 4.18). In the *ZmUbi$_{pro}$::ZmSCR1* lines, 2 of the 3 independent transgenic events displayed an increase in vein density relative to the corresponding WT controls, however, this was not reflected in the 3rd independent line, nor in the *ZmPEPC$_{pro}$::ZmSCR1* or *ZjPCK$_{pro}$::ZmSCR1* lines (Figure 4.18). Comparison of all *ZmUbi$_{pro}$::ZmSCR1* overexpression plants to a much larger dataset of WT vein measurements indicated that all three lines displayed normal vein spacing (data not shown).

In maize, loss of *ZmSCR1/1h* function resulted in both the formation of extra BS cells outside of the normal single BS cell layer, and veins with no separating M cells. To test whether overexpression of *ZmSCR1* induced alterations in rice BS patterning the number of BS cells surrounding four intermediate vascular bundles per plant was recorded in all of the *ZmSCR1* over-expressing lines. No additional BS cells nor any difference in the number of BS cells surrounding the vein was found between WT and over-expressing plants (Figure 4.18). To test whether M cell patterning was affected by overexpression of *ZmSCR1*, the number of M cells separating vascular bundles was recorded for four distinct inter-veinal regions per plant. No significant differences were found between WT and overexpressing plants (Figure 4.18). Given these data, it seems unlikely that introduction of *ZmSCR1* into rice is likely to result in a shift to more C_4-like leaf anatomy, regardless of the promoter or expression context in question.

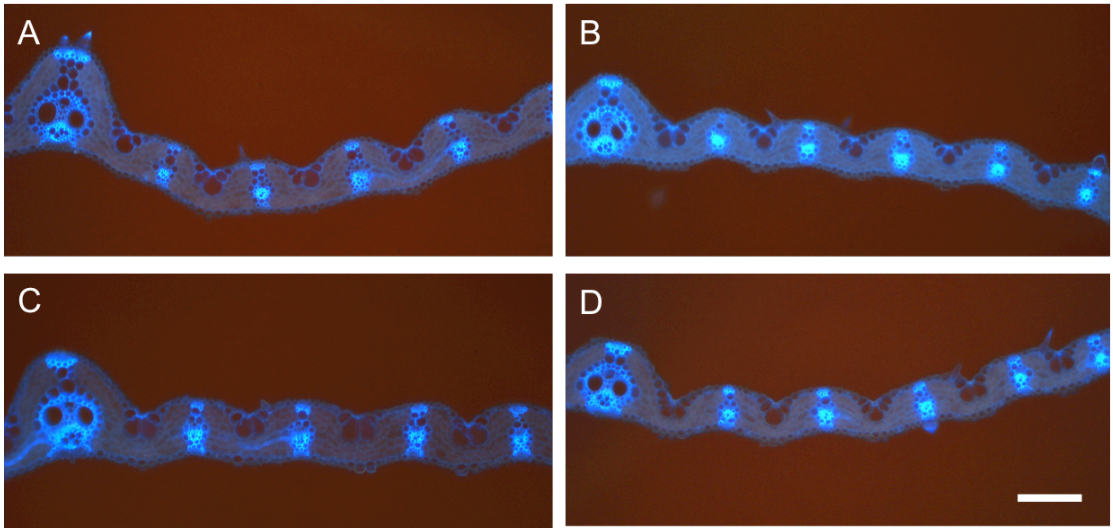


Figure 4.17 Transverse fresh sections of Kitaake WT and *ZmSCR1* overexpressing rice leaves. A) Kitaake WT **B)** construct 154 line 1B (*ZmUbi*$_{pro}$*::ZmSCR1*) **C)** construct 174 line 2I (*ZmPEPC*$_{pro}$*::ZmSCR1*) **D)** construct 176 6O (*ZjPCK*$_{pro}$*::ZmSCR1*). In (A-D) sections were taken between 10 and 12 cm from the leaf tip of rice fully expanded leaf 5, and imaged under UV illumination. Scale bar = 100µm.

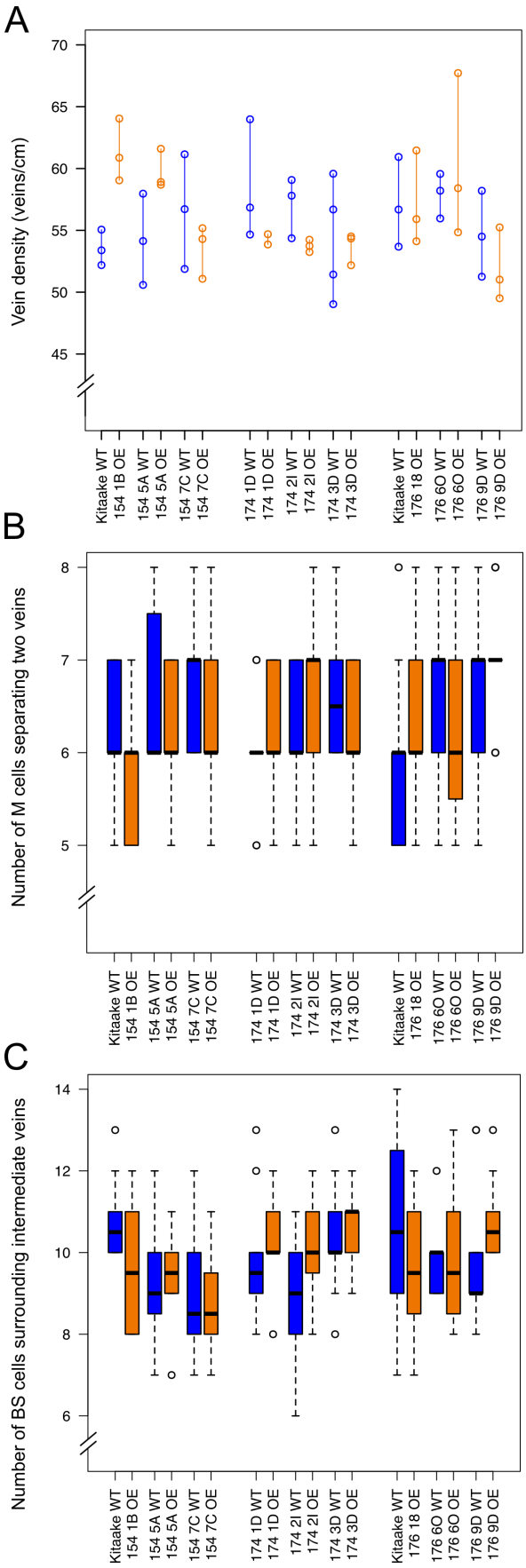


Figure 4.18 Quantification of Kranz-related anatomical traits in rice lines overexpressing *ZmSCR1*. Figure legend is on the next page.

Figure 4.18 Quantification of Kranz-related anatomical traits in rice lines overexpressing *ZmSCR1*. A) Lineplot displaying total vein density **B,C)** Boxplots displaying the number of M cells separating two veins (B) and the number of BS cells surrounding an intermediate vein (C). Analysis in (A-C) was undertaken on fully expanded leaf 5 from transverse fresh sections made between 10-12cm from the leaf tip. In (A) veins were counted across the full width of the leaf, and t-tests were undertaken between segregating WT and *ZmSCR1* OE lines to test for statistical differences. Only 154 1B OE vs Kitaake WT was significant ($p < 0.05$). In (B) and (C) between 2 and 6 measures were taken per plant. Mixed effect poisson regression models were fitted for both M and BS cell counts. Likelihood ratio p-values were obtained by comparison to a null model, but no significant effect ($p<0.05$) attributable to genotype was detected. 154- *ZmUbi*$_{pro}$*::ZmSCR1*, 174- *ZmPEPC*$_{pro}$*::ZmSCR1*, 176- *ZjPCK*$_{pro}$*::ZmSCR1*. In each panel blue plots indicate segregating WT comparisons (unless indicated Kitaake), orange indicate overexpressing (OE) plants. In nearly all cases the number of biological replicates (n) = 3. Three exceptions to this occurred: 1) 174 1D OE where n=2, 2) 174 3D WT where n=4, and 3) 176 18 in panel B where n=2. Breaks in y-axis scale are indicated by a double diagonal line.

4.2.5. The role of *ZmSHR2* in maize

4.2.5.1. Neither plant nor leaf development is perturbed in the *Zmshr2-1* mutant

To determine whether *ZmSHR2* plays a role in Kranz anatomy development, a line segregating for the *Zmshr2-1* UniformMu insertion was analysed. Plants homozygous for the insertion did not display any changes in gross morphology relative to segregating WT plants. No obvious differences in plant height, leaf development or flowering time were recorded, although no assessment of root development was made in this line. Mature leaf cross sections were taken both using the vibratome (Figure 4.19), and these were used to quantify a number of Kranz anatomy related traits. No significant differences were found in relation to vascular patterning in the *Zmshr2-1* mutant, which is surprising considering how specifically *ZmSHR2* transcripts mark early differentiating veins in leaf primordia (Olga Sedelnikova, personal communication) (Figure 4.20). Neither total vein density nor the number of intermediate veins found between two lateral veins differed significantly between WT and *Zmshr2*-1 (Figure 4.20A,B). In addition, no difference in the ratio of rank-1 to rank-2 intermediate veins was found (Figure 4.20C).

No evidence for *ZmSHR2* involvement in BS and M cell patterning was obtained from the *Zmshr2-1* single mutant. Both WT and *Zmshr2-1* exhibited similar frequencies of extra BS cell formation (Figure 4.21). The degree to which this occurred was surprisingly high in the segregating WT lines, which were higher than other segregating WT lines examined. Whilst it is known that extra BS do sometimes occur at low frequency in UniformMu lines (Slewinski et al. 2012), the difference found here between the segregating *Zmshr2-1* WT and other examined segregating WT lines likely reflect either an artefact of the relatively small sample size, or the fact that there are at least seven additional Mu insertions in the *Zmshr2-1* seed stock (Appendix 2).

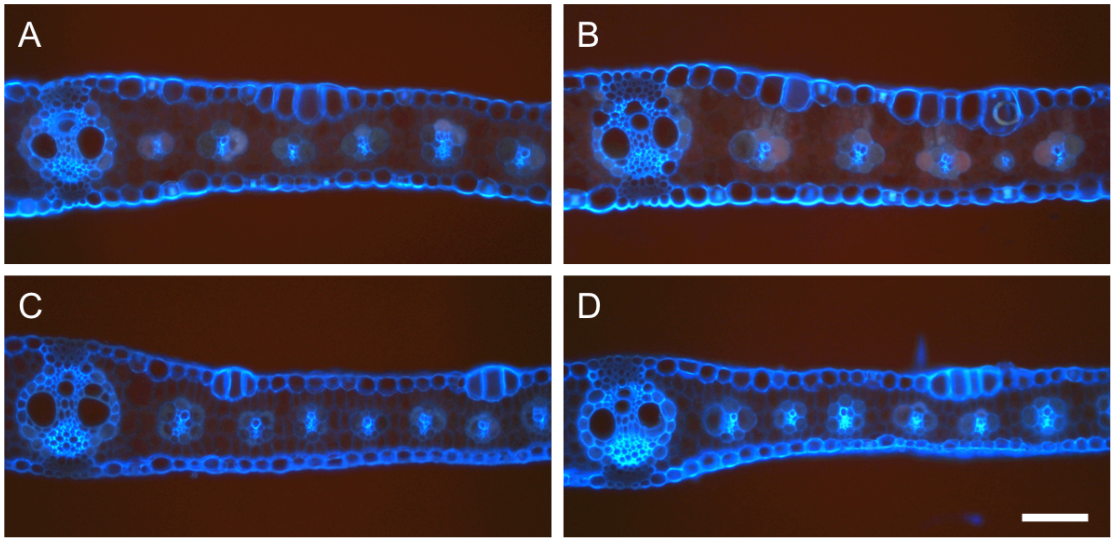


Figure 4.19 Transverse fresh sections of *Zmshr2-1* and *Zmscr1-1; Zmshr2-1* maize foliar leaves. A) WT (segregating from *Zmshr2-1* background) **B)** *Zmshr2-1* **C)** WT (segregating from *Zmscr1-1; Zmshr2-1* background) **D)** *Zmscr1-1; Zmshr2-1*. In (A-D) fresh transverse sections of fully expanded foliar leaf 4 (A,B) and leaf 5 **(**C,D) were taken from the midpoint along the proximal/distal axis, and imaged using UV illumination. Scale bar = 100μm.

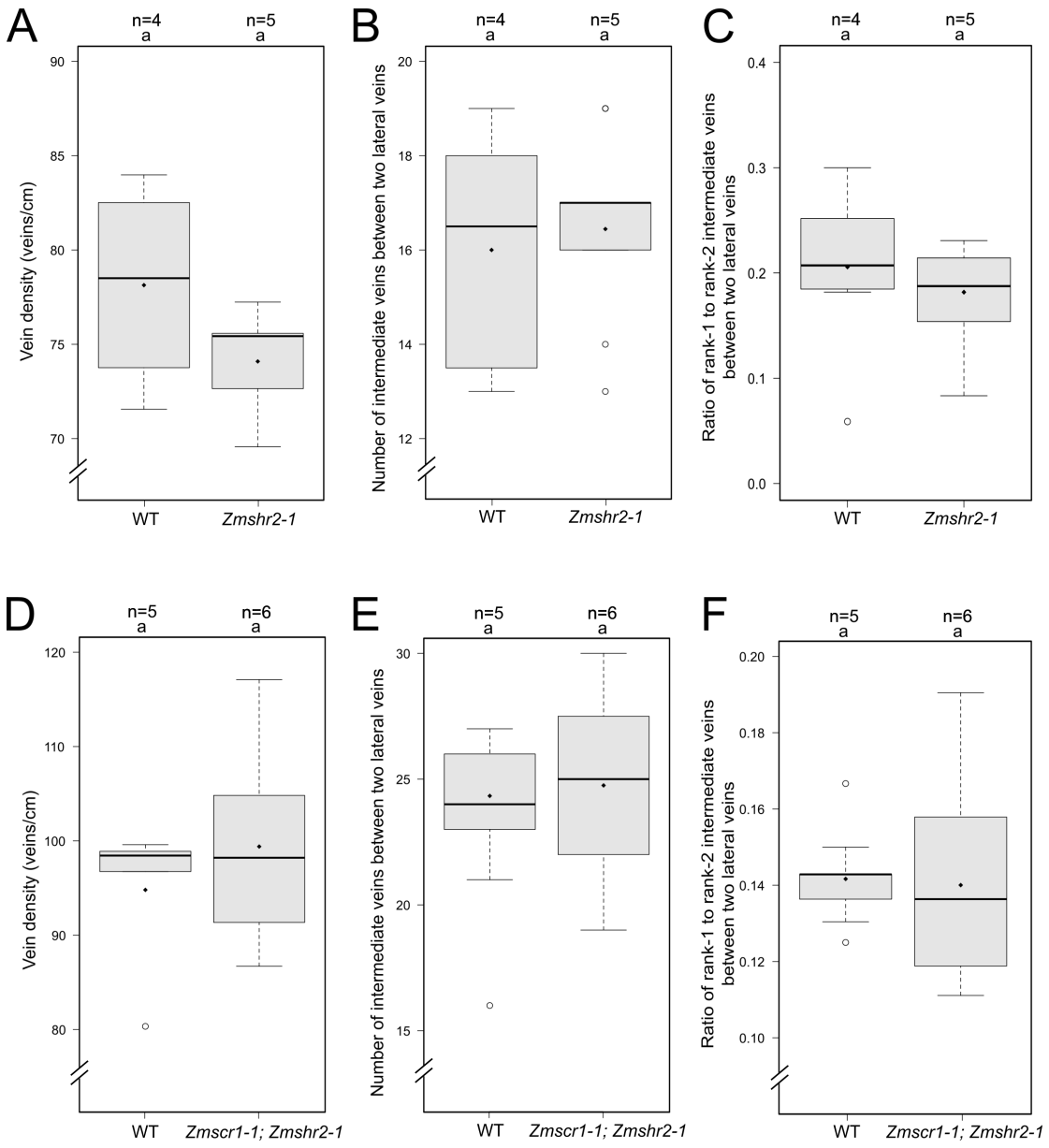


Figure 4.20 Boxplots displaying quantification of vein-patterning traits in *Zmshr2-1* and *Zmscr1-1; Zmshr2-1* foliar leaves. Figure legend is on the next page.

Figure 4.20 Boxplots displaying quantification of vein-patterning in *Zmshr2-1* and *Zmscr1-1; Zmshr2-1* foliar leaves. A,D) Vein density (veins/cm) **B,E)** Number of intermediate veins between laterals veins **C,F)** Ratio of rank-1 to rank-2 intermediate veins between laterals veins. In (A-C) comparisons are between WT segregating in the *Zmshr2-1* line, and *Zmshr2-1*. In (D-F) comparisons are between WT segregating in the *Zmscr1-1; Zmshr2-1* line, and *Zmscr1-1; Zmshr2-1*. All measures in (A-F) were taken from fresh transverse sections from the midpoint along the proximal/distal axis of fully expanded leaf 4 (A-C) or leaf 5 (D-F), from two pairs of laterals veins adjacent to the midvein. Statistical groups are indicated by letters above the plot, with different letters indicating statistically different ($p<0.05$) groups, as calculated by a t-test (A), mixed effect poisson regression model and subsequent likelihood ratio calculation by comparison to a null model (B,E), Wilcoxon rank sum test (D) or linear mixed effect regression model and subsequent likelihood ratio calculation by comparison to a null model (C,F). Biological replicates are indicated by n= above significance groupings. Means are indicated by black diamonds for each genotype, outliers are indicated by empty circles. Breaks in y-axis scale are indicated by a double diagonal line.

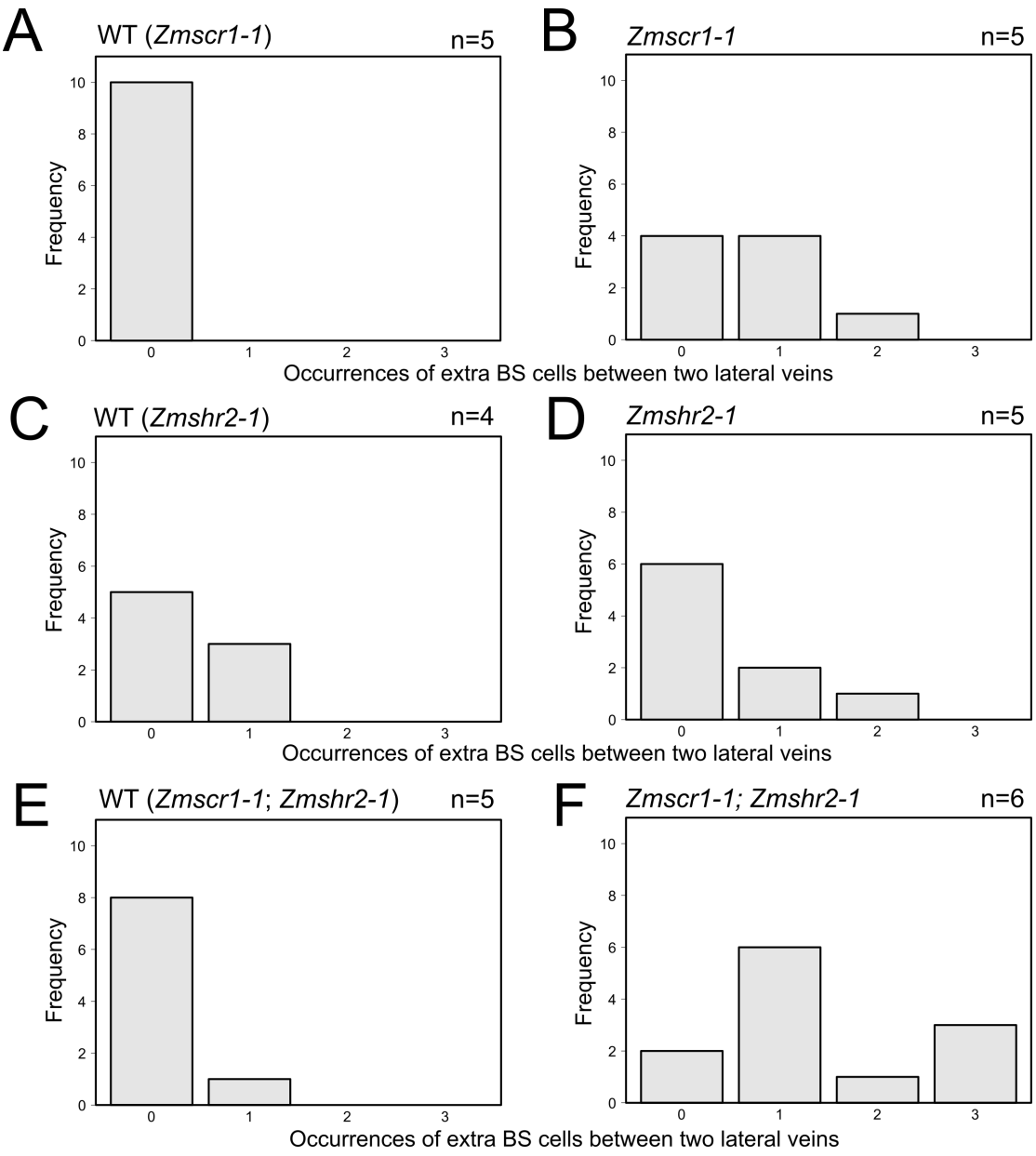


Figure 4.21 Barplots displaying the frequency of extra BS cell occurrences between two lateral veins in *Zmscr1-1, Zmshr2-1* and *Zmscr1-1; Zmshr2-1* foliar leaves. Figure legend is on the next page.

Figure 4.21 Barplots displaying the frequency of extra BS cell occurrences between two lateral veins in *Zmscr1-1, Zmshr2-1* and *Zmscr1-1; Zmshr2-1* foliar leaves. A) Segregating WT from *Zmscr1-1* line **B)** *Zmscr1-1* **C)** Segregating WT from *Zmshr2-1* line **D)** *Zmshr2-1* **E)** Segregating WT from *Zmscr1-1; Zmshr2-1* line **F)** *Zmscr1-1; Zmshr2-1*. Frequency of extra BS cells occurrence was recorded from fresh transverse sections taken from the mid-point along the proximal/distal axis of fully expanded leaf 5 (A,B,E,F) or leaf 4 (C,D) for each biological replicate. Extra-BS occurrences were counted between two adjacent lateral veins next to the midvein, with two pairs of lateral veins used per leaf. The total number of biological replicates is indicated by n= at the top right of each plot. Genotypes are indicated at the top left of each plot. Pearson's Chi-squared tests were used to determine statistical significance between WT and mutant comparisons. Statistically significant differences ($p<0.05$) were found between (A) and (B,) and (E) and (F), but not between (C) and (D).

No additional insertions were mapped in either the *Zmscr1-1* or *Zmscr1h-1* seed stocks (Appendix 2). Furthermore, no significant differences in occurrences of veins with no separating M cells patterning were found (Figure 4.22). Taken together, these data indicate that mutations in *Zmshr2-1* alone are insufficient to functionally validate its proposed role in Kranz development.

4.2.5.2. The *Zmscr1-1; Zmshr2-1* double mutant did not show additional phenotypic perturbations

In *Arabidopsis*, the *Atscr;Atshr* double mutant exhibits an identical phenotype to the single *Atshr* mutant, however no equivalent test has been undertaken in monocots (Helariutta et al. 2000). As there are three maize *SHR* genes (*ZmSHR1, ZmSHR2* and *ZmSHR2h*) and two maize *SCR* genes (*ZmSCR1* and *ZmSCR1h*) that are orthologous to *AtSHR* and *AtSCR* respectively, it is unclear how *SCR/SHR* function (if conserved in maize) is distributed across these genes. To determine whether loss-of-function of two potential members of this pathway yielded a more severe phenotype than if either is examined individually, the double mutant *Zmscr1-1; Zmshr2-1* was constructed.

Similar to the *Zmscr1-1* and *Zmshr2-1* single mutants, the *Zmscr1-1; Zmshr2-1* double mutant displayed no obvious morphological perturbations. Plant growth proceeded normally with height, growth and flowering time unaffected by the presence of mutant alleles. Phenotypic analysis was undertaken with plants segregating WT for both insertions and the homozygous double mutant, with these measurements compared to previously obtained analysis of the *Zmscr1-1* and *Zmshr2-1* single mutants. The double mutant did not display a more severe phenotype with regards to vascular patterning, in that vein density, the number of intermediate veins between laterals and the ratio of rank-1 to rank-2 intermediate veins were all found to be non-significantly different to WT (Figure 4.20D-F). This confirms the previous conclusion that vein density does not differ in the *Zmscr1-1*

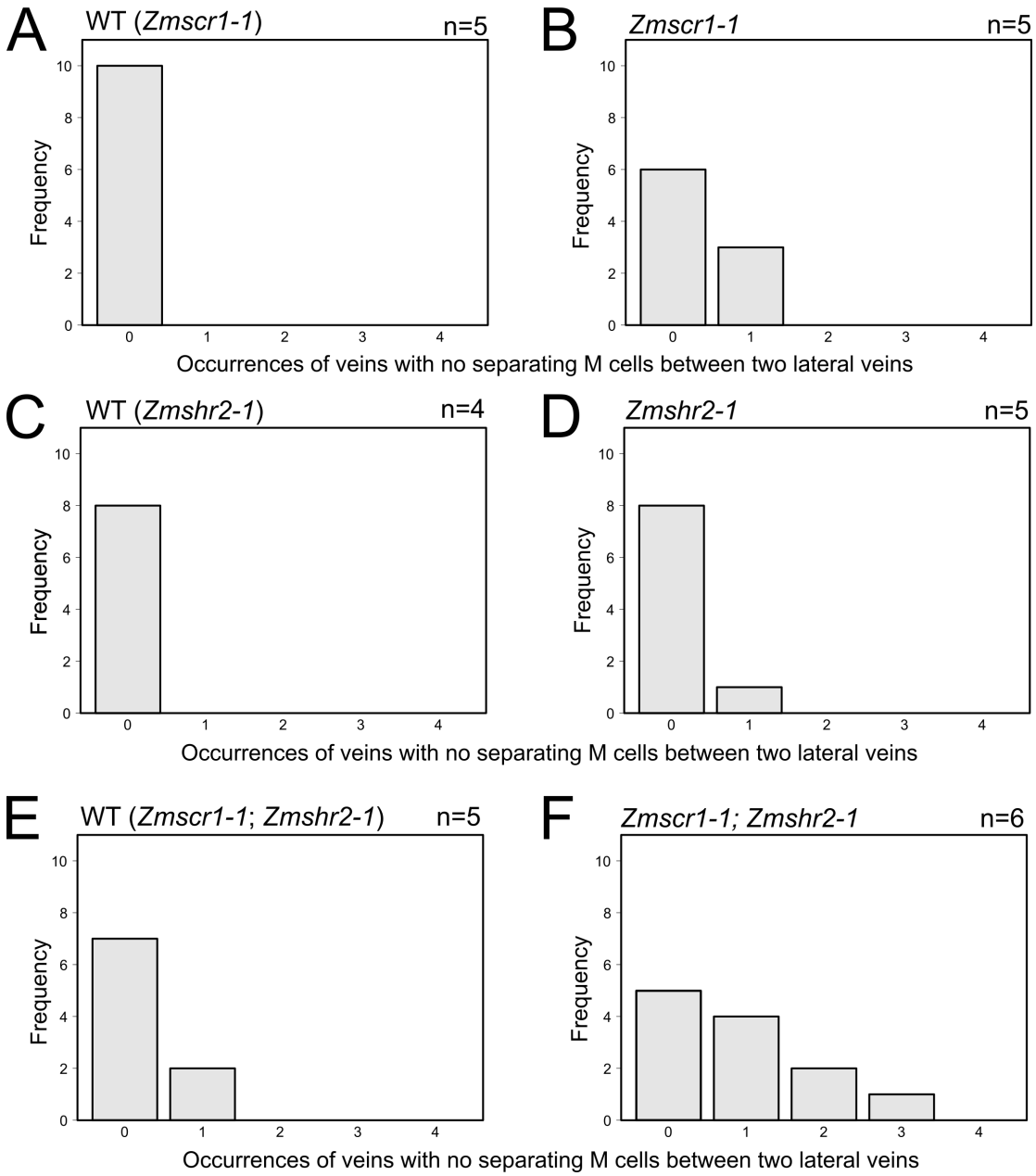


Figure 4.22 Barplots displaying the frequency of occurrences of veins with no separating M cells between two lateral veins in *Zmscr1-1, Zmshr2-1* and *Zmscr1-1; Zmshr2-1* foliar leaves. Figure legend is on the next page.

Figure 4.22 Barplots displaying the frequency of occurrences of veins with no separating M cells between two lateral veins in *Zmscr1-1, Zmshr2-1* and *Zmscr1-1; Zmshr2-1* foliar leaves. A) Segregating WT from the *Zmscr1-1* line **B)** *Zmscr1-1* **C)** Segregating WT from the *Zmshr2-1* line **D)** *Zmshr2-1* **E)** Segregating WT from the *Zmscr1-1; Zmshr2-1* line **F)** *Zmscr1-1; Zmshr2-1***.** The frequency of occurrences of veins with no separating M cells was recorded from fresh transverse sections taken from the mid-point along the proximal/distal axis of fully expanded leaf 5 (A,B,E,F) or leaf 4 (C,D) for each biological replicate. Occurrences of veins with no separating M cells were counted between two adjacent lateral veins next to the midvein, with two pairs of lateral veins used per leaf. The total number of biological replicates is indicated by n= at the top right of each plot. Genotypes are indicated at the top left of each plot. Pearson's Chi-squared tests were used to determine statistical significance between WT and mutant comparisons. No statistical differences ($p<0.05$) were found.

background without corresponding mutations in *Zmscr1h*, despite the previous findings to the contrary (Slewinski et al. 2012).

Interestingly, preliminary analysis suggested that the extra incidents of BS cell formation and veins with no separating M cells were found at higher frequency in the *Zmscr1-1; Zmshr2-1* double mutant than in the single *Zmscr1-1* background, suggesting involvement of ZmSHR2 in Kranz development (Figure 4.21, Figure 4.22). However, these incidents are recorded for the intermediate veins between two laterals, and as such variation in the number of intermediate veins between laterals will affect these measurements. Phenotypic analysis for the *Zmscr1-1; Zmshr2-1* double mutant was undertaken at a slightly later developmental stage than for other experiments, and as such leaf width was increased such that the number of intermediate veins between laterals was higher relative to other experiments. When expressed per intermediate vein, it is clear that the *Zmscr1-1; Zmshr2-1* double mutant does not exhibit higher occurrences of either extra BS cells or veins with no separating M cells, than the single *Zmscr1-1* mutant (Figure 4.23). Therefore, this analysis does not provide any functional evidence of ZmSHR2 involvement in Kranz patterning.

4.3. Discussion

4.3.1. *ZmSCR1* and *ZmSCR1h* have been recruited from their ancestral C_3 function to redundantly regulate the development of Kranz anatomy

The analysis presented in this chapter validates the previous finding that *ZmSCR1* plays a role in regulating Kranz anatomy. Furthermore, it demonstrates that *ZmSCR1* functions redundantly with *ZmSCR1h* in both root and foliar leaf development, indicating that since the maize whole genome-duplication there has not been significant functional divergence of this homeolog gene-pair in these contexts. It was previously reported that *AtSCR* is expressed in the BS of developing *Arabidopsis*

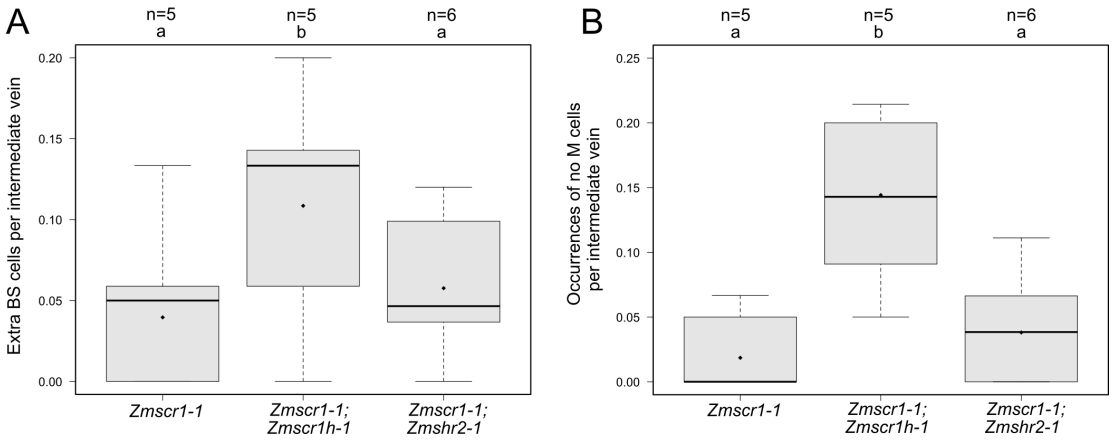


Figure 4.23 Boxplots displaying normalisation of BS and M cell patterning traits for *Zmscr1-1; Zmshr2-1* double mutants. A,B) Boxplots displaying occurrences of extra BS cells (A) and occurrences of veins with no separating M cells (B), both normalised to the number of intermediate veins present between each set of laterals. Statistical groups are indicated by letters above the plot, with different letters indicating statistically different ($p<0.05$) groups, calculated using the Kruskal-Wallis rank sum test. Biological replicates are indicated by n= above significance groupings. Means are indicated by black diamonds for each genotype.

leaves, and that mutations in *Atscr* resulted in subtle defects in BS differentiation (Cui et al. 2014). This indicates that *SCR* plays a role in C_3 BS development. However, in maize, mutations in both *SCR* orthologs result in far more severe perturbations to leaf development than those seen in *Arabidopsis*, with alterations in vein density, intermediate vein organisation and cellular patterning. These perturbations are not reflected in the more C_3-like husk leaves, indicating that *ZmSCR1* and *ZmSCR1h* have evolved Kranz-specific functions. BS patterning could not be quantified in husk leaves, so it remains possible that the subtle BS perturbations observed in *Arabidopsis* are also present in maize C_3-like leaves.

ZmSCR1 and *ZmSCR1h* are among the very first genes confirmed to play a role in the development of Kranz anatomy. However, even in the *Zmscr1-1; Zmscr1h-1* double mutant, there are more vascular bundles displaying normal Kranz formation than those exhibiting perturbations. This contrasts to the root, where perturbations appear to be found consistently across all roots examined. This may, in part, be explained by the inherent increase in complexity associated with leaf patterning. This occurs due to the nature of leaf development, which occurs along three distinct axes: proximal/distal, medial/lateral and ad/abaxial. In contrast, root development takes place across only two axes: radial and longitudinal. This means that understanding the patterning mechanisms underpinning leaf development is more complicated, and this added complexity might render it more difficult to perturb leaf development. This difficulty is reflected both in the paucity of genes confirmed as having leaf-patterning roles, and the difficulties associated with identifying vein-spacing mutants. As such, the confirmed roles of *ZmSCR1* and *ZmSCR1h* in regulating the development of Kranz anatomy are exciting, and point to the recruitment of the SCR/SHR regulatory system into the leaves of C_4 species.

4.3.2. *ZmSCR1* and *ZmSCR1h* are proposed to act alongside ZmSHR to control Kranz development

It is not yet clear, mechanistically, how *ZmSCR1* and *ZmSCR1h* regulate Kranz development. Given that many of the Kranz perturbations relate to vascular patterning and BS cell formation, it is intriguing that in situ hybridisation shows both *ZmSCR1* and *ZmSCR1h* to be preferentially expressed in the M cells of developing leaves (Olga Sedelnikova, personal communication). Both *ZmSCR1* and *ZmSCR1h* have an N-terminal domain, and as such it is most parsimonious to assume that they function cell autonomously (Gallagher and Benfey 2009). In an *Arabidopsis SCR* RNAi line, reductions in the amount of AtSCR resulted in an increased domain of AtSHR movement, which in turn altered cellular patterning (Cui et al. 2007; Wu et al. 2014). It might be that in the *Zmscr1-1; Zmscr1h-1* double mutant, ZmSHR movement is increased such that vein differentiation and BS formation are affected in a dosage dependent manner. This hypothesis forms the basis of the model for Kranz development proposed previously by Fouracre *et al.* (2014), in which ZmSHR1 and ZmSHR2 were proposed to act as the previously hypothesised mobile signals emanating from the vasculature that trigger Kranz cell patterning (Langdale and Nelson 1991; Fouracre et al. 2014). The functional validation of the roles of *ZmSCR1* and *ZmSCR1h* is supportive of this hypothesis, however a number of additional experiments will be necessary to confirm or refute it fully.

4.3.3. *ZmSCR1* and *ZmSCR1h* act redundantly in maize root development

Previous studies found evidence of functional conservation between *Arabidopsis* and maize *SCR* genes in the context of the root, as maize *ZmSCR1* was able to complement the *Atscr* ground-tissue phenotype (Lim et al. 2005). However, the function of maize *SCR* genes in *Arabidopsis* does not necessarily reflect the function of these genes in their endogenous context. In common with *Arabidopsis,* mutations

in *ZmSCR1* and *ZmSCR1h* appear to perturb ground-tissue formation in maize roots. In *Arabidopsis* only a single layer of ground-tissue is formed in the *Atscr* mutant, with features of both the endodermis and cortex. However, in the *Zmscr1-1; Zmscr1h* double mutant this does not appear to be the case, with multiple cortical cell layers still present, no obvious endodermis and an apparent reduction in the casparian strip (which is still present in the *Atscr* mutant). Despite apparent differences in these phenotypes, variation in WT root structure between dicots and monocots, alongside the absence of cell-specific markers for cortex and endodermal cells in maize, make it difficult to assess whether these phenotypes are in fact analogous. It is clear, however, that in common with their roles in leaf development, *ZmSCR1* and *ZmSCR1h* function redundantly in maize roots and do not appear to have functionally diverged from each other in this context.

4.3.4. Functional validation of a role for *ZmSHR2* in Kranz development will require concurrent loss-of-function analysis of *ZmSHR2h*

Unlike *ZmSCR1* and *ZmSCR1h*, no functional evidence was obtained that *ZmSHR2* plays a role in the development of Kranz anatomy. However, *ZmSHR2* has a maize-specific homeolog duplicate (*ZmSHR2h*). Unfortunately, no transposon insertion was available in *ZmSHR2h*, meaning that no assessment of potential redundancy between these genes could be made. This is disappointing, particularly given that the double *Zmscr1-1; Zmscr1h-1* mutant displayed far more severe phenotypes than either of the respective single mutants. To test whether ZmSHR2 and ZmSHR2h might function redundantly in Kranz development, an RNAi line is currently being generated that will knock down expression of both *ZmSHR2* and *ZmSHR2h* simultaneously.

In addition to the maize-specific duplication of *ZmSHR2*, there was an additional SHR duplication at the base of the monocots. In combination, this has resulted in

three maize *SHR* genes that are all orthologous to *AtSHR*. It is not clear to what extent these distinct monocot *SHR1* and *SHR2* clades have diverged in function. Although both *ZmSHR1* and *ZmSHR2* were allocated to the same expression clusters in early foliar and husk leaf development by Wang *et al.* (2013b), in situ expression analysis has subsequently revealed distinct expression domains for both genes during early foliar leaf development (Olga Sedelnikova, personal communication), and *ZmSHR1,* but not *ZmSHR2* is expressed in mature maize BS cells (Li et al. 2010; Chang et al. 2012). Elucidation of the operation of the SCR/SHR pathway during Kranz anatomy formation must involve full consideration of all three maize *SHR* genes.

4.3.5. *ZmSCR1* and *ZmSCR1h* as confirmed Kranz regulators: prospects for C_4 rice engineering

Given that *ZmSCR1* and *ZmSCR1h* are among the first confirmed developmental regulators of Kranz anatomy, they are of considerable interest to attempts to engineer the C_4 pathway into rice. It was hoped that overexpression of Kranz regulators in rice might yield a shift towards more C_4-like anatomy. Previous work has indicated that rice can tolerate changes in vein density, albeit as a result of changes in M cell size rather than cell number (Smillie et al. 2012; Feldman et al. 2014). However, overexpression of *ZmSCR1* in rice did not result in any changes in leaf anatomy. This does not necessarily mean that *ZmSCR1* will not be useful in engineering Kranz anatomy, however, these results likely point to the need to introduce multiple genes simultaneously in specific regulatory networks if rice leaf anatomy is to be precisely manipulated. Given that elucidation of the components of the Kranz regulatory network is still in its infancy, this goal remains a long way off.

Chapter 5: The role of maize IDD transcription factors during leaf development

5.1. Introduction

The INDETERMINATE DOMAIN (IDD) genes encode a highly conserved subfamily of C2H2 transcription factors (Colasanti et al. 2006). Roles in a diverse range of biological processes have been described for the *Arabidopsis* IDD transcription factors, including gravitropism (Morita et al. 2006; Tanimoto et al. 2008; Wu et al. 2013), sugar metabolism (Seo et al. 2011), auxin biosynthesis and transport (Cui et al. 2013), and germination (Feurtado et al. 2011). In addition, a number of IDD genes have been implicated in the regulation of cellular patterning. For example*, AtIDD4* and *AtIDD11* are well established downstream targets of the HDZIPIII and KANADI transcription factors, both of which regulate ad/abaxial patterning in *Arabidopsis* shoots (Brandt et al. 2012; Reinhart et al. 2013). Furthermore, the IDD genes *JACKDAW (AtJKD), BALDIBIS (AtBIB), MAGPIE (AtMGP)* and *NUTCRACKER (AtNUC)* regulate cellular patterning in *Arabidopsis* roots via modulation of the SCR/SHR pathway (Welch et al. 2007; Ogasawara et al. 2011; Long et al. 2015b). Despite the founding member of the IDD subfamily being the maize floral transition regulator *INDETERMINATE1* (*ZmID1)* (Kozaki et al. 2004; Colasanti et al. 2006), however, little else is known about the function of the IDD genes in maize. Since the discovery of *ZmID1*, the only additional maize IDD mutant published is *NAKED ENDOSPERM (Zmnkd)*, which was identified based on a distinct shrivelled kernel phenotype caused by impaired aleurone cell patterning (Yi et al. 2015). The *Zmnkd* phenotype was subsequently found to result from transposon insertions in the duplicate genes *ZmNKD1* and *ZmNKD2* (Yi et al. 2015). Given the range of important developmental processes regulated by IDD transcription factors in

Arabidopsis, there is a need for further research into the roles of these genes in monocots.

Three IDD genes were identified in the RNA-seq analysis of both foliar and husk early leaf development in maize, which was designed to detect candidate regulators of Kranz anatomy (Wang et al. 2013b; Fouracre et al. 2014). One of these genes was the previously identified *ZmNKD1*, whereas the other two, named *ZmRVN1* (previously *ZmIDD2*) and *ZmIDD14* had no known function in maize. All three were identified initially based on their allocation to the 'foliar ascending 2 (FA2)' and 'husk neutral (HN)' expression clusters (Wang et al. 2013b). The FA2 cluster was designed to include genes whose expression was consistent with roles in vascular and Kranz patterning. Genes were included if they were expressed significantly more highly in the FP3/4 sample (plastochron-3 and 4 primordia) than in FP1 (shoot apical meristem plus plastochron-1 and 2 primordia), with this expression maintained in FP5 (plastrochron-5 primordia) (Wang et al. 2013b). The HN cluster included genes not differentially expressed across any of the equivalent three stages in husk primordia. The majority of genes present in the FA2 cluster had no known function, which was expected as little is known about the precise regulation of vein and BS/M cellular patterning in maize. The expression profile of all three of these IDD genes was, therefore, consistent with a role in Kranz patterning.

There are a number of additional sources of evidence that support the hypothesis that IDD genes may play a role in the development of Kranz anatomy. First, the SCR/SHR pathway has emerged as an important regulator of Kranz development (Slewinski et al. 2012, 2014, Chapter 4). As a number of IDD transcription factors are known to modulate the SCR/SHR pathway in the *Arabidopsis* root, the three maize IDD genes identified are good candidates for regulators of Kranz cellular patterning. Second, *ZmNKD1* is preferentially expressed in the M cells of mature leaves (Li et al. 2010; Chang et al. 2012), and *in situ* hybridisation has revealed that this expression

pattern is similar in developing leaf primordia (Olga Sedelnikova, personal communication). Conversely, *ZmRVN1* is preferentially expressed in the BS of mature maize leaves (Li et al. 2010; Chang et al. 2012), albeit at low levels. In addition, there is some evidence for preferential BS expression in developing leaf primordia (Olga Sedelnikova, personal communication). These expression patterns are consistent with roles for both of these genes in modulating the activity of either *ZmSCR1* or *ZmSHR1*, in order to correctly specify BS and M cell differentiation. Finally, *ZmNKD1* is already known to play a role in cellular patterning, with *Zmnkd1; Zmnkd2* mutant kernels exhibiting multiple cell layers surrounding the endosperm (Yi et al. 2015). These cells only sporadically acquire aleurone cell identity, in contrast to the normal single-cell aleurone layer found in the WT (Yi et al. 2015). Taken together, the available evidence points to a role for these IDD transcription factors in Kranz cell patterning, however, functional validation is necessary to test this.

The aim of this chapter was to investigate the function of maize IDD transcription factors identified as candidate Kranz regulators, in both maize and rice. This was achieved through the analysis of maize transposon insertion mutants, and by overexpression in rice using a range of different promoters. No transposon insertions were identified for *ZmIDD14*, and a preliminary analysis indicated that constitutive expression of *ZmIDD14* in rice did not perturb leaf anatomy (Appendix 9). Therefore, *ZmIDD14* was deprioritised for further characterisation. In contrast, maize insertions were available for *ZmNKD1* and *ZmRVN1*, and these genes were thus characterised in detail.

5.2. Results

5.2.1. *ZmNKD1* and *ZmRVN1* have maize-specific duplicates

Prior to functional validation, a maximum likelihood phylogenetic tree was constructed in order to better understand the relationships underpinning the IDD

genes (Figure 5.1A, Appendix 8). *ZmNKD1* was already established to have a duplicate pair in maize, termed *ZmNKD2* (Yi et al. 2015). The rice and sorghum genomes were previously found to contain only one *NKD* gene, whereas no orthologous *NKD* genes were found in dicots (Yi et al. 2015). This phylogenetic structure was confirmed here, with only single copy *NKD* genes present in the genomes of sorghum, *Setaria italica* and rice, and no orthologs found in either *Arabidopsis* or tomato (Figure 5.1B). This indicates that *ZmNKD1* and *ZmNKD2* resulted from a duplication after the divergence of maize and sorghum. This gene-pair was not included in the homeolog dataset compiled using synteny across the maize genome, however, the genes reside on chromosomes 2 and 10, within regions of known shared synteny (Wei et al. 2007; Schnable et al. 2011). It is possible, therefore, that *ZmNKD1* and *ZmNKD2* are a homeolog gene-pair resulting from the recent maize whole genome duplication.

The IDD phylogeny supports the previous inclusion of *ZmRVN1* in the homeolog dataset, as it was found to have a duplicate specific to maize (Figure 5.1C). *ZmRVN1* and *ZmRVN1h* are orthologous to *AtIDD7* and *AtIDD11* (Figure 5.1C). Although neither *AtIDD7* or *AtIDD11* have known functions, both have *REVOLUTA (REV)* binding sites upstream (Brandt et al. 2012), and *AtIDD11* expression was found to be down-regulated by REV (Reinhart et al. 2013). Mutations in *REV* are known to affect vascular and ad/abaxial patterning (Talbert et al. 1995; Emery et al. 2003; Prigge et al. 2005), which implicates AtIDD7 and AtIDD11 (and possibly *ZmRVN1* and *ZmRVN1h*) in both of these patterning processes.

Interestingly, the predicted gene model of *ZmRVN1h* is much shorter than *ZmRVN1*, and has homology only to the third exon of *ZmRVN1* (Figure 5.2A,B). To investigate whether this was due to an incorrectly assembled gene-model, or reflected a real difference between the two genes, an alignment was constructed between the *ZmRVN1* gene sequence, the predicted shorter *ZmRVN1h* gene sequence, and the

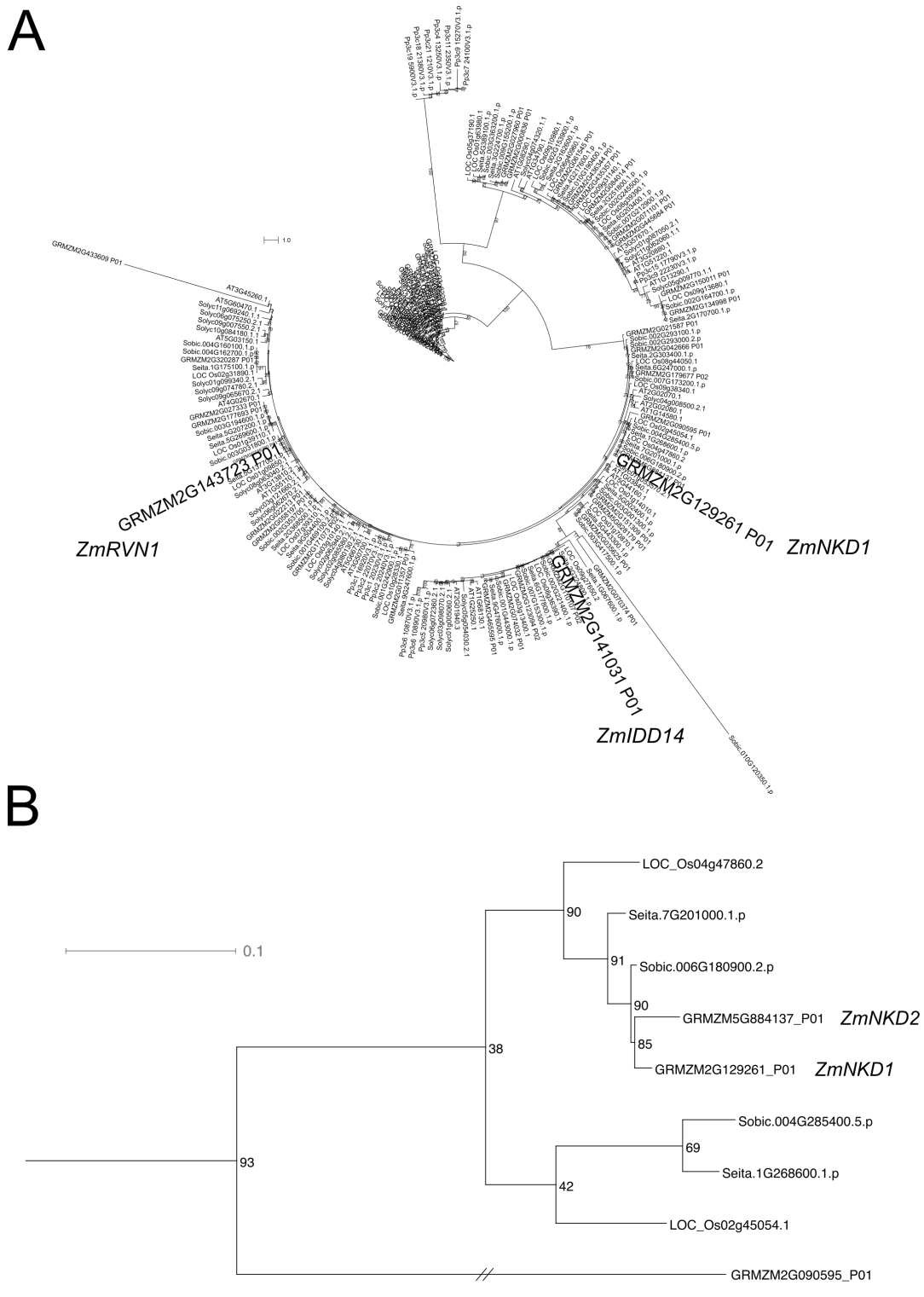


Figure 5.1 Maximum likelihood phylogenetic tree of the IDD transcription factors (A and B). Part (C) and figure legend are on the next page.

C

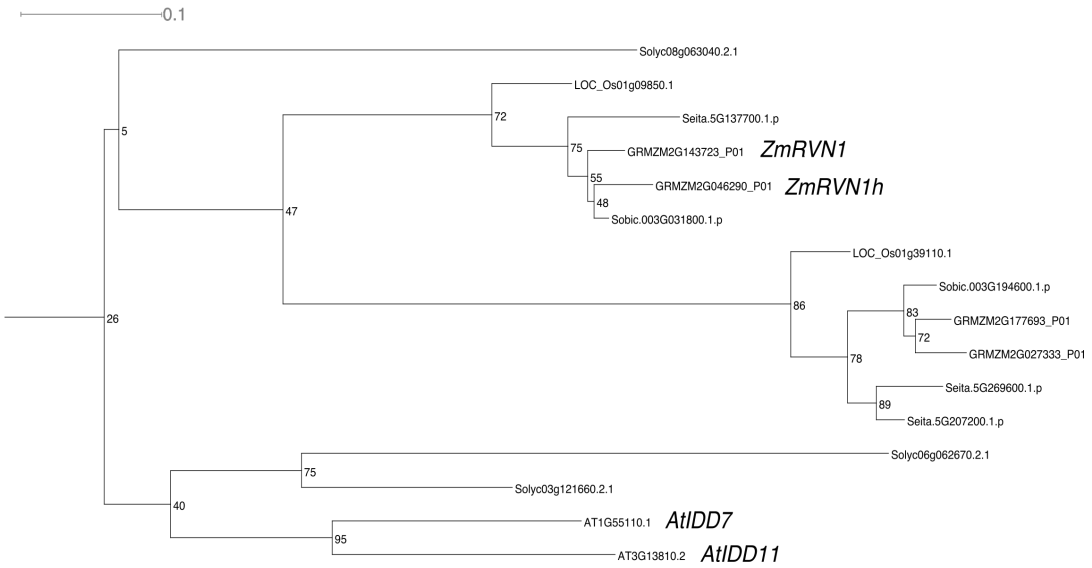
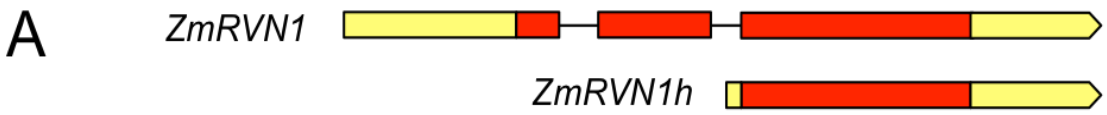


Figure 5.1 Maximum likelihood phylogenetic tree of the IDD transcription factors. A) Complete phylogenetic tree, with the position of *ZmNKD1, ZmRVN1* and *ZmIDD14* highlighted. **B)** The *NKD* clade of the tree. One branch length (indicated by a double diagonal line) has been halved for display purposes. **C)** The *RVN* clade of the tree. IDD sequences were acquired using the *ZmNKD1* and *ZmNKD2* protein sequences as a query for a BLASTp search (evalue = 1e-3) against the proteomes of all seven species. The retrieved sequences were aligned using MergeAlign, and the alignment subsequently trimmed using trimAL. A maximum likelihood tree was constructed using RaxML. Scale bars indicate the number of substitutions per site. Bootstrap values are displayed on each branch. Maize, sorghum, *Setaria italica,* rice, *Arabidopsis*, tomato and *Physcomitrella patens* gene accessions begin GRM, Sobic, Seita, LOC_Os, AT, Solyc and Ppat respectively. An electronic version of this tree is included in Appendix 8.

ZmRVN1h locus including the region 2kb upstream of the predicted *ZmRVN1h* start site (Figure 5.2B). This revealed that there is considerable sequence homology between the upstream *ZmRVN1h* region and the first two exons of the *ZmRVN1* gene sequence. Furthermore, an open reading frame homologous to that of *ZmRVN1* could be constructed for *ZmRVN1h*, with the same start and stop codons, and correctly positioned intron donor and acceptor sites. To test whether this longer gene model was real and expressed, the sequenced reads from the RNA-seq dataset used to detect candidate regulators Kranz anatomy were interrogated using Integrative Genomics Viewer (IGV) (Thorvaldsdóttir et al. 2013; Wang et al. 2013b). Reads from both the FP1/2 and FP3/4 samples of early foliar leaf development were detected for the shorter predicted *ZmRVN1h* sequence (Figure 5.2C). However, additional reads were present upstream of the predicted start codon, and these mapped to the intron donor and acceptor sites from the new hypothesised longer *ZmRVN1h* model. Thus, a new gene model has been constructed for *ZmRVN1h* that exhibits high homology and gene structure with *ZmRVN1* (Figure 5.2D). It is not clear whether the ORF used to construct the truncated *ZmRVN1h* model represents a distinct transcript, or is simply a mistake in gene-model prediction. In either case it appears that *ZmRVN1h* is expressed at very low levels in leaf primordia, and an attempt to amplify the transcript representing the revised gene model from leaf primordia enriched cDNA failed (data not shown). Further experimental work is necessary to elucidate the precise *ZmRVN1h* gene model.

5.2.2. Maize transposon insertion alleles were identified for *ZmNKD1*, *ZmNKD2* and *ZmRVN1*

To establish whether *ZmNKD1* or *ZmRVN1* function in Kranz anatomy development, loss-of-function maize mutants were obtained. Hypomorphic and null mutant alleles were available for both *ZmNKD1* and *ZmNKD2.* The *Zmnkd1-R* and *Zmnkd2-R*



B

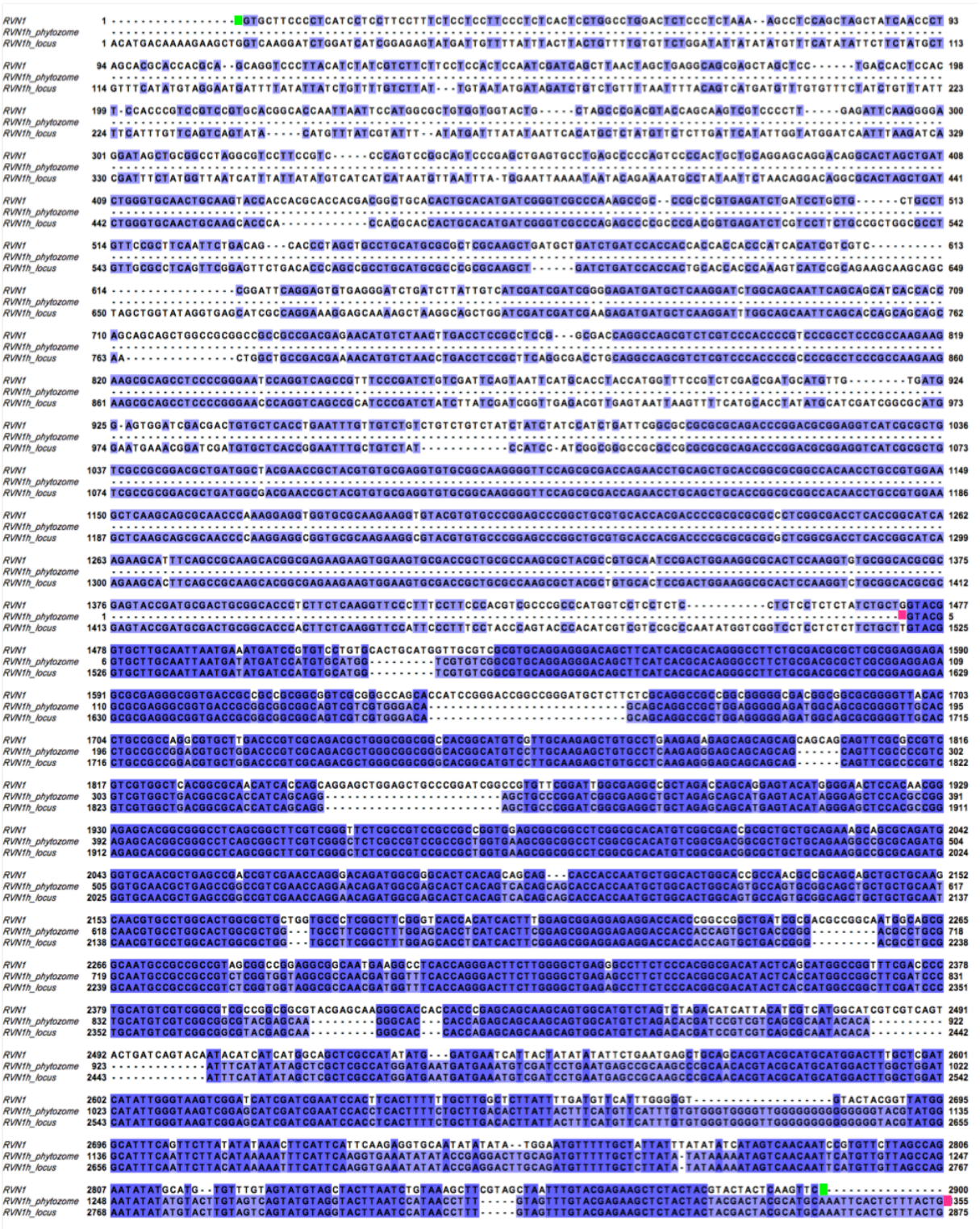


Figure 5.2 Revision of *ZmRVN1h* gene model. Figure legend is on the next page.

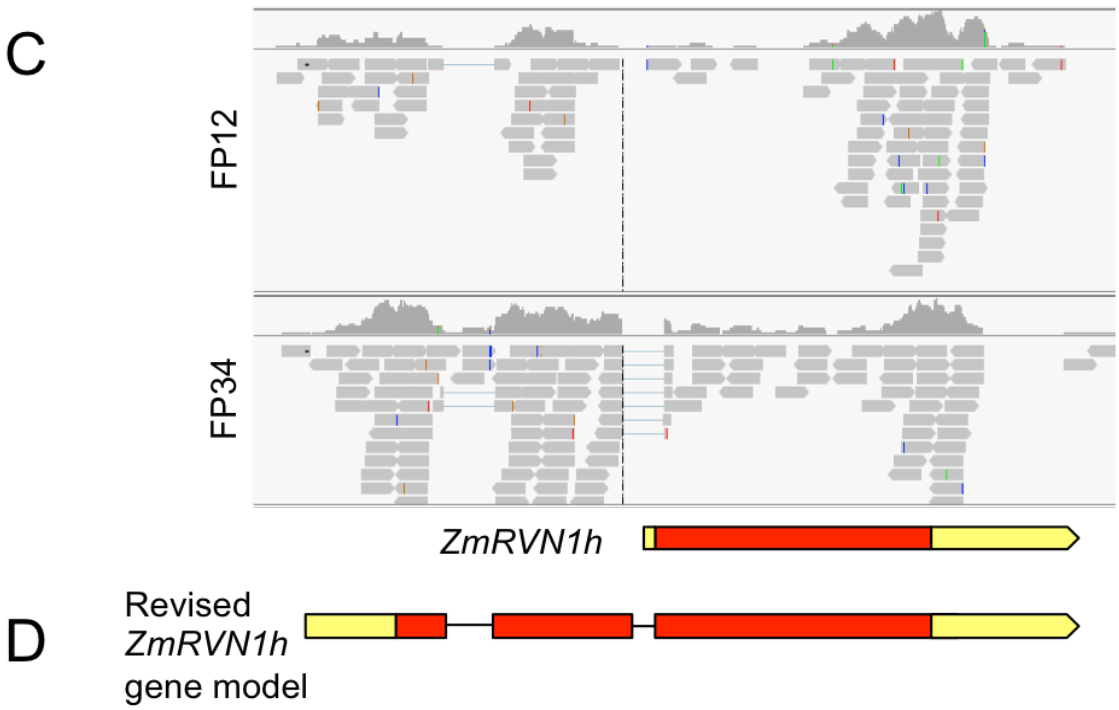


Figure 5.2 Revision of *ZmRVN1h* gene model. A) Cartoon alignment of *ZmRVN1* and *ZmRVN1h* predicted gene models. Sequences taken from Phytozome11. The 5' end is displayed on the left, yellow regions refer to the UTRs, red to the CDS and black lines to introns. **B)** DNA sequence alignment using Clustal Omega of *ZmRVN1* (top), predicted phytozome *ZmRVN1h* gene model (middle), and the *ZmRVN1h* locus (bottom), including region upstream of the predicted gene sequence. Green squares mark the start and finish of the *ZmRVN1* gene model. Pink squares mark the start and finish of the predicted *ZmRVN1h* gene sequence. Sequence similarity is indicated by blue highlighting, with darker blue indicating higher similarity. **C)** Mapped reads for *ZmRVN1h* from two RNA-seq samples. Both samples were from maize early foliar leaf development (Wang et al. 2013b). The top panel shows reads from the FP1/2 sample, the bottom panel shows reads from the FP3/4 sample. In both cases there are multiple reads upstream of the predicted *ZmRVN1h* gene model. Coloured lines in the reads indicate mismatches. **D)** A revised *ZmRVN1h* gene model, with intron donor and acceptor sequences used to place intron sites.

hypomorphic alleles were previously identified in a mixed genetic background. When alleles for both genes are homozygous, kernels fail to develop properly, with altered aleurone cell division leading to a distinctive shrivelled appearance (Yi et al. 2015). The responsible lesions have not been confirmed in either case, however, the *Zmnkd1-R* mutant phenotype is suspected to result from a C to T base change that leads to a histidine to tyrosine conversion (Yi et al. 2015) (Figure 5.3). This histidine is part of the first zinc finger of the C2H2 motif and is highly conserved across other IDD genes. In the case of *Zmnkd2-R* the causal lesion is suspected to be a *copia*-like retrotransposon present in the 5'UTR of the gene (Yi et al. 2015) (Figure 5.3). The two suspected null alleles contain *Ds* transposon insertions. In *Zmnkd1-Ds* the element is inserted in the fourth exon of the gene, within the CDS, whereas in *Zmnkd2-Ds2* the element is inserted in an intron within the 5'UTR of the gene (Figure 5.3). Although only *ZmNKD1* emerged from previous transcriptome analysis as a candidate regulator of Kranz anatomy, *ZmNKD2* is also expressed during early leaf development (albeit at lower levels) and preferentially expressed in mature M cells (Li et al. 2010; Chang et al. 2012). Given that *ZmNKD1* and *ZmNKD2* are known to function redundantly in aleurone cell patterning (Yi et al. 2015), it was hypothesised that to reveal a potential role for *ZmNKD1* in Kranz development it would be necessary to analyse the double rather than single *Zmnkd* mutants. Both *Zmnkd1-R; Zmnkd2-R* and *Zmnkd1-Ds; Zmnkd2-Ds* double mutant lines were therefore obtained from Phil Becraft (Iowa State University). In the case of *ZmRVN1*, four *Mutator* transposon insertions were identified from the UniformMu (*Zmrvn1-1* and *Zmrvn1-2*) and Mu-Illumina (*Zmrvn1-3* and *Zmrvn1-4*) projects (Appendix 2). Three of these insertions were in the second exon of the CDS, with the fourth being in the 5'UTR of the gene. One allele (*Zmrvn1-4*) was prioritised for functional validation due to the transposon insertion site in the CDS, and the availability of homozygous seed (Figure 5.3).

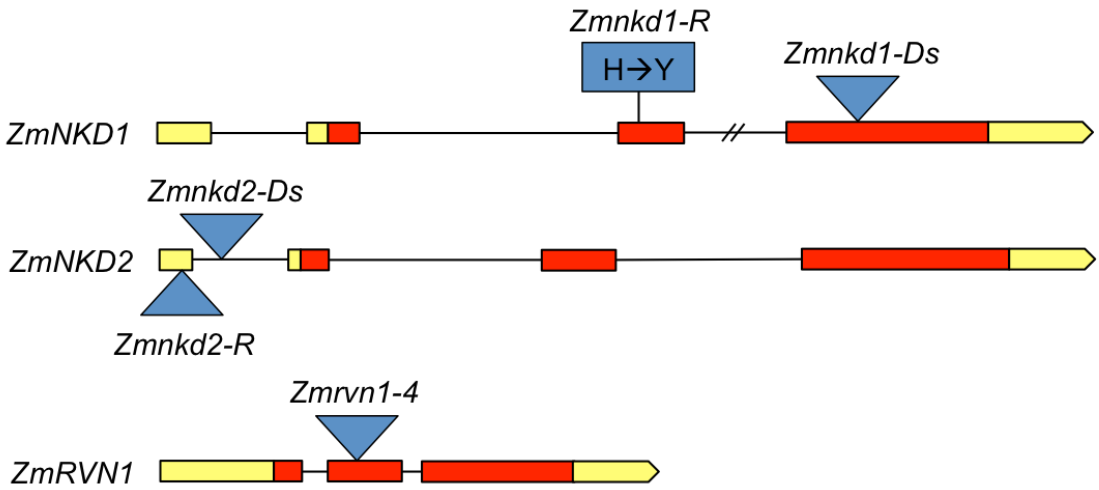


Figure 5.3 Schematic of transposon insertions into IDD transcription factors. Gene models are displayed for each transcription factor: in all cases the 5' end is displayed on the left, yellow regions refer to the UTRs, red to the CDS and black lines to introns. Blue triangles represent transposon insertions (not to scale), the blue box indicates an amino acid substitution that results from the suspected causal mutation in the *Zmnkd1-R* allele. Insertion positions were taken from Yi *et al.* (2015) (for *Zmnkd1-R/Ds* and *Zmnkd2-R/Ds*), or from the Mu-Illumina project (for *Zmrvn1-4*) (Alice Barkan, personal communication). Allele names are displayed above the transposon. All gene models are to scale, except the third intron of *ZmNKD1*, which has been truncated (indicated by a double diagonal line) for display purposes.

5.2.3. The role of *ZmNKD1* and *ZmNKD2* in maize

5.2.3.1. *Zmnkd1; Zmnkd2* double mutants exhibit alterations in leaf morphology

Seed from both the hypomorphic *Zmnkd1-R; Zmnkd2-R* and the null *Zmnkd1-Ds; Zmnkd2-Ds* lines were grown in the greenhouse for phenotypic analysis. Mutant seed were grown alongside WT plants from the same inbred background for comparison (B73 for *Zmnkd1-R; Zmnkd2-R*, W22 for *Zmnkd1-Ds; Zmnkd2-Ds*). Plants were genotyped and phenotypic characterisation was undertaken 38 days after planting (Figure 5.4). Plant morphology appeared normal (Figure 5.5), although the previously reported reductions in growth and delayed flowering were apparent at later developmental stages. Both double mutants exhibited slightly shorter and narrower leaves than their WT counterparts (Figure 5.6A,B), which is likely a reflection of the slightly reduced growth previously reported. The effect on leaf length is similar in both comparisons, however, the effect on leaf width was more pronounced in the *Zmnkd1-Ds; Zmnkd2-Ds* - W22 WT comparison (Figure 5.6B). This is likely to be due to the combined effects of W22 WT leaves generally being wider than B73, and the *Zmnkd-Ds* alleles being likely nulls. These alterations in leaf morphology confirm the role of *ZmNKD1* and *ZmNKD2* in general plant growth and development.

5.2.3.2. *Zmnkd1; Zmnkd2* double mutants exhibit more frequent occurrences of extra BS cell formation

To determine whether leaf patterning was perturbed in *Zmnkd* double mutants, fresh sections were taken from the midpoint along the proximal/ distal axis of fully expanded leaf 6, 38 days after planting (Figure 5.7). No difference in vein density was found between WT and either double mutant background (Figure 5.8A), however, there was a trend in both cases for significantly reduced numbers of intermediate veins separating two lateral veins (Figure 5.8B). This trend is only

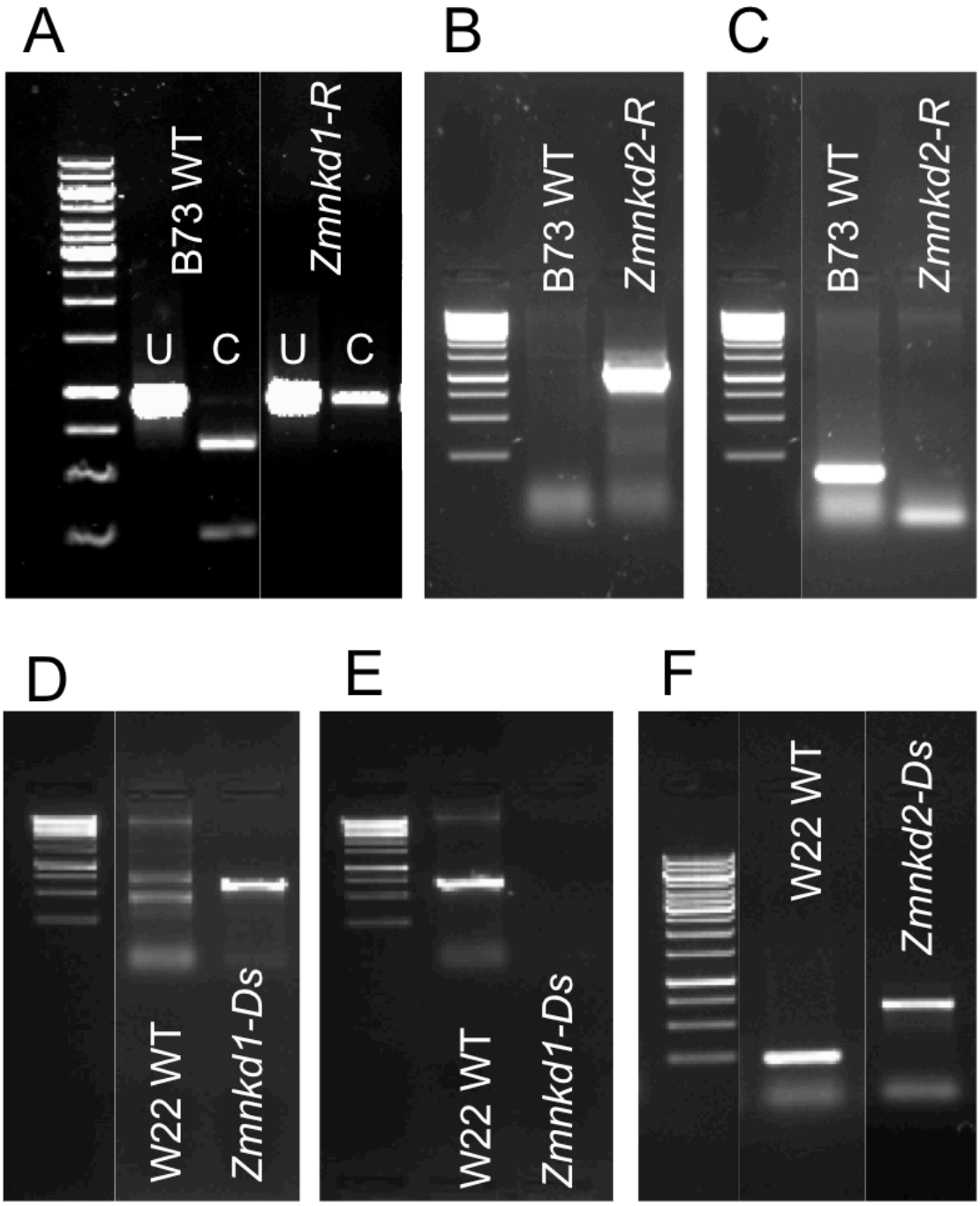


Figure 5.4 Gel images showing PCR amplified products from WT and homozygous *Zmnkd* mutant alleles. Figure legend is on the next page.

Figure 5.4 Gel images showing PCR amplified products from WT and homozygous *Zmnkd* mutant alleles. **A)** A single nucleotide polymorphism (SNP) renders the PCR product of the *Zmnkd1-R* allele resistant to cleavage by the restriction enzyme *Hin*dIII, whereas the B73 allele is cleaved. U = uncut by *Hin*dIII, C = cut by *Hin*dIII. **B)** PCR amplification from the *copia*-like retrotransposon present in *Zmnkd2-R*. No PCR product is obtained from B73 WT gDNA. **C)** PCR amplification using WT primers that flank the insertion site of the *copia*-like retrotransposon present in *Zmnkd2-R*. No PCR product is obtained from homozygous *Zmnkd2-R* gDNA, likely because the extension time was insufficient to amplify across the transposon. **D)** PCR amplification from the *Ds* transposon present in *Zmnkd1-Ds*. A clear amplicon of the correct size is only found from *Zmnkd1-Ds* gDNA. **E)** PCR amplification using WT primers that flank the insertion site of the *Ds* transposon present in *Zmnkd1-Ds*. No PCR product is obtained from homozygous *Zmnkd1-Ds* gDNA, likely because the extension time was insufficient to amplify across the transposon. **F)** PCR amplification using WT primers that flank the insertion site of the *Ds* transposon present in *Zmnkd2-Ds*. If the *Zmnkd2-Ds* allele is present, a larger amplicon is observed than in W22 WT. In all cases the smallest visible fragment of the DNA ladder is 250bp, and the brighter fragment three above that is 1kb.

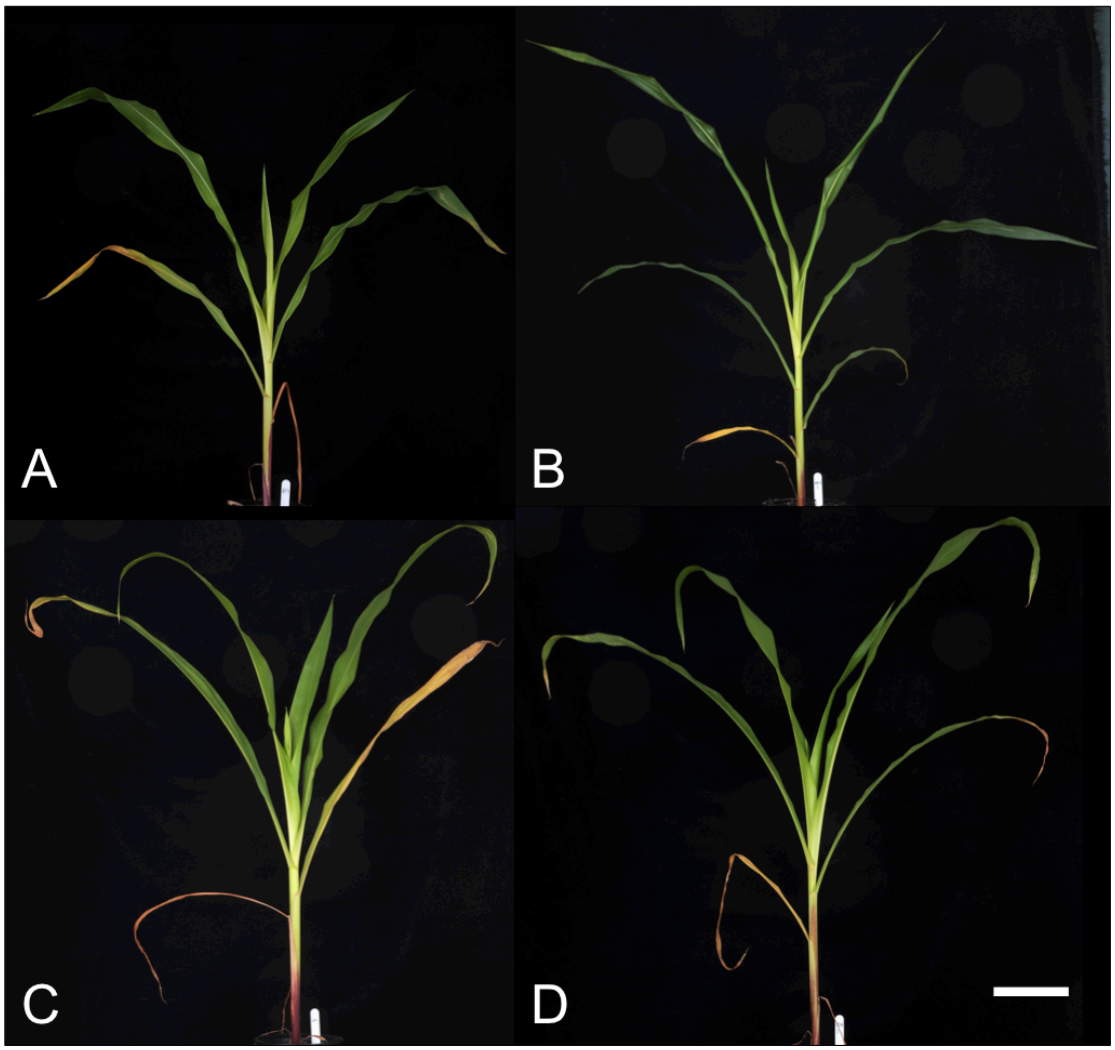


Figure 5.5 Whole plant phenotype of *Zmnkd* mutants and WT plants. **A)** B73 WT **B)** *Zmnkd1-R; Zmnkd2-R* **C)** W22 WT **D)** *Zmnkd1-Ds; Zmnkd2-Ds*. White scale bar = 10cm. Scale is the same in (A-D). Images were taken 38 days after planting.

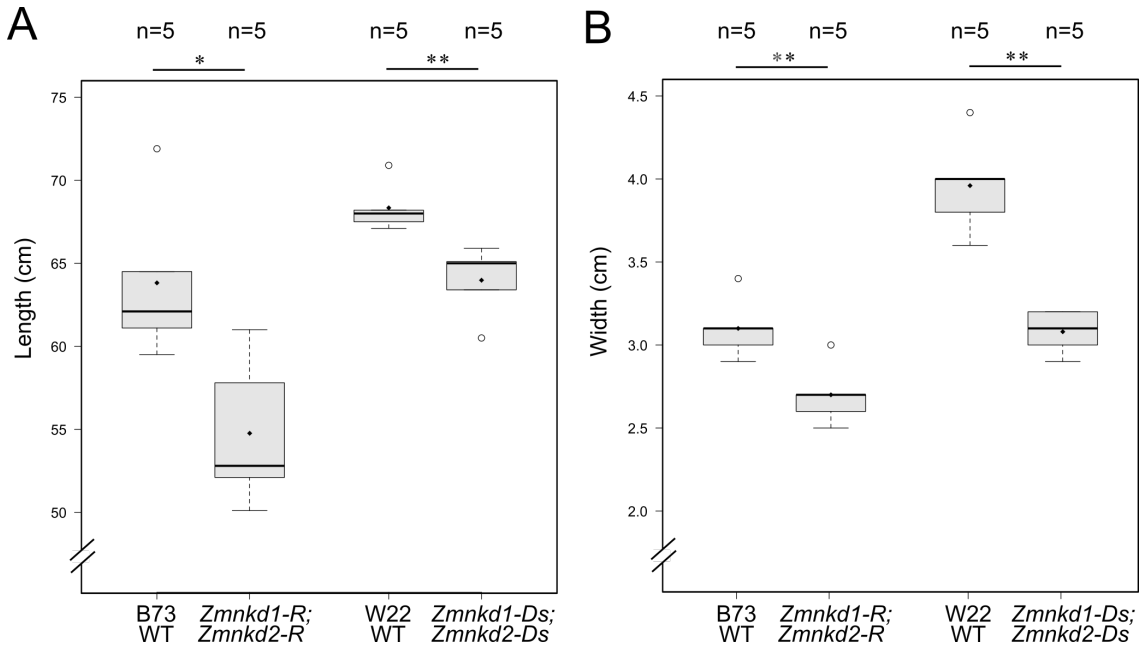


Figure 5.6 Boxplots displaying leaf morphology traits for *Zmnkd* mutants and WT plants. A) Fully expanded leaf 6 length. Leaf length measured from the ligule to the tip of the blade. **B)** Fully expanded leaf 6 width. Leaf width taken at the mid-point of the proximal/distal axis. *Zmnkd1-R; Zmnkd2-R* was previously introgressed into the B73 inbred background, and *Zmnkd1-Ds; Zmnkd2-Ds* into the W22 inbred background (Phil Becraft, personal communication), so mutant comparisons were made to the appropriate WT line. Statistical significance calculated by t-tests between WT and *Zmnkd* mutant plants are indicated by asterisks above the plot (* = $p \leq 0.05$, ** = $p \leq 0.01$). Sample sizes are indicated by n= above plots. Means are indicated by black diamonds. Outliers are indicated by empty circles. Breaks in y-axis scale are indicated by a double diagonal line.

statistically significant in the *Zmnkd1-Ds; Zmnkd2-Ds* - W22 WT comparison, and is most likely a result of the wider W22 leaves (as C_4 leaf width increases extra intermediate veins are specified between laterals) rather than a vein patterning effect *per se*. Finally, the ratio of rank-1 to rank-2 intermediate veins was significantly higher in *Zmnkd1-R; Zmnkd2-R* relative to B73 WT, but not in *Zmnkd1-Ds; Zmnkd2-Ds* relative to W22 WT (Figure 5.8C). There is no obvious explanation for this, and it may simply reflect a B73 background effect or statistical artefact. Interestingly, in both double mutants there is an increased occurrence of extra BS cells forming between two lateral veins (Figure 5.7, Figure 5.9). This effect is statistically significant between W22 WT and *Zmnkd1-Ds; Zmnkd2-Ds*, but not between B73 WT and *Zmnkd1-R; Zmnkd2-R*. However, the trend is consistent in both cases. No increased occurrence of veins with no separating M cells was observed in either double mutant (data not shown).

5.2.3.3. *Zmscr1-1; Zmnkd1-R; Zmnkd2-R* triple mutants did not show more severe leaf patterning defects than single *Zmscr1-1* or double *Zmnkd1-R; Zmnkd2-R* mutants

It is intriguing that mutations in both the *ZmNKD1/2* and *ZmSCR1/1h* genes result in increased occurrences of extra BS cell formation outside of the single layer of BS cells that surround veins. Both pairs of genes are known to be preferentially expressed in the M of developing leaves (Olga Sedelnikova, personal communication), and it is known that AtSCR and various IDD genes work together in patterning the *Arabidopsis* root. In *Arabidopsis*, when the *Atscr* mutation was combined with mutations in various IDD genes, the single mutant phenotypes were subtly altered. In the case of the *Atscr; Atjkd* the double mutant abolished the ectopic periclinal divisions found in the single *jkd* mutant (Welch et al. 2007), indicating that

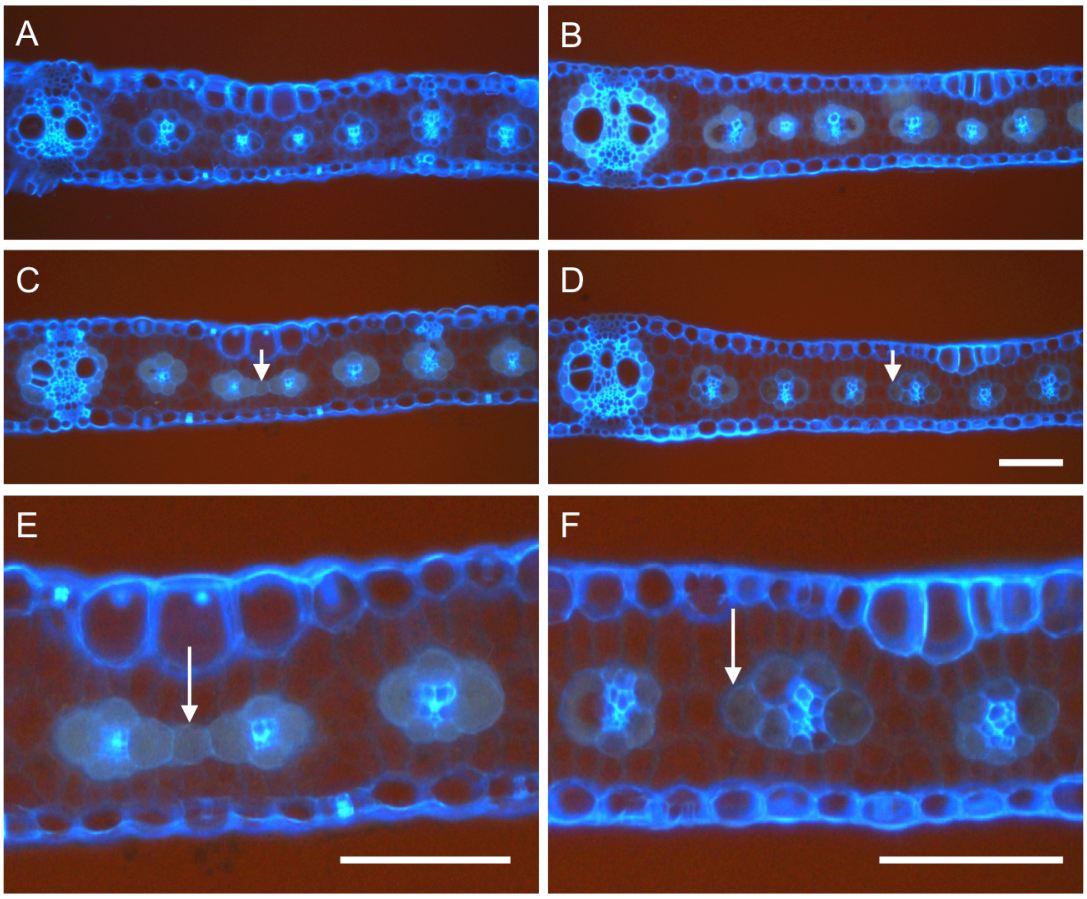


Figure 5.7 Transverse fresh sections of *Zmnkd* mutants and WT foliar leaves. A) B73 WT **B)** W22 WT **C,E)** *Zmnkd1-R; Zmnkd2-R* **D,F)** *Zmnkd1-Ds; Zmnkd2-Ds*. Fresh transverse sections were taken at the midpoint along the proximal/distal axis of fully expanded leaf 6, and imaged under UV illumination. *Zmnkd1-R; Zmnkd2-R* was previously introgressed into the B73 inbred background, and *Zmnkd1-Ds; Zmnkd2-Ds* into the W22 inbred background (Phil Becraft, personal communication), so mutant comparisons were made to the appropriate WT line. Scale bars = 100µm. Scale bar in (D) is the same in (A-C). White arrows indicate extra BS cell formation.

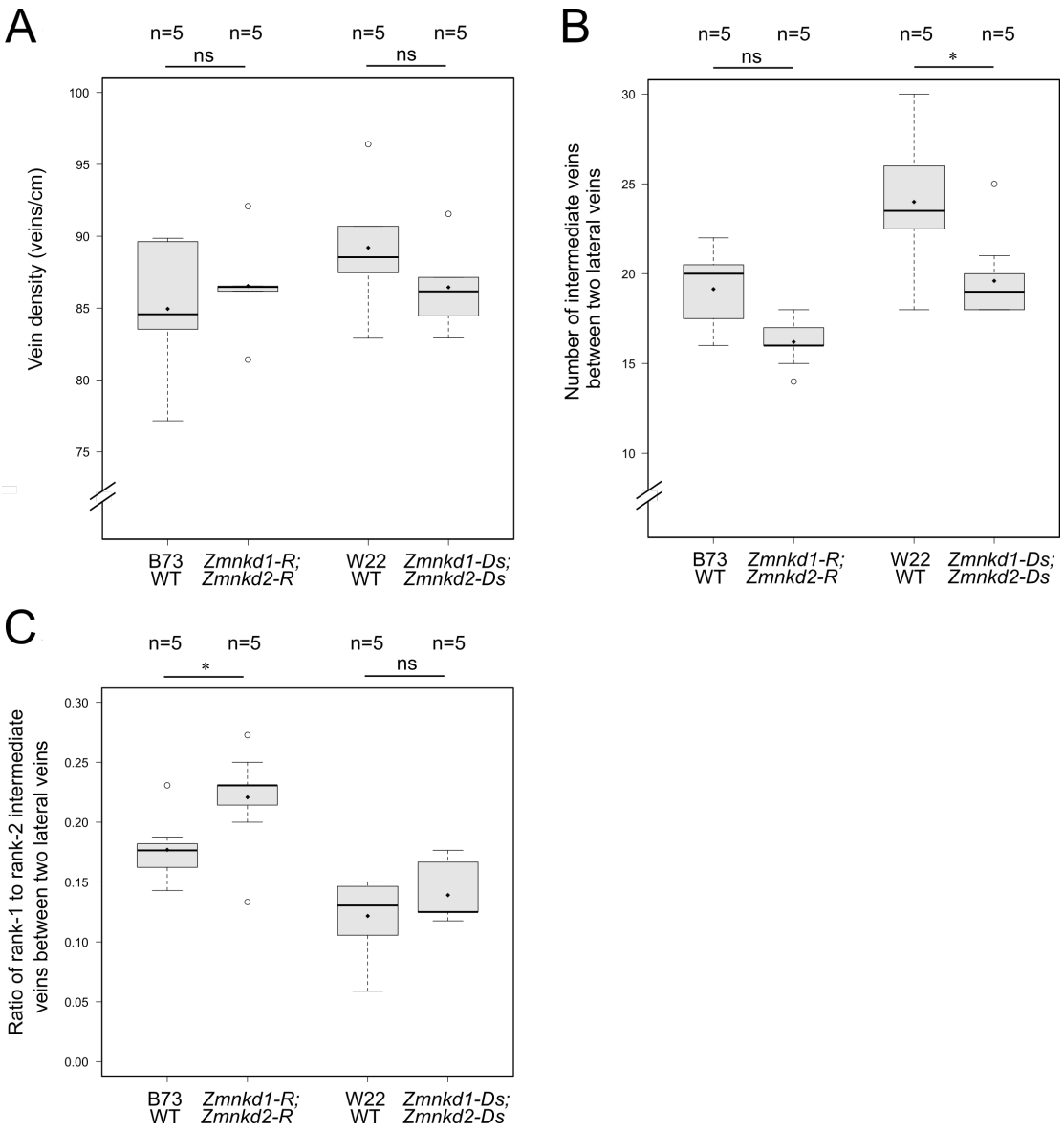


Figure 5.8 Boxplots displaying vein-patterning traits in *Zmnkd* and WT foliar leaves. Figure legend is on the next page.

Figure 5.8 Boxplots displaying vein-patterning traits in *Zmnkd* and WT foliar leaves. A) Vein density **B)** Number of intermediate veins between two lateral veins, and **C)** Ratio of rank-1 to rank-2 intermediate veins between two lateral veins. Quantification in (A-C) was undertaken using transverse sections from the mid-point of the proximal/distal axis of fully expanded leaf 6, and traits were quantified for two pairs of lateral veins adjacent to the midvein. *Zmnkd1-R; Zmnkd2-R* was previously introgressed into the B73 inbred background, and *Zmnkd1-Ds; Zmnkd2-Ds* into the W22 inbred background (Phil Becraft, personal communication), so mutant comparisons were made to the appropriate WT lines. Statistical significance was calculated by t-tests (A), mixed effect poisson regression models and subsequent likelihood ratio calculation by comparison to a null model (B), or mixed effect linear regression models and subsequent likelihood ratio calculation by comparison to a null model (C). Statistical significance between WT and *Zmnkd* mutant plants is indicated by asterisks above the plot (ns = $p > 0.05$, * = $p \leq 0.05$,). Sample sizes are indicated by n= above plots. Means are indicated by black diamonds. Outliers are indicated by empty circles. Breaks in y-axis scale are indicated by a double diagonal line.

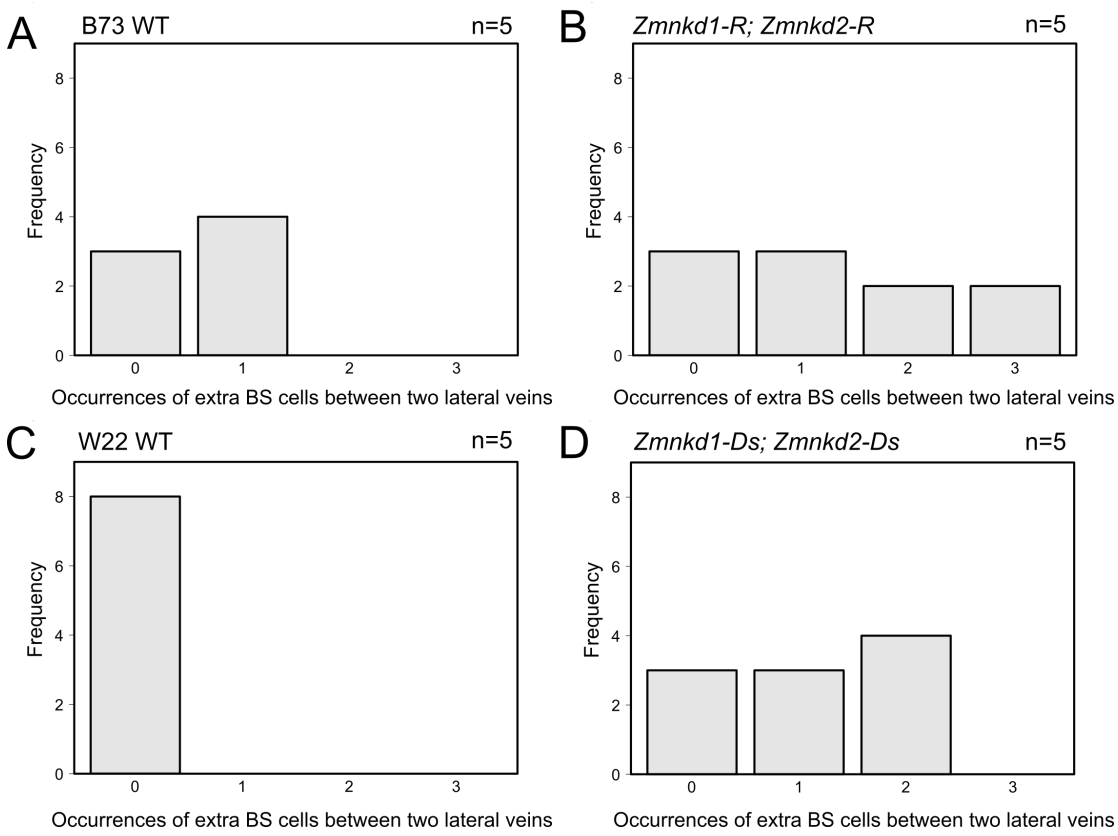


Figure 5.9 Barplots displaying the frequency of extra BS cell occurrences between two lateral veins in *Zmnkd* and WT foliar leaves. A) B73 WT **B)** *Zmnkd1-R; Zmnkd2-R* **C)** W22 WT **D)** *Zmnkd1-Ds; Zmnkd2-Ds*. Measurements in (A-D) were taken using fresh transverse sections from the midpoint along the proximal/distal axis of fully expanded leaf 6 for each biological replicate. Extra-BS occurrences were counted between two adjacent lateral veins next to the midvein, with two pairs of lateral veins used per leaf. The total number of biological replicates is indicated by n= at the top right of each plot. Genotypes are indicated at the top left of each plot. *Zmnkd1-R; Zmnkd2-R* was previously introgressed into the B73 inbred background, and *Zmnkd1-Ds; Zmnkd2-Ds* into the W22 inbred background (Phil Becraft, personal communication), so comparisons were made between (A) and (B), and between (C) and (D). In (A,B) and (C,D) statistical significance was calculated using Pearson's Chi-square test. No significant difference ($p>0.05$) was found for (A), however, in (B) $p < 0.05$.

these divisions were dependent on AtSCR. Furthermore, when the quadruple *Atscr; Atjkd; Atmgp-i; Atnuc-i* was generated, the single layer of ground-tissue had only cortex cell identity (Long et al. 2015b), in contrast to the single *Atscr* mutant in which this layer exhibited mixed cortex and endodermal cell identity (Laurenzio et al. 1996). Therefore, in order to investigate whether ZmSCR1 functions alongside ZmNKD1 and ZmNKD2 in Kranz anatomy cellular patterning, the triple *Zmscr1-1; Zmnkd1-R; Zmnkd2-R* mutant was constructed. As this mutant was produced by a cross between B73 and W22 inbred backgrounds, no detailed phenotypic quantification of the segregating F2 triple mutants was carried out. However, examination of leaf sections of these plants revealed no obvious alterations to the previously observed phenotypic perturbations in single *Zmscr1-1*, and double *Zmnkd1-R; Zmnkd2-R* mutants (Figure 5.10). Extra BS cells were observed in the triple mutant, but the frequency of occurrence was the same as in single *Zmscr1-1*, and double *Zmnkd1-R; Zmnkd2-R* mutants. Given that it is now known that ZmSCR1 and ZmSCR1h function redundantly in Kranz patterning (Chapter 4), to fully test the interaction between the *ZmSCR* and *ZmNKD* genes it will be necessary to construct the quadruple *Zmscr1-1; Zmscr1h-1; Zmnkd1-R; Zmnkd2-R* mutant, however, this was not possible during the timeframe of this project.

5.2.4. The role of *ZmRVN1* in maize

5.2.4.1. *Zmrvn1-4* single mutants do not display altered leaf anatomy

To assess whether *ZmRVN1* functions in Kranz anatomy development, plants homozygous for the *Zmrvn1-4* insertion were analysed. Plants were grown in the greenhouse and genotyped by PCR in order to validate WT and homozygous plants (Figure 5.11). To examine expression patterns, RNA was extracted from pools of WT and homozygous primordia samples, and used as template for cDNA synthesis. To test whether the transposon insertion is maintained in the *Zmrvn1-4* transcript, RT-PCR was carried out using primers that amplify from the transposon insertion. This

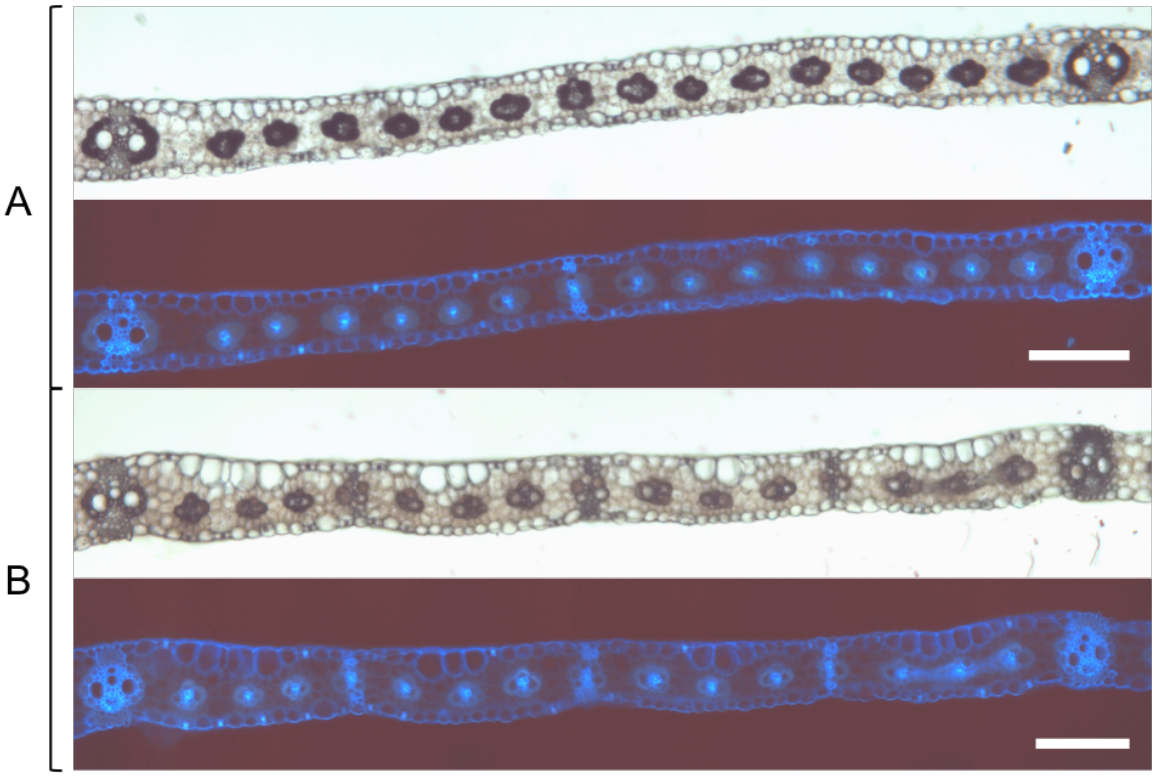


Figure 5.10 Transverse fresh sections of *Zmscr1-1; Zmnkd1-R; Zmnkd2-R* triple mutants and WT foliar leaves. A) *ZmSCR1/ZmSCR1*; *ZmNKD1/±ZmNKD1; ZmNKD2/±ZmNKD2*. **B)** *Zmscr1-1; Zmnkd1-R; Zmnkd2-R* triple mutant. In (A) and (B) sections were taken from the midpoint along the proximal/axis of fully expanded leaf 5, and imaged under brightfield and UV illumination. Scale bar = 200µm.

analysis confirmed that the *Mutator* transposon was present in the *Zmrvn1-4* mRNA (Figure 5.11). A second amplification with primers that flank the insertion, confirmed that no WT transcript was formed in the homozygous plants (Figure 5.11). The *Zmrvn1-4* allele is thus most likely a null, although protein work would be necessary to fully confirm this.

Phenotypic analysis was undertaken on plants 39 days after germination. No obvious differences in growth rate or gross morphology were observed between WT and homozygous *Zmrvn1-4* plants. To establish whether homozygous *Zmrvn1-4* plants had impaired Kranz anatomy, fresh sections were taken from the mid-section of fully expanded leaf 5 (Figure 5.12). No differences were found in terms of vein density or the number of intermediate veins separating two lateral veins (Figure 5.13A,B). There was a trend for homozygous *Zmrvn1-4* plants to exhibit a slightly higher ratio of rank-1 to rank-2 intermediate veins (Figure 5.13C), however, this was not a statistically significantly effect. As *ZmRVN1* is hypothesised to function in BS cell patterning, the number of BS cells surrounding both rank-1 and rank-2 intermediate veins was quantified (Figure 5.14). No significant difference was found, with both WT and homozygous *Zmrvn1-4* exhibiting the same median number of BS cells surrounding rank-1 (median = 7) and rank-2 (median = 5) intermediate veins. Furthermore, no increased occurrence of extra BS cell formation was observed in the *Zmrvn1-4* mutant (data not shown). Thus, no evidence was found that supported the hypothesis that *ZmRVN1* acts in BS patterning during Kranz development. Unfortunately, no insertions in the *ZmRVN1h* gene were identified in this study. A *Zmrvn1; Zmrvn1h* double mutant would be necessary to fully test the role of the *ZmRVN* genes in Kranz development.

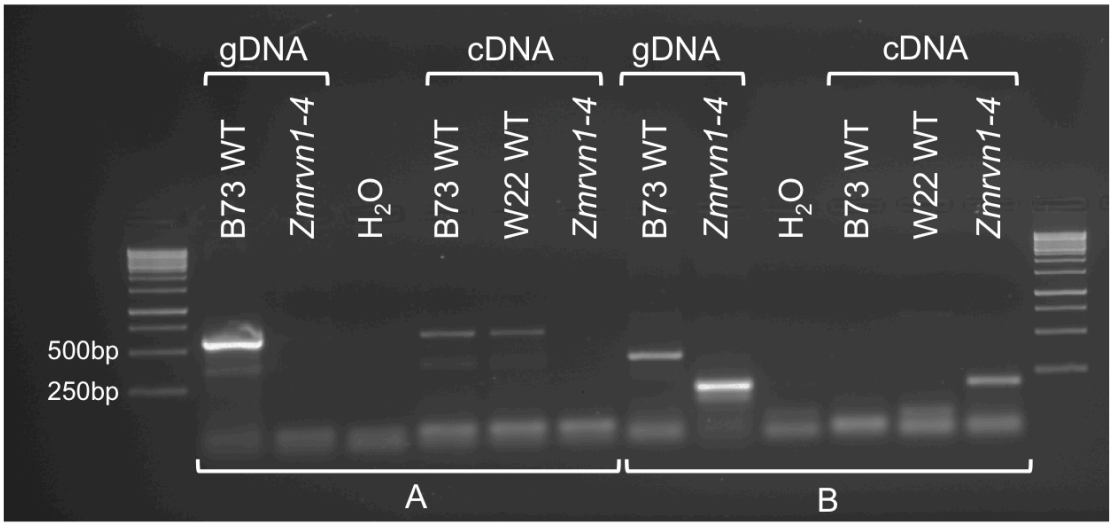


Figure 5.11 Gel image displaying genotyping and expression analysis of *Zmrvn1-4 Mutator* transposon insertion lines. A) PCR amplification from gDNA and cDNA extracted from leaf primordia, using primers flanking the *Zmrvn1-4* transposon insertion site. In WT gDNA and cDNA samples the correct amplicon (586bp) is observed, whereas in *Zmrvn1-4* no amplicon is observed (likely because extension time was insufficient to amplify across the transposon). **B)** PCR amplification from the *Zmrvn1-4 Mutator* transposon, using gDNA and cDNA extracted from leaf primordia. A larger fragment of incorrect size is seen in the B73 gDNA, however the correct amplicon (220bp) is present in homozygous *Zmrvn1-4* gDNA and cDNA.

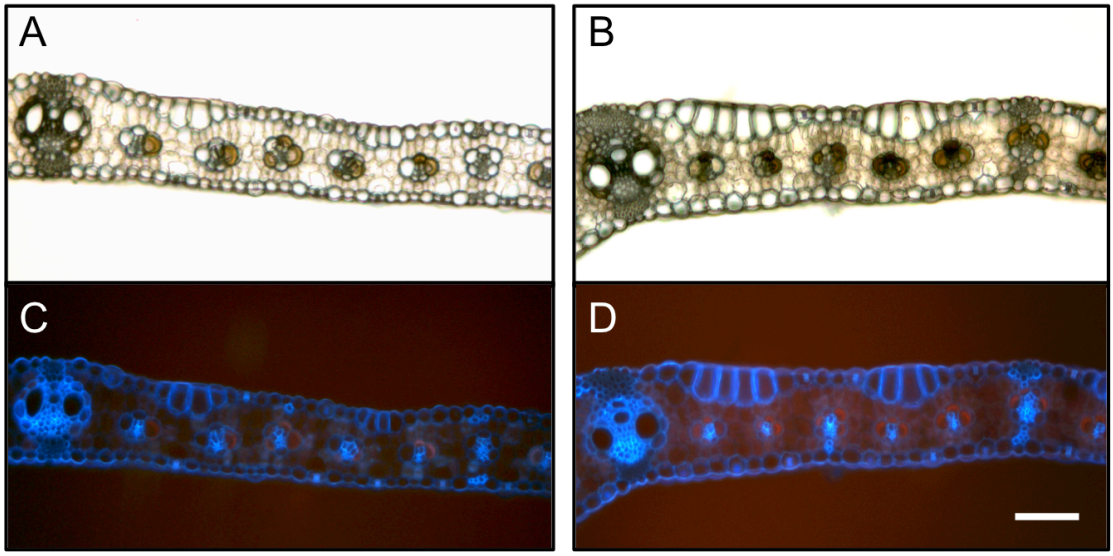


Figure 5.12 Transverse sections of *Zmrvn1-4* mutants and WT foliar leaves. A,C) WT segregating from the *Zmrvn1-4* line **B,D)** *Zmrvn1-4*. In (A-D) sections were taken from the midpoint along the proximal/distal axis of fully expanded leaf 5, and imaged under brightfield (A,B) and UV illumination (C,D). Scale bar = 100µm.

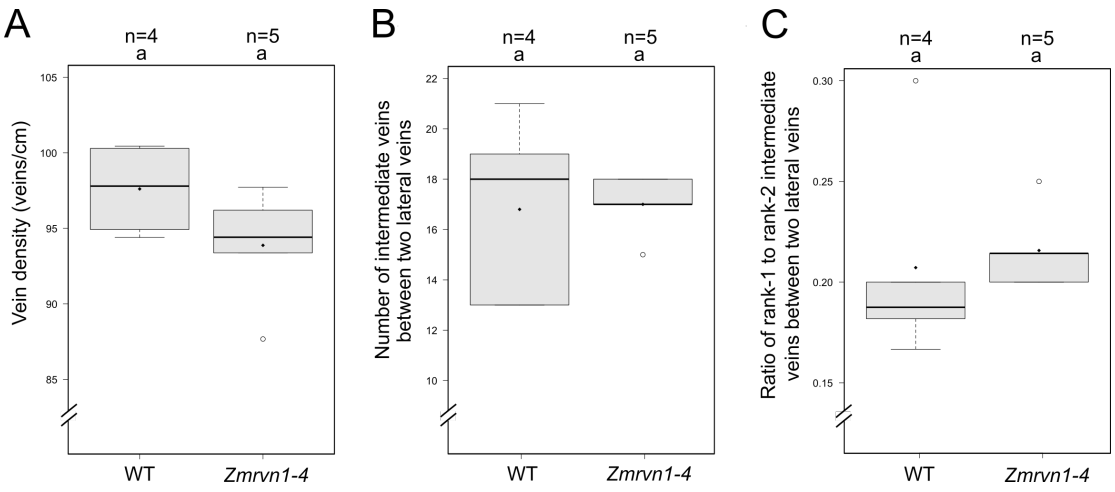


Figure 5.13 Boxplots displaying vein-patterning traits in *Zmrvn1-4* and WT foliar leaves. A) Vein density **B)** Number of intermediate veins between two lateral veins **C)** Ratio of rank-1 to rank-2 intermediate veins between two lateral veins. Measurements were taken using fresh transverse sections taken from the mid-point along the proximal/distal axis of fully expanded leaf 5, and traits were quantified for one set of lateral veins adjacent to the midvein. WT comparisons were segregating in the *Zmrvn1*-4 line. Statistical significance was calculated by t-tests (A), poisson regression models and subsequent likelihood ratio calculation by comparison to a null model (B), or Wilcoxon rank sum test (C). Statistical significance groups are indicated by letters above plot: the same letter indicates no significant difference. Biological replicates are indicated by n= above plots. Means are indicated by black diamonds. Outliers are indicated by empty circles. Breaks in y-axis scale are indicated by a double diagonal line.

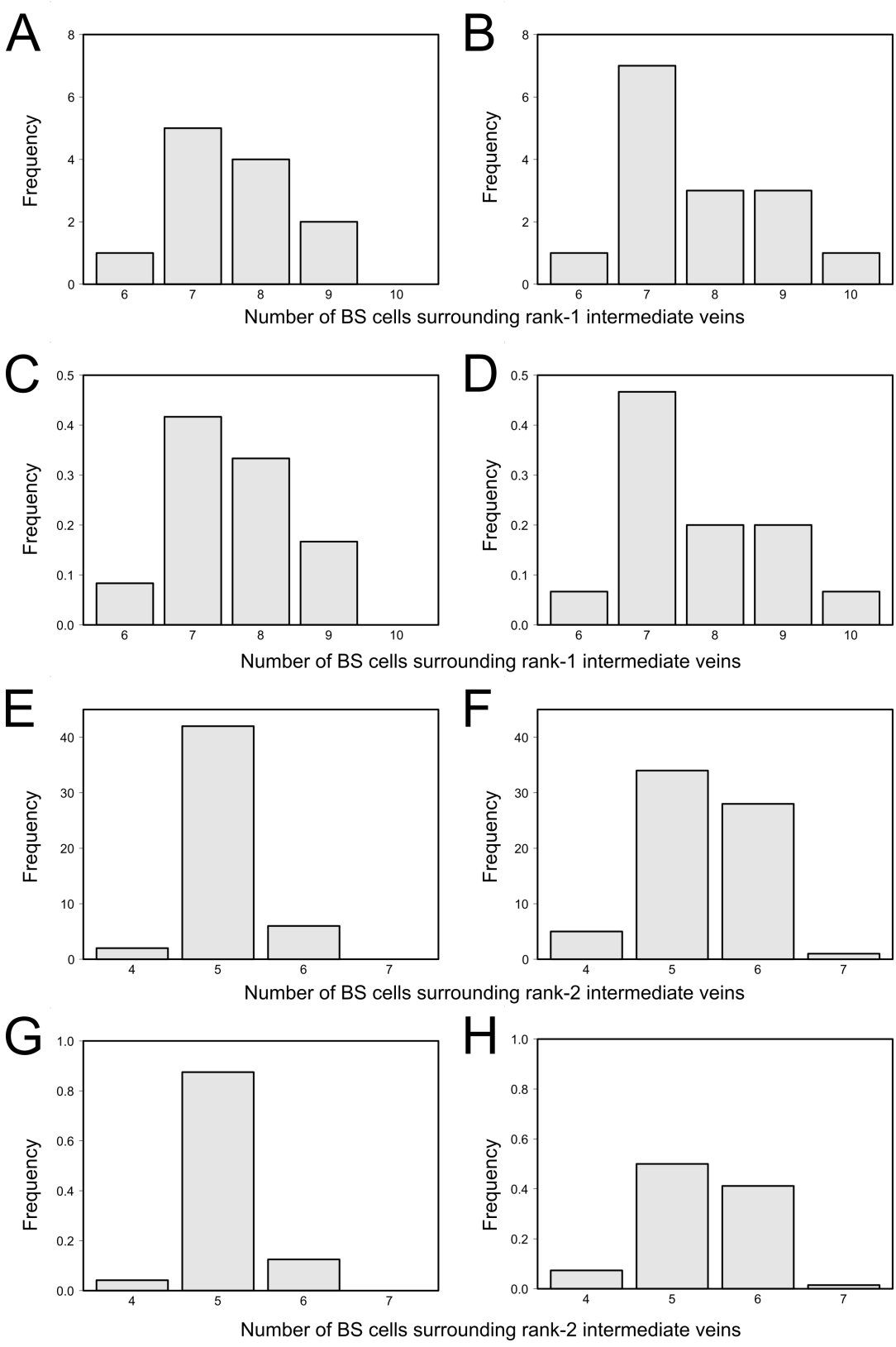


Figure 5.14 Barplots displaying BS cell number surrounding intermediate veins in *Zmrvn1-4* mutant and WT foliar leaves. Figure legend is on the next page.

Figure 5.14 Barplots displaying BS cell numbers surrounding intermediate veins in *Zmrvn1-4* mutant and WT foliar leaves. A-H) The number of BS cells surrounding rank-1 (A-D) and rank-2 (E-H) intermediate veins in both WT segregating in the *Zmrvn1-4* line (A,C,E,G), and *Zmrvn1-4* (B,D,F,H) foliar leaves. Counts were taken using fresh transverse sections from the mid-point of the proximal/distal axis of fully expanded leaf 5. BS cells were counted for all intermediate veins between one set of lateral veins in each biological replicate. For WT (A,C,E,G) n=4 and *Zmrvn1-4* (B,D,F,H) n=5. In (A,B,E,F) exact counts are displayed, with frequency indicating the total occurrences of veins with a certain number of surrounding BS cells. In (C,D,G,H) BS cell counts are normalised for sample size, with frequencies indicating the proportion of the total number of veins quantified that had a certain number of surrounding BS cells. Statistical significance ($p<0.05$) was determined using mixed effect poisson regression models and subsequent likelihood ratio calculation by comparison to a null model. No statistical differences between WT and *Zmrvn1-4* were found.

5.2.4.2. *Zmrvn1-4; Zmnkd1-R; Zmnkd2-R* triple mutants did not show more severe leaf patterning defects than single *Zmrvn1-4* or double *Zmnkd1-R; Zmnkd2-R* mutants

In *Arabidopsis* root development, many IDD genes function redundantly, as mutant phenotypes are only revealed once the function of multiple genes is disrupted simultaneously (Long et al. 2015b). As such, it was hypothesised that if *ZmNKD1* and *ZmRVN1* function in Kranz development, a stronger mutant phenotype might be found in the *Zmrvn1-4; Zmnkd1-R; Zmnkd2-R;* triple mutant. As this mutant was produced from a cross between lines from different inbred backgrounds, no detailed quantification was undertaken of the triple homozygous mutants in the initial F2 generation. However, no obvious alteration to the phenotypes already characterised in the *Zmnkd1-R; Zmnkd2-R* double mutant were observed. Kernels exhibiting the characteristic *Zmnkd* phenotype were still apparent in the segregating F2 seed, and the resultant plants displayed the slightly reduced growth previously attributed to mutations in *ZmNKD1* and *ZmNKD2*. Examination of leaf sections revealed no additional phenotypes relating to Kranz development (Figure 5.15). Extra BS cells were still found in the triple mutant, and whilst these may occur at lower frequency than in the *Zmnkd1-R; Zmnkd2-R* double mutant, this needs to be quantified properly in subsequent outcrossed generations where more individuals will be available for analysis. At this stage, there is no compelling evidence to suggest that *ZmRVN1* has a redundant role with *ZmNKD1/2* in Kranz development.

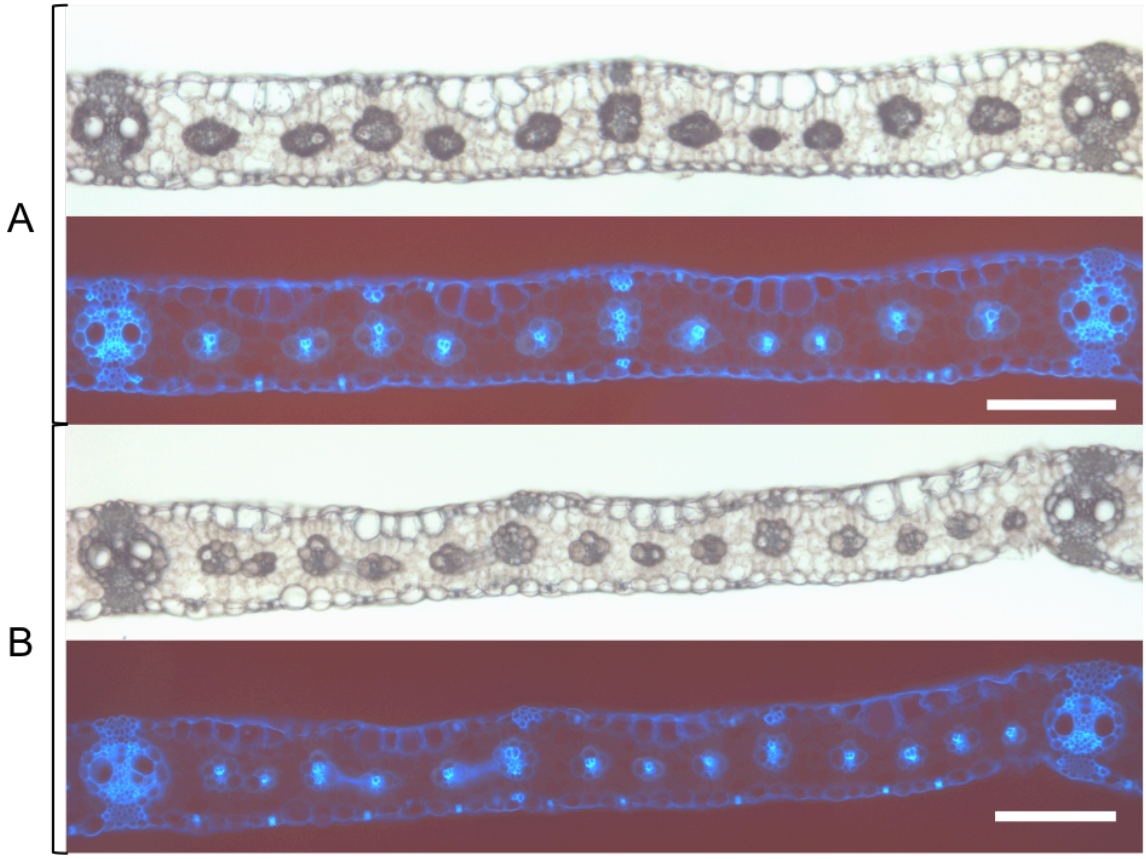


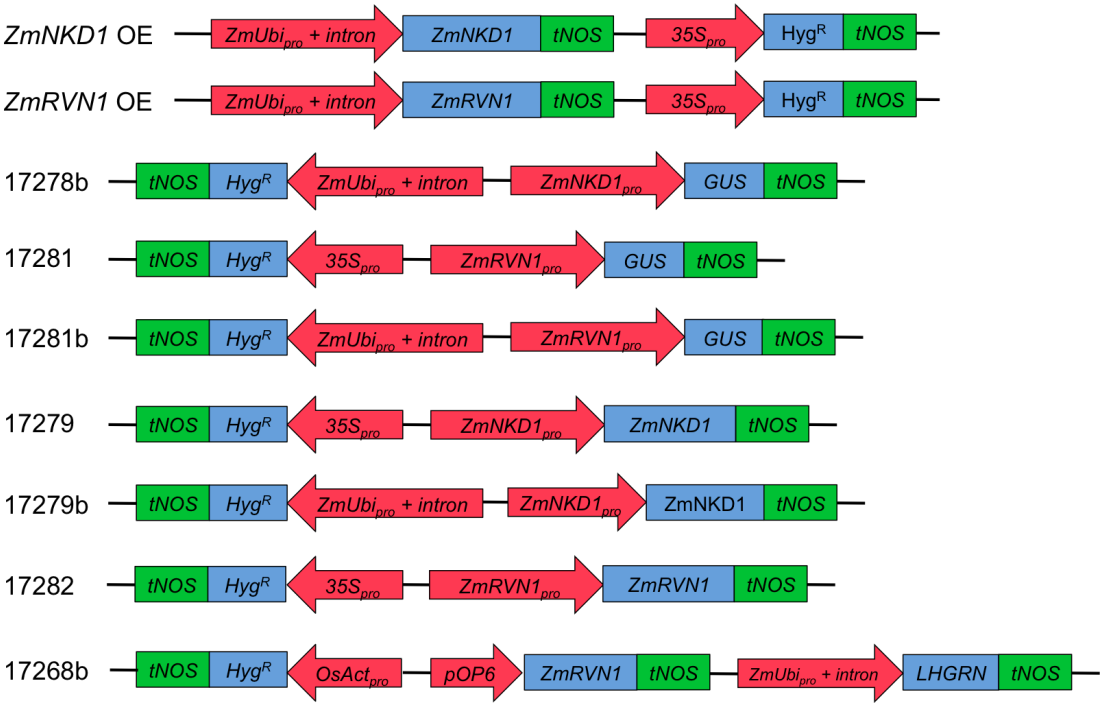
Figure 5.15 Transverse fresh sections of *Zmrvn1-4; Zmnkd1-R; Zmnkd2-R* triple mutants and WT foliar leaves. A) *ZmRVN1/Zmrvn1-4*; *ZmNKD1/±ZmNKD1; ZmNKD2/±ZmNKD2*. **B)** *Zmrvn1-4; Zmnkd1-R; Zmnkd2-R* triple mutant. In (A) and (B) sections were taken from the midpoint along the proximal/axis of fully expanded leaf 5, and imaged under brightfield and UV illumination. Scale bar = 200µm.

5.2.5. The effect of *ZmNKD1* and *ZmRVN1* expression in rice

5.2.5.1. Constitutive expression of *ZmNKD1* and *ZmRVN1* prevented the regeneration of rice callus

To determine whether expression of either ZmNKD1 or ZmRVN1 in rice affected leaf development, constructs were designed and cloned in which expression was driven by the maize ubiquitin promoter (Figure 5.16A). These constructs were transformed into rice callus multiple times, however, plants containing the transgene could not be recovered from tissue culture. In the case of *ZmUbi*$_{pro}$*::ZmRVN1*, a number of plants regenerated, but upon genotyping these were found to contain only the hygromycin resistance gene. The failure of callus transformed with constitutive expression constructs to regenerate was a recurring problem for a number of the high-priority Kranz candidate regulators previously identified, with constitutive expression of eight out of twelve proving problematic (Jane Langdale, personal communication) (Wang et al. 2013b; Fouracre et al. 2014). This is perhaps unsurprising, as many of these genes are predicted to function early in plant vascular development, and as such constitutive overexpression may simply result in the over-specification of vascular tissue in the regenerating tissue. However, since both ZmNKD1 and ZmRVN1 are hypothesised to function in M and BS cellular patterning rather than vascular specification, it is surprising that constitutive overexpression of these genes prevented rice callus regeneration. Given the high degree of protein sequence conservation between IDD genes, and their involvement in a wide-array of developmental processes, it is possible that overexpression interferes with early plant development to the extent that plant regeneration is impaired. Alternatively, it could be that the hypothesis concerning ZmNKD1 and ZmRVN1 involvement in BS/M cell patterning is incorrect, and in fact they regulate more fundamental developmental processes (or interfere with them in rice). Whatever the cause, the failure to recover plants with constitutive overexpression of ZmNKD1 or ZmRVN1

A



B

Construct ID	Transgenic line ID
17278b	2.1, 4.2
17281	1b, 5b, 6a
17282	2a, 4a, 5a
17268b	1b, 2a, 5b

Figure 5.16 Constructs used to investigate the role of *ZmNKD1* and *ZmRVN1* in rice. A) Cartoon construct diagrams. The construct ID is indicated on the left side. OE= overexpression. Red arrows indicate promoters (pro), with the arrowhead indicating the direction of expression. Blue boxes are coding sequences, and green boxes are terminators (in all cases *tNOS*). Black lines indicate linking DNA sequences between expression modules. **B)** Construct IDs and (where relevant for subsequent figure labelling) T0 transgenic line IDs. All distinct IDs represented independent transformation events (there is no difference between IDs that include only numbers and those that include a lower case letter).

meant that alternative approaches were sought to investigate the effect of expression on leaf development in rice.

5.2.5.2. Maize endogenous promoters were cloned for expression of *ZmNKD1* and *ZmRVN1* in rice

Given that callus transformed with constitutive *ZmNKD1* and *ZmRVN1* expression constructs failed to regenerate, it was decided to restrict the expression of these genes in rice to the spatial and temporal domain analogous to that of maize by using the endogenous maize promoters. To this end, the regions 3kb upstream of the *ZmNKD1* and *ZmRVN1* start codon (i.e. including 5'UTR sequences) were synthesised. As sequences were cloned using the Golden Gate system, it was necessary to alter certain type-IIS restriction sites (if present) in the sequence that would otherwise interfere with the cloning process. For the *ZmNKD1* promoter two such sequences were identified, and for *ZmRVN1* only one such sequence was identified. These sites were altered to remove the type-IIS restriction site, using either the maize duplicate or sorghum orthologs as a template, thus minimising the chance of changing a motif that is critical for C_4-like expression. To validate promoter activity in rice, constructs were generated in which the maize promoters were used to drive expression of the β-glucuronidase (GUS) reporter gene (Figure 5.16).

5.2.5.3. The *ZmNKD1* promoter is a strong shoot specific promoter in rice

To assess whether the *ZmNKD1* promoter is capable of driving transgene expression in rice, a $ZmNKD1_{pro}$*::GUS* construct was transformed into rice. The maize ubiquitin promoter was used to drive hygromycin expression in this construct, in order to avoid potential activity from the *35S* promoter (either as a promoter or enhancer), which could interfere with the activity of the *ZmNKD1* promoter (Yoo et al. 2005; Zheng et al. 2007) (Figure 5.16). Five T1 seed for each of three independent transgenic lines were de-husked, sterilised and then germinated on ½ MS media. After 7 days, seedlings were tested for GUS activity. One of the three independent events

displayed near ubiquitous GUS activity, with substrate accumulation throughout both the roots and shoots (data not shown). This was markedly different to the pattern found in the other two independent lines, and as such is not believed to be representative of the activity of the *ZmNKD1* promoter. The remaining lines exhibited no activity in the roots, but strong activity in the shoot, with substrate detectable in all visible tissue (Figure 5.17A,C). To establish whether the *ZmNKD1* promoter conferred gene expression during early leaf development, sections were taken of developing leaf primordia. This revealed strong GUS activity during early leaf development (Figure 5.17B,D). At the youngest stages, substrate was detected in vascular tissue, BS cells and particularly strongly in M cells. At slightly later stages, substrate localised preferentially to M cells, although this was more pronounced in one of the lines where virtually no substrate was detected in either veins or BS cells of outer leaf primordia (Figure 5.17D). This pattern matches the expression pattern observed by *in situ* hybridisation in developing maize leaf primordia, where *ZmNKD1* transcripts are detected preferentially in M cells (Olga Sedelnikova, personal communication). However, differences between lines, which may be the result of transgene copy number, insertion position or construct fragmentation and/or rearrangement during insertion, means that more independent transformation events need to be generated and analysed to assess the *ZmNKD1* promoter expression profile in rice.

5.2.5.4. Rice transformed with a *ZmNKD1$_{pro}$::ZmNKD1* construct did not set T1 seed

To determine whether *ZmNKD1$_{pro}$* driven expression of *ZmNKD1* affected rice leaf development, two *ZmNKD1$_{pro}$::ZmNKD1* constructs were cloned and transformed into rice (Figure 5.16). In one of these constructs (17279) the *35S* promoter was used to drive hygromycin resistance, whereas in the other (17279b) *ZmUbi$_{pro}$* was used. In both cases transformations were successful, with 14 and 4 independent

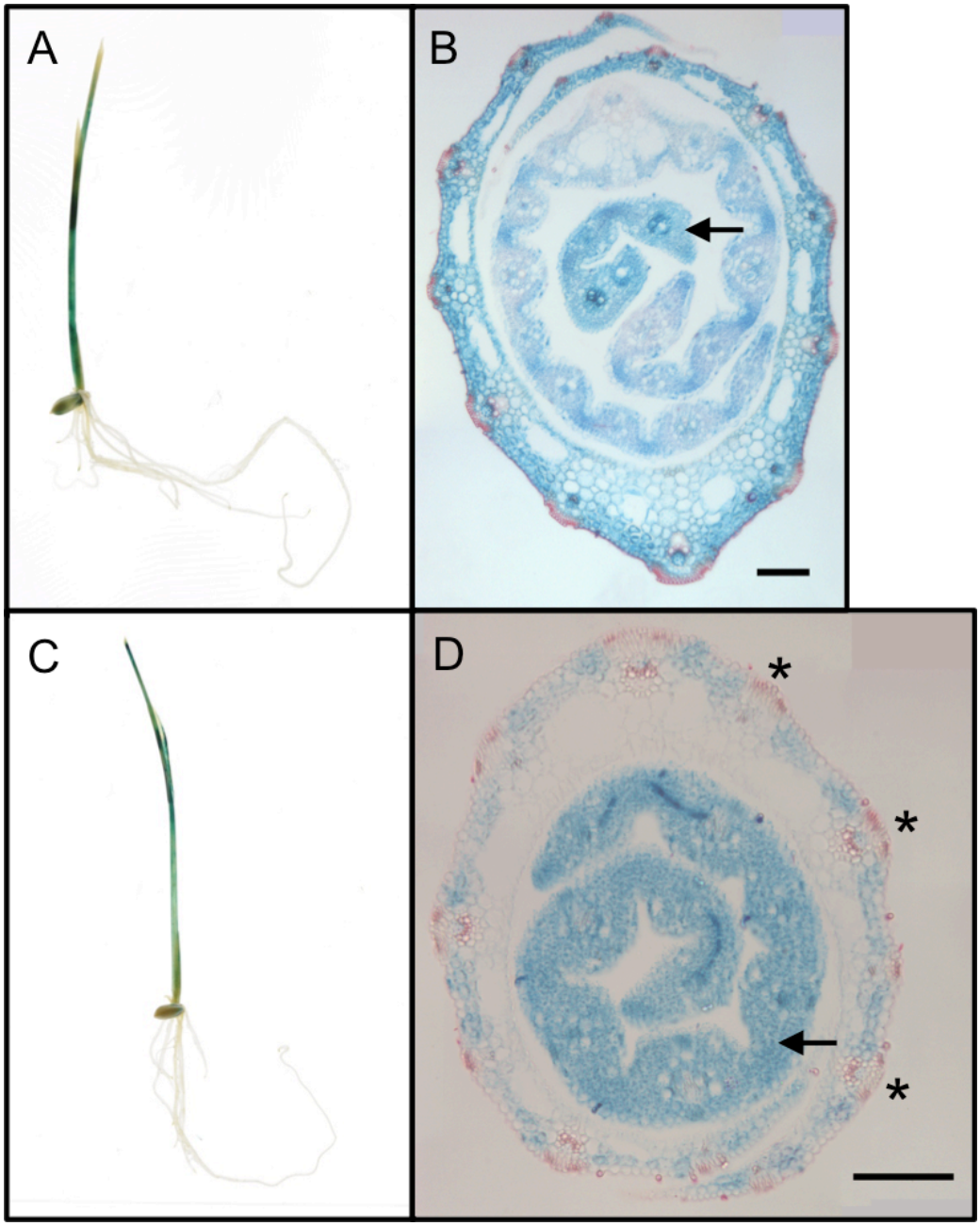


Figure 5.17 GUS activity driven by the *ZmNKD1* promoter in rice. A-D) GUS staining of rice seedlings transformed with construct 17278b (*ZmNKD1*$_{pro}$*::GUS*). Transgenic line 2.1 (A,B) and transgenic line 4.2 (C,D). GUS staining of 7 day old seedlings (A,C) and GUS staining of sectioned shoot apices showing leaf primordia at a variety of ages (B,D). In (B) and (D) red staining is residual eosin from wax infiltration. Arrows in (B) and (D) indicate young leaf primordia where GUS substrate is visible in all cell-types. Asterisks in (D) indicate examples of veins and BS cells where GUS substrate is only visible in M cells. Scale bars are 100µm.

transgenic lines obtained for each construct. These plants appeared healthy, and grew well in the greenhouse. However, lines from both constructs failed to set viable T1 seed, with all seed remaining empty flowering. T0 transgenic kitaake seed set was variable across all transformed lines, especially during the winter when day-length is shorter and less natural light available. However, the fact that this effect was consistent across two independent constructs and multiple independent transformation events likely indicates an effect (at least in part) of the *ZmNKD1* transgene when expressed using the maize *ZmNKD1* promoter. As these lines failed to set seed no detailed phenotypic analysis could be undertaken on T1 plants. Therefore, to assess whether *ZmNKD1* affected leaf development in these lines, transverse sections were taken from the flag leaves of two independent T0 individuals containing the *ZmUbi$_{pro}$::Hyg^{R}-ZmNKD1$_{pro}$::ZmNKD1* construct (Figure 5.18). In the absence of an appropriately staged WT control, these sections were compared to T0 plants containing the *ZmUbi$_{pro}$::Hyg^{R}-ZmNKD1$_{pro}$::GUS* construct (Figure 5.18). No obvious perturbations in leaf anatomy were apparent in these sections, thus it appears that expression of *ZmNKD1* in rice does not alter leaf development.

5.2.5.5. The *ZmRVN1* promoter confers early expression in rice leaves only when the *35S* promoter is used upstream to drive the hygromycin resistance gene

Two distinct constructs were used to test whether the *ZmRVN1* promoter was able to drive transgene expression in rice (Figure 5.16). These constructs differed only in the promoter used to drive the hygromycin selectable marker, which in one case was the *35S* promoter (17281), and in the other was the maize ubiquitin promoter (17281b). All other aspects of the constructs were identical, with the *ZmRVN1* promoter used to drive GUS expression (Figure 5.16). For the *35S$_{pro}$::Hyg* construct, three independent

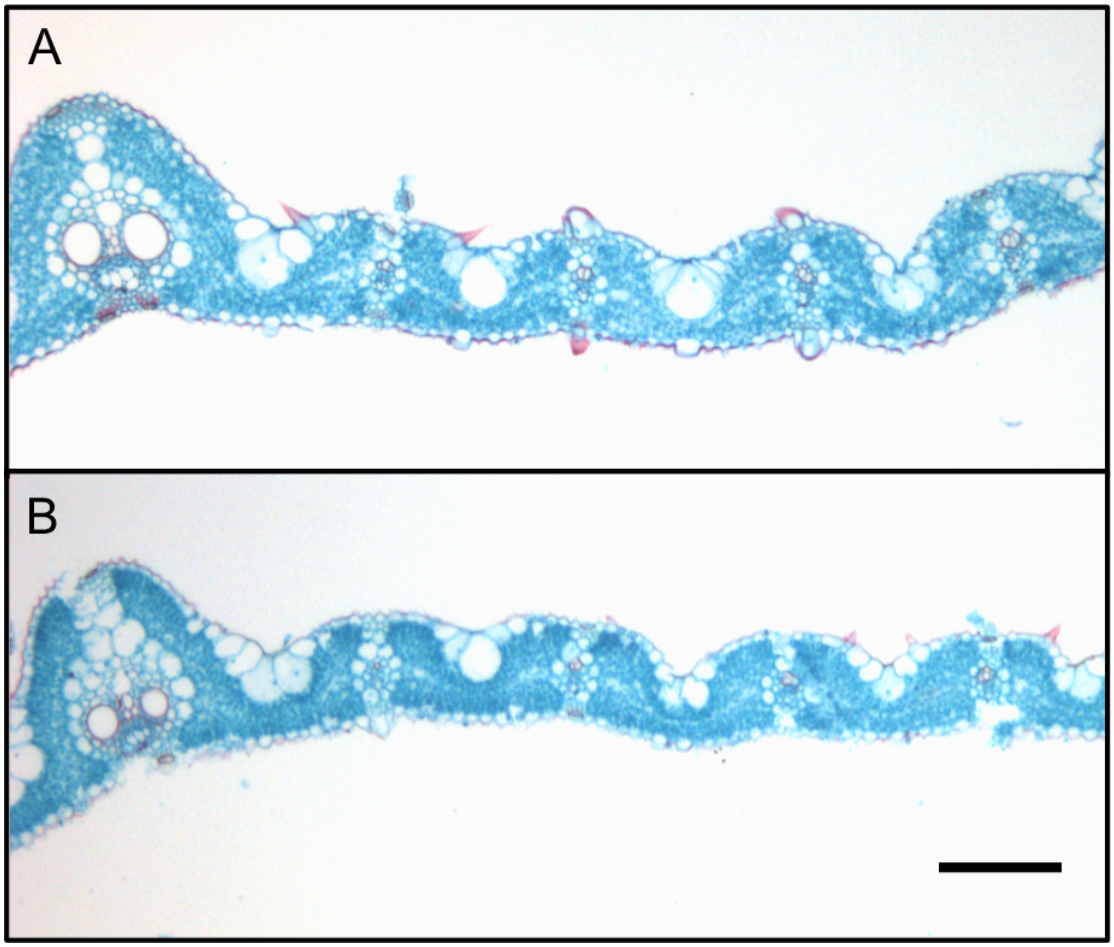


Figure 5.18 Transverse sections of *ZmNKD1$_{pro}$::GUS* and *ZmNKD1$_{pro}$::ZmNKD1* T0 rice leaves. A) *ZmNKD1$_{pro}$::GUS* (17278b) **B)** *ZmNKD1$_{pro}$::ZmNKD1* (17279b). Sections were taken from the midpoint of the proximal/distal axis of the fully expanded flag leaf. Fixed leaf tissue was sectioned (10µm) and stained with Safranin-O and Fast-Green. Scale bar = 100µm.

transgenic lines were tested for GUS activity, with all three events exhibiting consistent patterns. In seedlings, GUS substrate accumulation patterns were similar to that of *ZmNKD1*, with little activity detected in the roots but consistent (albeit weaker) activity in the shoot (Figure 5.19A,C,E). However, when sections of developing leaf primordia were examined, a pattern very different to that of the *ZmNKD1* promoter was revealed (Figure 5.19B,C,E,F,H). Here, GUS activity was detected in all veins, at all stages of development. In the youngest veins this activity was only detected abaxially, but in older veins both ab- and adaxial regions were stained. In addition, in more developed veins some activity was detected in the surrounding BS cells, and in the most developed leaves examined, M cells were also stained. Interestingly, however, when this analysis was repeated using two independent lines of the *ZmUbi$_{pro}$::Hyg* version of the construct, no shoot activity was detected, nor any expression in developing leaf primordia (data not shown). Given that activity was detected in the seed, indicating that the construct is functional, it appears that the observed leaf-staining pattern in 17281 lines is due to interference from the *35S* promoter. Interaction between the *35S* promoter and other promoters in a construct, even when the *35S* promoter has been used in the opposite orientation, has previously been reported (Yoo et al. 2005; Zheng et al. 2007). However, the observed expression profile observed in this case appears distinct from the constitutive expression conferred by *35S$_{pro}$* when it is used to directly drive gene expression (Battraw and Hall 1990). This interaction is most likely caused by the enhancer sequence in *35S$_{pro}$*, and it is not clear whether the observed *ZmRVN1* promoter expression profile in this context bears any relationship to the endogenous maize behaviour of the construct.

5.2.5.6. *ZmRVN1$_{pro}$::ZmRVN1* did not perturb rice leaf anatomy

It is not clear how the GUS activity profile observed in young rice leaves expressing the *35S$_{pro}$::Hyg - ZmRVN1$_{pro}$::GUS* construct relates to the endogenous maize

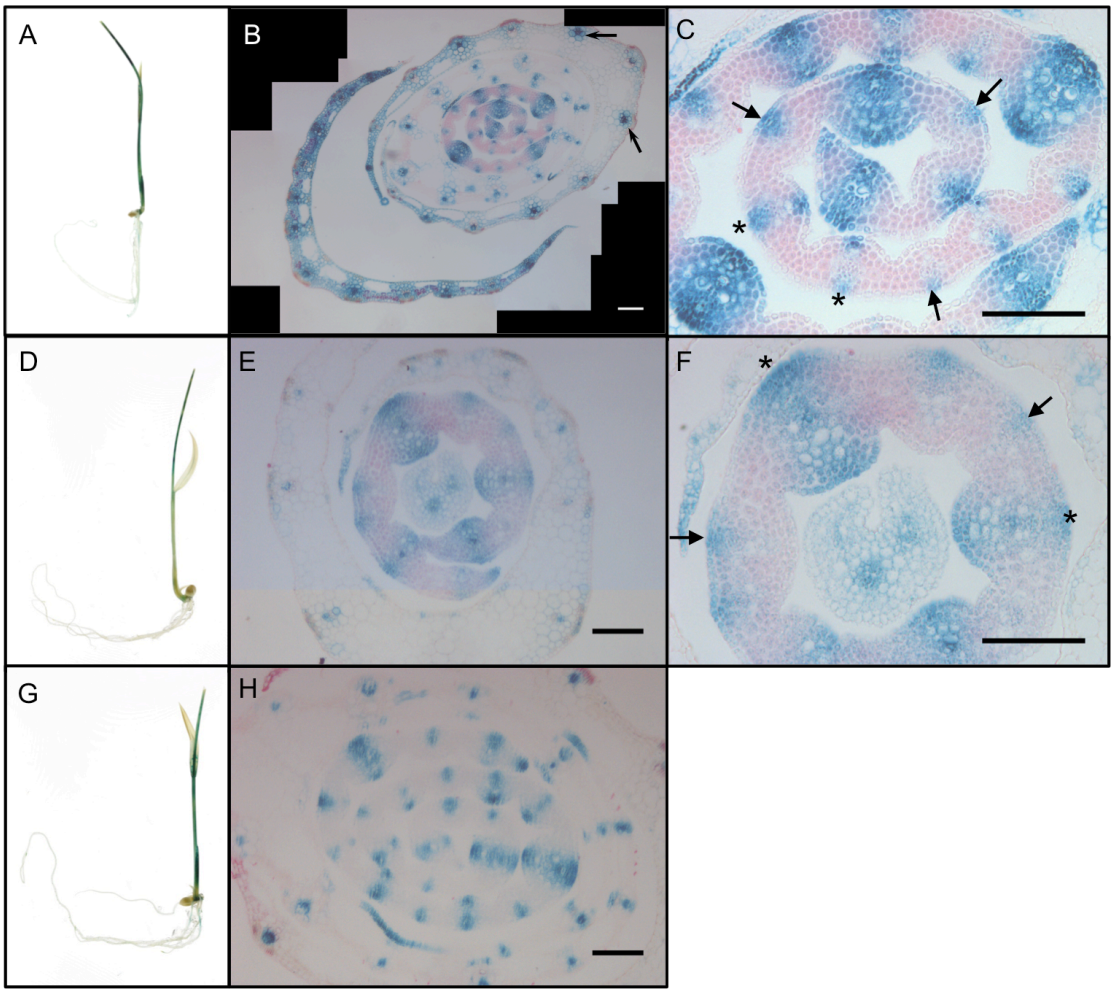


Figure 5.19 GUS activity driven by the *ZmRVN1* promoter in rice. A-H) GUS staining of rice seedlings transformed with construct 17281 (*ZmRVN1*$_{pro}$*::GUS* with *35S*$_{pro}$*::Hyg*^{R}). Transgenic line 1b (A-C), transgenic line 5b (D-F) and transgenic line 6a (G,H). GUS staining of 7 day old seedlings (A,D,H) and GUS staining of sectioned shoot apices showing leaf primordia at a variety of ages (B,C,E,F,H**).** In (B,C,E,F,H**)** red staining is residual eosin from wax infiltration. Arrows in (B) indicate veins with preferential BS GUS substrate accumulation. Arrows in (C) and (F) indicate developing veins where GUS substrate is only visible in the abaxial region. Asterisks in (C) and (F) indicate older veins where GUS substrate is visible both ab- and adaxially to the vein. All scale bars in (B,C,E,F,H**)** are 100µm.

expression profile of *ZmRVN1*. Despite this uncertainty, when used in this context, the promoter allows the effect of expressing transgenes in the vasculature and BS of young rice leaves to be investigated. To determine whether the expression of *ZmRVN1* in the veins and BS of young rice leaves alters leaf anatomy, a *$35S_{pro}$::Hyg-$ZmRVN1_{pro}$::ZmRVN1* construct (17282) was cloned and transformed into rice callus (Figure 5.16). Unlike constitutive expression of *ZmRVN1* using the maize ubiquitin promoter, which prevented regeneration, plants transformed with this construct regenerated, with the resultant seedlings phenotypically indistinguishable from WT. T1 seed was collected and used for phenotypic analysis. Around 25 seed from three independent transgenic lines were planted in soil. These lines (2a, 4a, 5a) were genotyped using primers that amplified a region between the *ZmRVN1* promoter and CDS, in order to identify transgene positive and segregating WT plants (Appendix 10). In the case of line 5a, only one segregating WT plant was obtained from 25 total plants, likely indicating that this line contained multiple copies of the transgene. Kitaake WT was used as a phenotypic comparison for this line. An attempt to check expression by RT-PCR was made using RNA isolated from mature leaf 5, however, no transgene expression was detected. Given that the *35S* promoter is known to confer strong expression in mature rice leaves (Battraw and Hall 1990), the absence of transcripts suggests that in constructs 17281 and 17282, the *35S* promoter acts as an enhancer of the *ZmRVN1* promoter rather than driving expression itself. As RNA isolation from leaf primordia is a destructive process, and T2 seed were needed, no further expression analysis was undertaken on the T1 plants.

Transgene positive plants displayed no phenotypic differences to WT in terms of gross morphology, with growth and flowering unaffected by putative expression of *ZmRVN1*. To assess whether expression of *ZmRVN1* in the developing veins of leaf primordia perturbed leaf anatomy, fresh transverse sections of mature leaf 6 were examined (Figure 5.20). No significant difference in vein density was found between

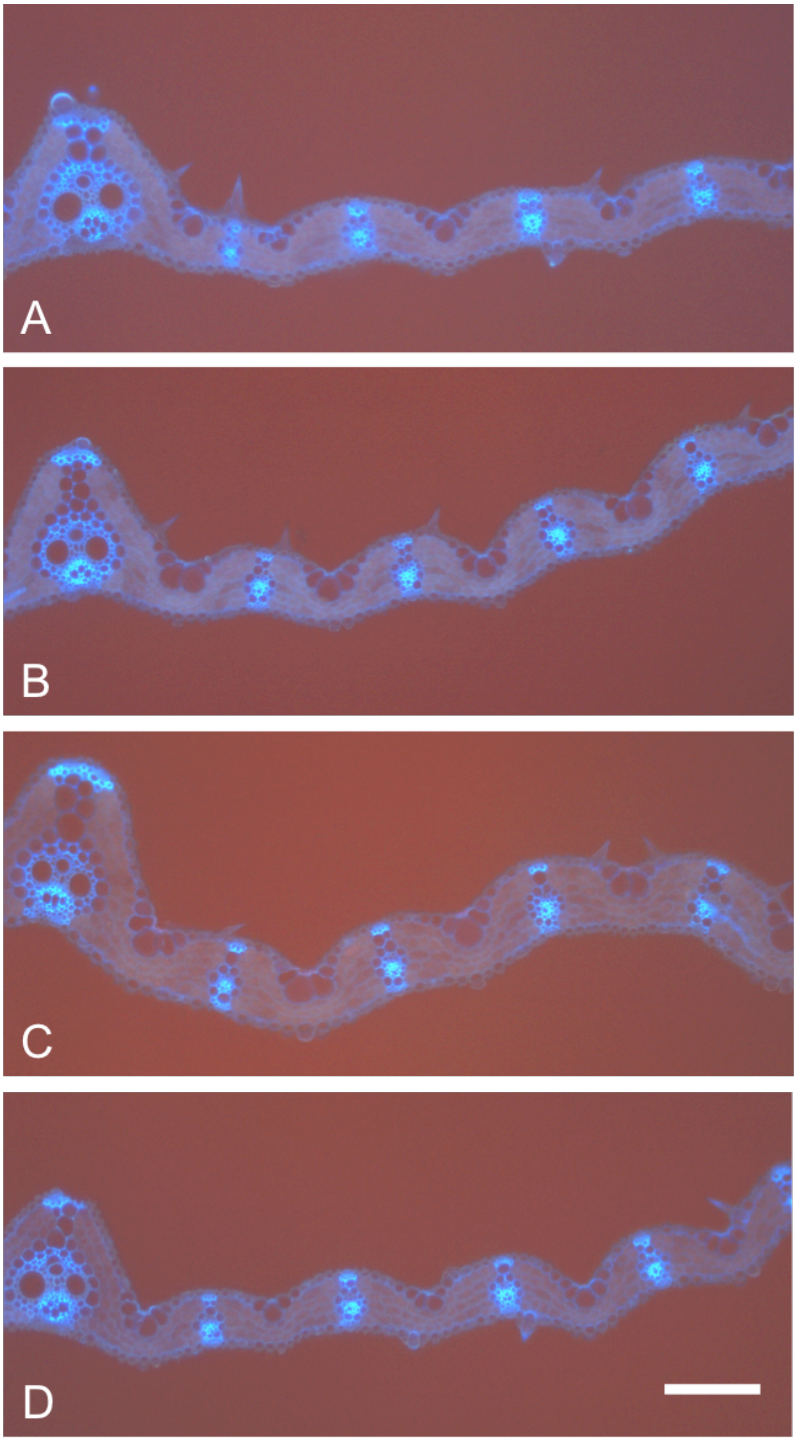


Figure 5.20 Transverse fresh sections of *ZmRVN1*$_{pro}$*::ZmRVN1* and WT rice leaves. A) Kitaake WT **B-D)** Independent T1 rice lines transformed with construct 17282 (*ZmRVN1*$_{pro}$*::ZmRVN1* with *35S*$_{pro}$*::Hyg*^{R}). Transgenic line 2a (B), 4a (C) and 5a (D). Sections were taken from the midpoint along the proximal/distal axis of fully expanded leaf 6 and imaged using UV illumination. Scale bar = 100µm.

transgenic and WT plants, indicating that expression of *ZmRVN1* in this context did not trigger the specification of more veins (Figure 5.21). Furthermore, no significant difference was found in terms of the number of M cells separating veins, or the number of BS cells surround intermediate veins in the mature leaf (Figure 5.22, Figure 5.23). Therefore, presumed expression of *ZmRVN1* in this context does not alter leaf anatomy.

5.2.5.7. Inducible expression of *ZmRVN1* resulted in no alterations to leaf patterning in young rice leaves

As constitutive overexpression of *ZmRVN1* prevented the regeneration of rice callus, an inducible expression system was used to investigate the effect of *ZmRVN1* overexpression on leaf development. The *pOP6/LHGRN* system was chosen for this purpose, as it enables transgene expression to be induced specifically and quickly by the addition of dexamethasone (Dex) (Craft et al. 2005; Samalova et al. 2005), and has recently been validated to work well in rice (Daniela Vlad, personal communication). A construct (17268b) was designed and cloned to place the LHGRN activator under the control of the maize ubiquitin promoter, with the inducible *pOP6* promoter used to drive *ZmRVN1* expression in response to Dex (Figure 5.16). The rice actin promoter was used to drive hygromycin expression in order to avoid the potential cross-activity from the *35S* promoter. This construct was transformed into rice and regenerated T0 plants were grown to maturity in the absence of induction, with T1 seed then collected for experimentation.

Three independent transgenic lines were chosen for phenotypic analysis, with between 15 and 20 seed for each event de-husked and sterilised, before being placed on a 28x28cm petri-dish containing ½ MS media and 1μM Dex. After 10 days the 1st leaf was harvested for PCR genotyping. In three lines all of the seedlings (18, 13 and 14 plants respectively) contained the transgene, with no segregating WT plants obtained (data not shown). This indicates that all of these lines were likely

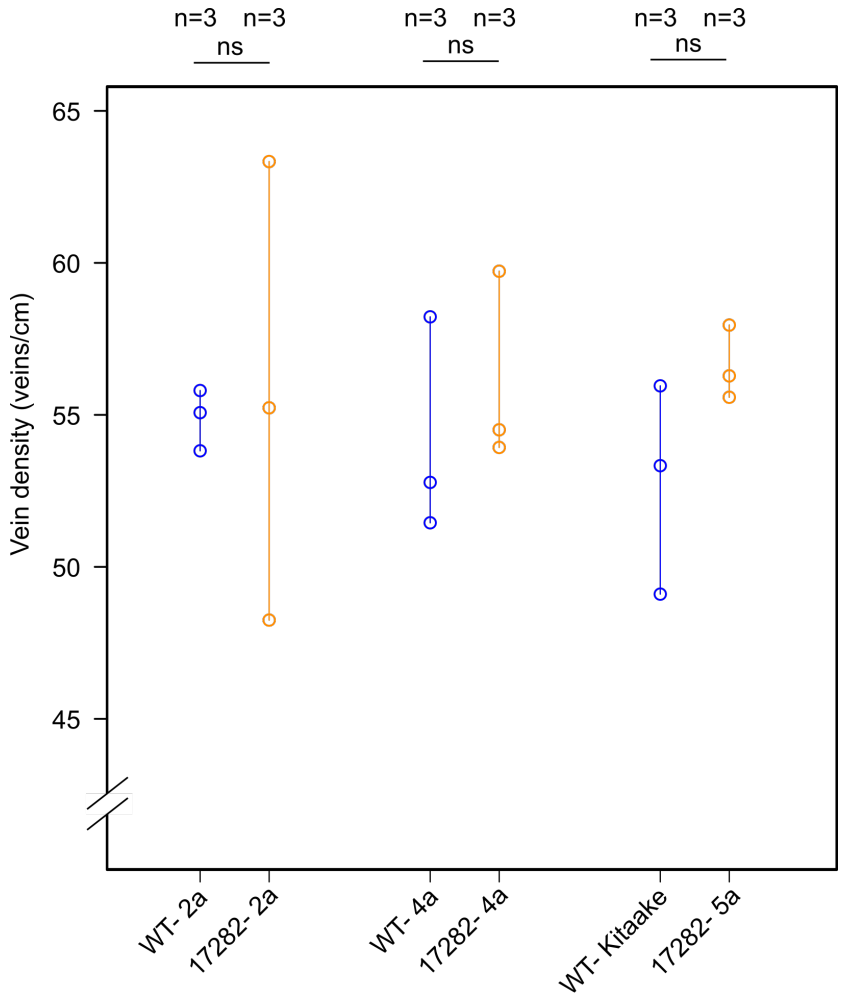


Figure 5.21 Lineplots showing vein density in leaves of *ZmRVN1$_{pro}$::ZmRVN1* rice lines. Vein density (veins/cm) of T1 rice lines transformed with construct 17282 (*ZmRVN1$_{pro}$::ZmRVN1* with *35S$_{pro}$::Hyg*). Vein density was calculated across the full width of fully expanded leaf 6, using sections taken from the mid-point along the proximal/distal axis. Three independent transgenic lines were quantified (17282- 2a, 4a and 5a), with comparisons to segregating WT plants (2a and 4a) or Kitaake WT (5a) made. Statistical significance, as calculated by t-tests, and samples sizes (n=) are indicated above the plot. ns = not significant ($p > 0.05$). Breaks in y-axis scale are indicated by a double diagonal line.

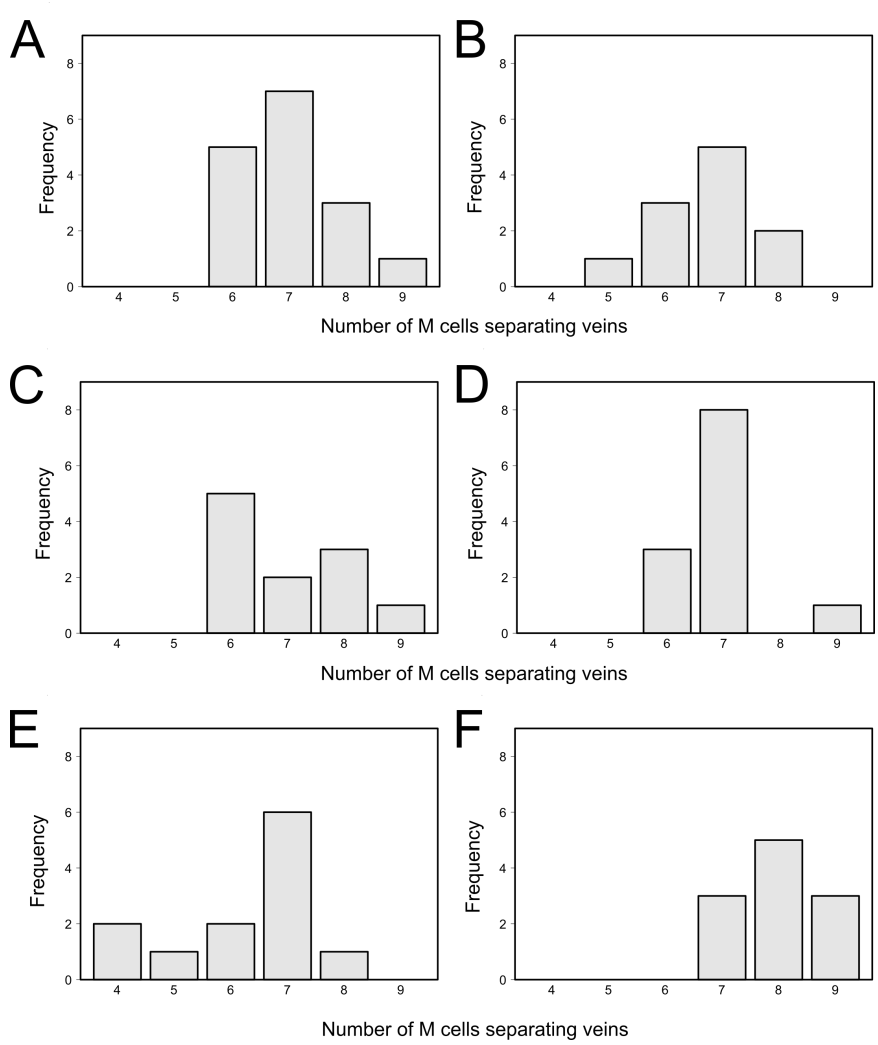


Figure 5.22 Barplots showing the number of M cells separating veins in leaves of *ZmRVN1*$_{pro}$*::ZmRVN1* rice lines. A) Segregating WT from construct 17282 (*ZmRVN1*$_{pro}$*::ZmRVN1* with *35S*$_{pro}$*::Hyg*) line 2a **B)** 17282 line 2a **C)** Segregating WT from construct 17282 line 4a **D)** 17282 line 4a **E)** Kitaake WT **F**) 17282 line 5a. M cell counts were obtained using fresh transverse sections taken from the mid-point along the proximal/distal axis of fully expanded leaf 6. Three plants (n=3) for each line was analysed, with three M counts per individual plant. Statistical significance was calculated using mixed effect poisson regression models and subsequent calculation of a likelihood ratio by comparison to a null model. No statistically significant differences between WT and 17282 lines were found ($p > 0.05$).

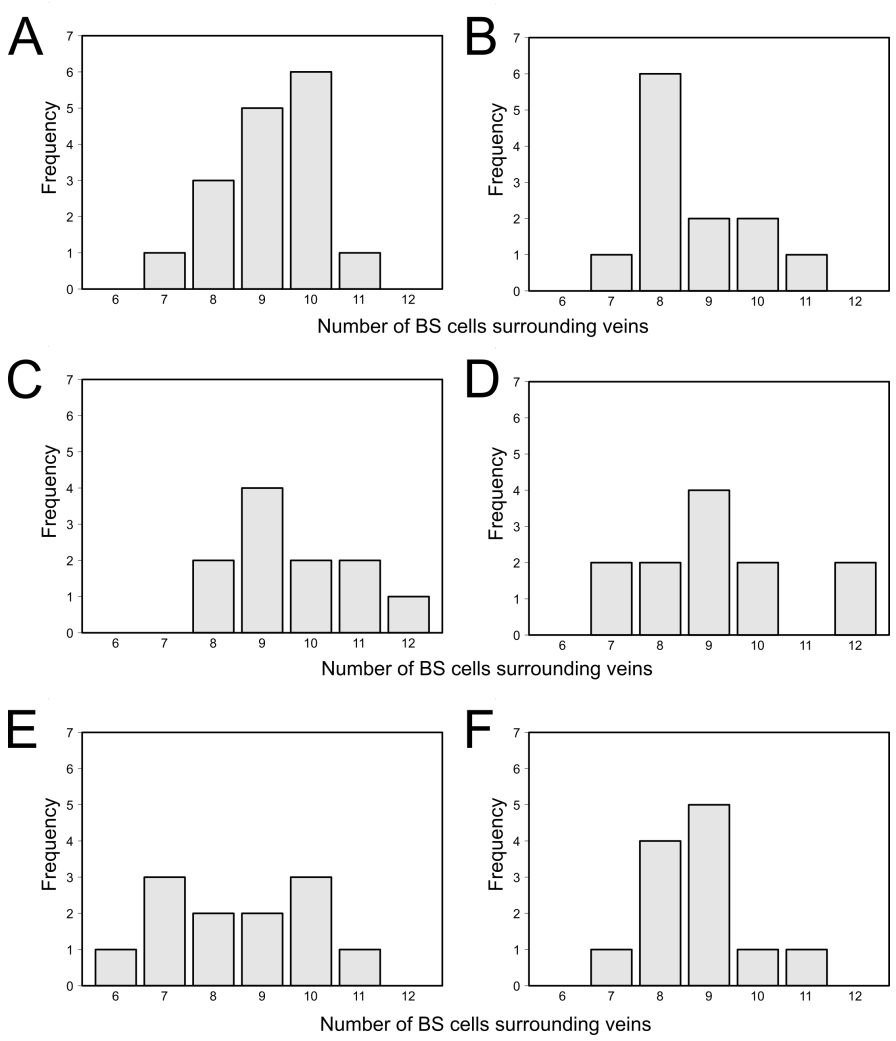


Figure 5.23 Barplots showing the number of BS cells surrounding veins in *ZmRVN1$_{pro}$::ZmRVN1* rice lines. A) Segregating WT from construct 17282 (*ZmRVN1$_{pro}$::ZmRVN1* with *35S$_{pro}$::Hyg*) line 2a **B)** 17282 line 2a **C)** Segregating WT from construct 17282 line 4a **D)** 17282 line 4a **E)** Kitaake WT **F**) 17282 line 5a. BS cell counts were obtained using fresh transverse sections taken from the mid-point along the proximal/distal axis of fully expanded leaf 6. Three plants (n=3) for each line was analysed, with four M counts per individual plant. Statistical significance was calculated by mixed effect poisson regression models and subsequent calculation of a likelihood ratio by comparison to a null model. No statistically significant differences between WT and 17282 lines were found ($p > 0.05$).

multiple insertion events. In the absence of segregating WT plants, transgenic plants were compared to Kitaake WT plants treated in the same way. Dex-induced transgenic plants were morphologically indistinguishable from WT (Figure 5.24). After 13 days leaf 2 was harvested from four transgene positive plants for each line, with the region halfway along the length of the leaf used for phenotypic analysis. The remainder of the leaf was harvested for RNA extraction and cDNA synthesis. RT-PCR confirmed that in the WT plants no *ZmRVN1* expression was observed in the presence of Dex, whereas in the three independent transgenic lines *ZmRVN1* expression was detected for all plants (Figure 5.25). To assess whether Dex-induced expression of *ZmRVN1* perturbed rice leaf anatomy, fresh transverse sections of leaf 2 were examined (Figure 5.26). Dex-induced *ZmRVN1* overexpressing plants showed no difference in vein density when compared to Kitaake WT (Figure 5.27), which is in line with the initial hypothesis that *ZmRVN1* acts to pattern the BS rather than in vein patterning. However, no alterations in M or BS cell patterning were observed in the induced *ZmRVN1* overexpressing lines; the number of M cells separating veins (Figure 5.28A), and the number of BS surrounding intermediate veins (Figure 5.28B), were not significantly different from Kitaake WT.

5.3. Discussion

5.3.1. There is extensive redundancy between IDD genes

Whilst some evidence was obtained that *ZmNKD1* and *ZmNKD2* play a role in leaf patterning (see section 5.3.2), generally neither loss-of-function analysis in maize nor gain-of-function analysis in rice revealed compelling evidence of a key role for *ZmNKD1* or *ZmRVN1* in regulating the development of Kranz anatomy. In the case of *ZmRVN1*, the lack of an obvious phenotypic perturbation in *Zmrvn1-4* mutants might be due to the presence of the maize-specific homeolog *ZmRVN1h.* A revised gene model was constructed for *ZmRVN1h*, which is expressed in early leaf

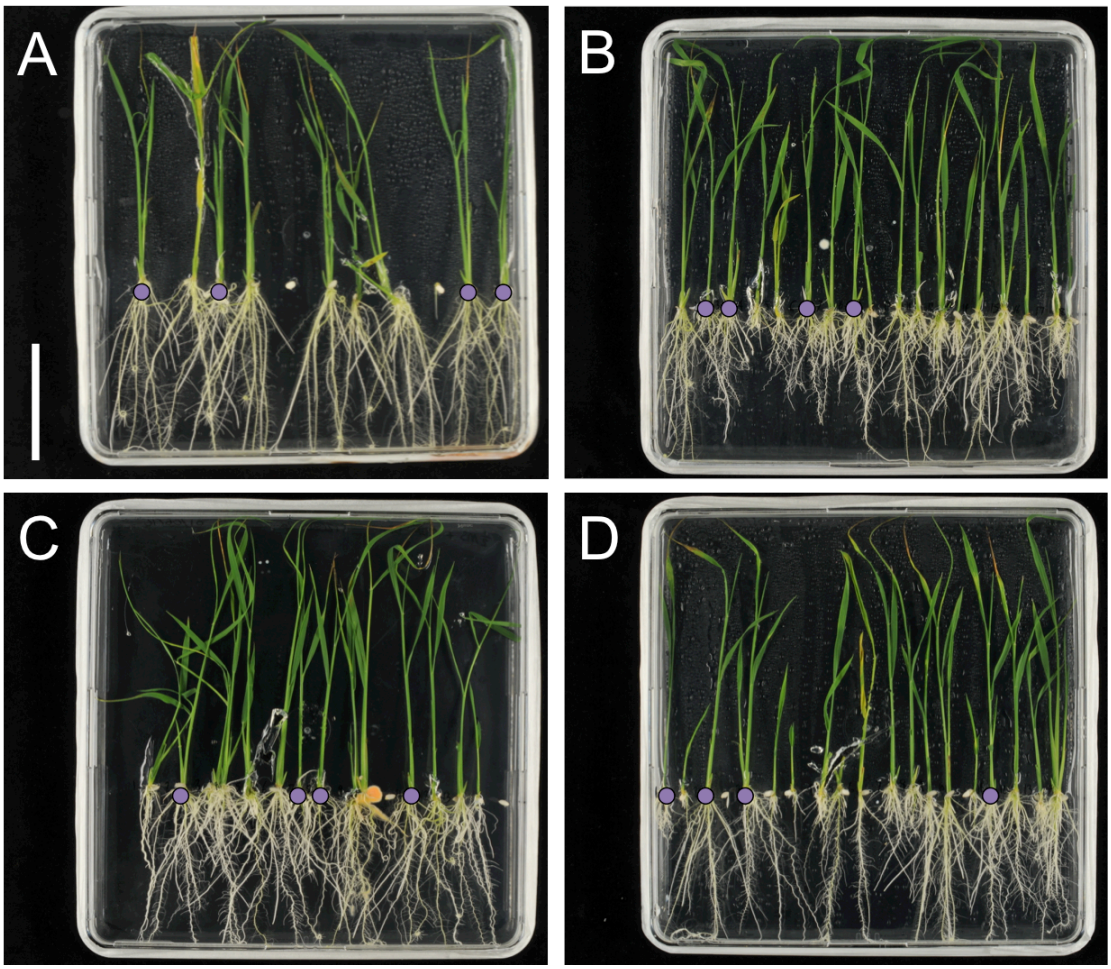


Figure 5.24 Images of Kitaake WT and transgenic rice plants following Dex-induction of *ZmRVN1* expression. A) Kitaake WT **B-D)** Three independent transgenic lines for construct 17268b (*pOP6::ZmRVN1 – ZmUbi$_{pro}$::LHGRN*): 1b (B), 2a (C), 5b (D). Kitaake T1 seed was germinated and grown on 28x28cm petri dishes contained ½ MS media and 1µM Dex. Photos were taken 13 days after planting. Purple dots indicate plants for which *ZmRVN1* expression was tested and subsequent phenotypic analysis undertaken. Scale bar = 7cm.

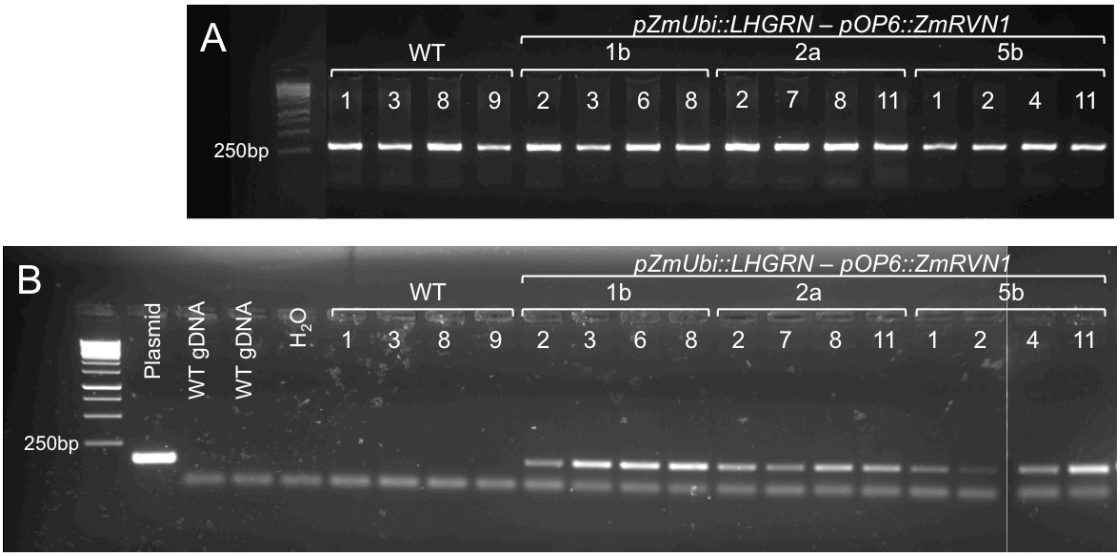


Figure 5.25 Gel images showing RT-PCR amplification products in WT and transgenic inducible *ZmRVN1* lines. A) Primers amplifying from the rice ubiquitin gene. **B)** Primers amplifying from the *ZmRVN1* gene. cDNA was amplified from RNA isolated from leaf 2. Plants were germinated on 1µM Dex and sampled 13 days after planting. In (A) and (B) four Kitaake WT samples are displayed alongside four different plants from three independent transgenic lines of construct 17268b (*pOP6::ZmRVN1 – ZmUbi$_{pro}$::LHGRN*) (1b, 2a and 5b). Gel images are positioned such that the same samples are vertically aligned in (A) and (B). In (B) plasmid, WT gDNA and H_2O controls are displayed.

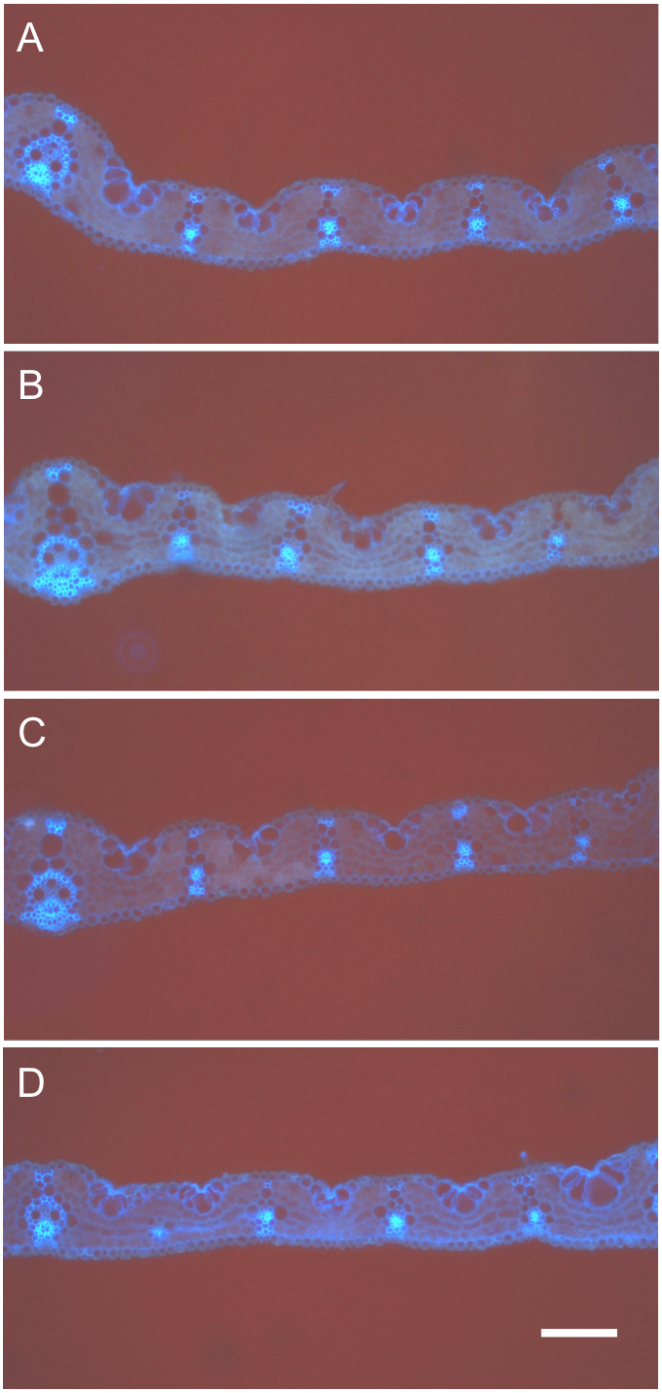


Figure 5.26 Transverse fresh sections of Kitaake WT and transgenic rice leaves following Dex induction of *ZmRVN1* expression. A) Kitaake WT **B-D)** Three independent Dex induced 17268b (*ZmUbi$_{pro}$::LHGRN – pOP6::ZmRVN1*) rice lines: 1b (B), 2a (C), 5b (D). Sections were taken from the midpoint along the proximal/distal axis of fully expanded leaf 2 and imaged using UV illumination. Plants were germinated on 1μM Dex and sampled 13 days after planting. Scale bar = 100μm.

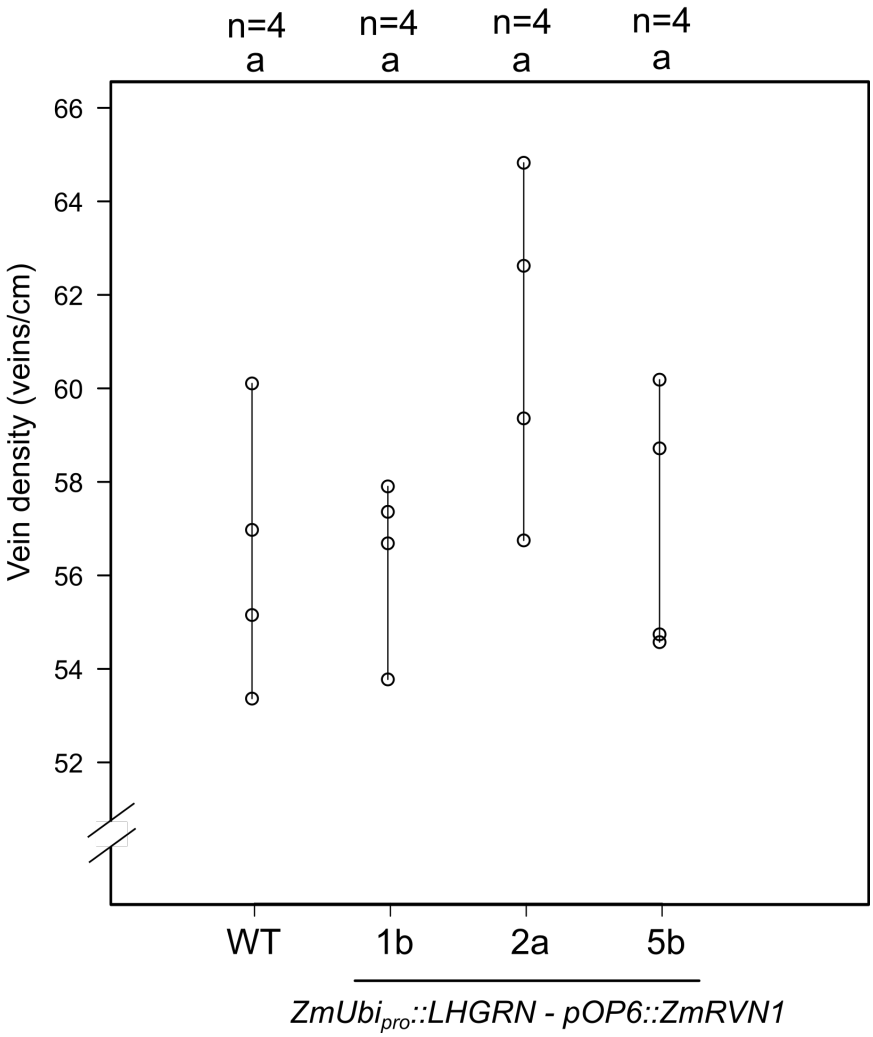


Figure 5.27 Lineplot showing vein density in Kitaake WT and transgenic line rice leaves following Dex induction of *ZmRVN1* expression. Vein density (veins/cm) calculated across the full width of fully expanded leaf 2, from sections taken from the mid-point along the proximal/distal axis. Three independent transgenic lines of construct 17268b ((*pOP6::ZmRVN1 – ZmUbi$_{pro}$::LHGRN*) were quantified (1b, 2a and 5b), with comparisons to Kitaake WT made. Plants were germinated on 1µM Dex and sampled 13 days after planting. Statistical significance, as calculated by a one-way ANOVA, and samples sizes (n=) are indicated above the plot. Statistically similar groups are indicated by the same letter. Breaks in y-axis scale are indicated by a double diagonal line.

-

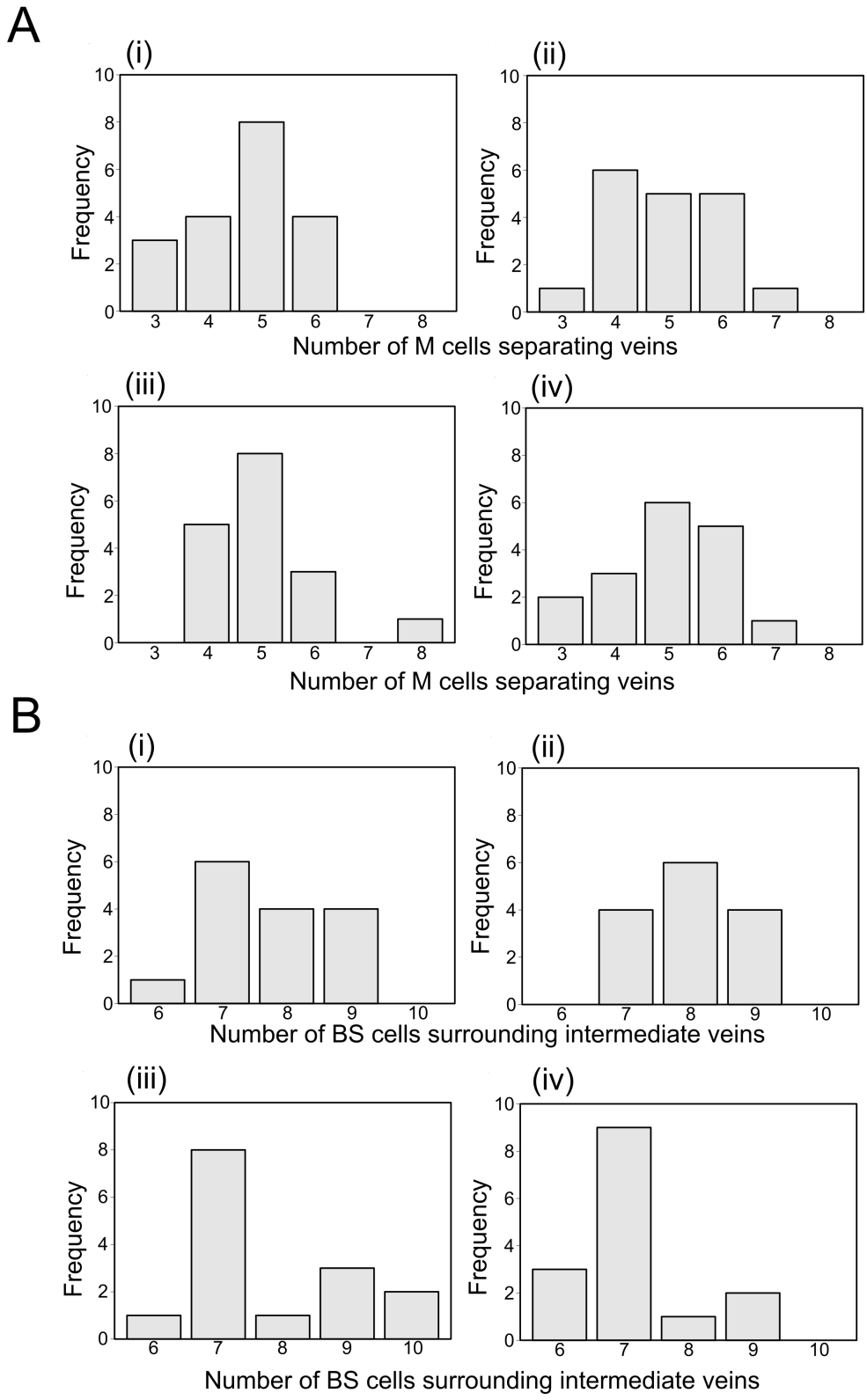


Figure 5.28 Barplots displaying M and BS cell patterning in Kitaake WT and transgenic rice lines following Dex induction of *ZmRVN1* expression. Figure legend is on the next page.

Figure 5.28 Barplots displaying M and BS cell patterning in Kitaake WT and transgenic rice lines following Dex induction of *ZmRVN1* expression. A) Number of M cells separating veins. **B)** Number of BS cells surrounding intermediate veins. In both (A) and (B) Kitaake WT (i) and three independent transgenic lines of construct 17268b (*pOP6::ZmRVN1 – ZmUbi$_{pro}$::LHGRN*) (ii-iv) are displayed: 1b (ii), 2a (iii) and 5b (iv). Cell counts in both (A) and (B) were obtained using fresh transverse sections taken from the mid-point along the proximal/distal axis of fully expanded leaf 2. For each line n=4, with at least 4 M and BS cell counts taken per plant. Plants were germinated on 1μM Dex and sampled 13 days after planting. In both (A) and (B) statistical significance was calculated using mixed effect poisson regression models and subsequent calculation of a likelihood ratio by comparison to a null model. No significant differences were found between WT and transgenic lines ($p > 0.05$). Breaks in y-axis scale are indicated by a double diagonal line.

development (albeit at very low levels). However, even when both *ZmNKD1* and *ZmNKD2* were mutated together only very subtle phenotypic perturbations were revealed in the leaf, and no further severity was observed in *Zmrvn1-4; Zmnkd1-R; Zmnkd2-R* triple mutants.

There is extensive protein similarity in both the *Arabidopsis* and maize IDD genes (Colasanti et al. 2006), indicating the potential for functional redundancy. In *Arabidopsis* roots, mutations in a number of IDD genes elicit no obvious phenotypic perturbations when single mutants are examined, but upon examination of higher order mutants more severe phenotypes are revealed (Long et al. 2015b). This is true of both *BALDIBIS* and *MAGPIE*, for which confirmed roles in root patterning were only identified upon examination of double and triple mutants alongside other IDD genes (Long et al. 2015b). Such redundancy is likely further exacerbated by the recent whole genome duplication in maize, which resulted in ~20% of maize gene content consisting of homeolog gene-pairs, the majority of which show conserved expression profiles in maize early leaf development (Swigonova et al. 2004b; Schnable et al. 2011; Hughes et al. 2014). Indeed, whilst *ZmNKD1*, *ZmRVN1* and *ZmIDD14* were the only IDD genes identified in the bioinformatics filtration used by Wang *et al.* (2013b) to identify putative Kranz anatomy regulators, all three have duplicated pairs, and a number of other IDD genes are expressed at the relevant stages for Kranz development (Wang et al. 2013b).

An added complication is that leaf development involves three axis of development, whereas root development involves only two. For the reasons outlined in section 4.3.1, this may underpin the difficulty in perturbing leaf patterning processes. In the case of the IDD genes, where the reported root phenotypes are relatively subtle, the prospect of deciphering even more subtle phenotypes in the leaf is a significant challenge. This is made even more challenging by the inherent difficulties associated with functional work on C_4 monocots, where currently maize is the only viable system

for undertaking loss-of-function analysis. Advances in the use of *Setaria viridis* as a model C_4 monocot might alleviate this problem in the near future (Brutnell et al. 2010; Li and Brutnell 2011). The phenotyping strategy employed in this study was to examine fresh transverse sections. This allowed precise M and BS cell counts to be quantified and spatial organisation (which is fundamental to Kranz anatomy) to be examined. The work associated with obtaining fresh sections, however, meant that although some subtle effects could be elucidated, the sampling area was relatively small. An alternative strategy, that might identify even more subtle defects, would be to examine longitudinal cleared leaves in order to count cells across a far wider area of the leaf. This strategy could in future help to elucidate further roles, if any, of the IDD genes in Kranz anatomy development.

5.3.2. Evidence for functional interaction between *ZmNKD1/2* and *ZmSCR1/1h*

Rigorous quantitative analysis revealed a subtle but statistically significant increase in occurrences of extra BS cells in the *Zmnkd1-Ds; Zmnkd2-Ds* double mutant. Further support for this was provided by the *Zmnkd1-R; Zmnkd2-R* double mutant, which exhibits the same trend. Whilst in the latter case the trend was not statistically significant, this may reflect the fact that both *Zmnkd1-R* and *Zmnkd2-R* are thought to be hypomorphic alleles, whereas both *Zmnkd1-Ds* and *Zmnkd2-Ds* are thought to be nulls. This phenotype is the same as that exhibited by the *Zmscr1-1; Zmscr1h-1* double mutant, which raises the intriguing possibility that *ZmNKD1/2* and *ZmSCR1/1h* function are linked during the development of Kranz anatomy. This alignment of phenotypic perturbations is supported by the expression profiles of all four genes, as all are preferentially expressed in the M cells of developing maize leaf primordia. Although none of the vein patterning defects associated with the *Zmscr1-1; Zmscr1h-1* double mutant are found in the *Zmnkd1-Ds; Zmnkd2-Ds* background, it is notable that occurrences of extra BS cells are found (at lower frequency) in the

Zmscr1-1 single mutant, which also does not show any vein patterning defects. This supports the hypothesis that *ZmSCR1/ZmSCR1h* expression may be reduced sufficiently in the *Zmnkd1-Ds; Zmnkd2-Ds* background to reveal the increased occurrence of extra BS cells, but not so much as to perturb vein patterning. If correct, this would explain why no additional phenotypic perturbations are revealed in the *Zmscr1-1; Zmnkd1-R; Zmnkd2-R* triple mutant, as it is possible that *ZmSCR1* expression is already reduced to a similar degree in the *Zmnkd1-R; Zmnkd2-R*. This would place *ZmNKD1* and *ZmNKD2* upstream of *ZmSCR1* and *ZmSCR1h*, which is similar to the situation in the *Arabidopsis* root where *AtJKD* initially activates *AtSCR* expression (Welch et al. 2007; Long et al. 2015b). To fully test this hypothesis it will be necessary to assess *ZmSCR1* and *ZmSCR1h* expression in the *ZmNKD* mutant backgrounds.

Some support for the hypothesis that *ZmNKD1* and *ZmNKD2* act upstream to *ZmSCR1* and *ZmSCR1h* is provided by analysis of *nkd* mutant kernel transcriptomes (Phil Becraft, personal communication). Across three biological replicates *ZmSCR1h* transcript levels were significantly reduced in *Zmnkd1-R; Zmnkd2-R* double mutant kernels, albeit from already low expression levels. However, *ZmSCR1*, which is expressed between 20 and 100 times more highly than *ZmSCR1h* in WT kernels, remained unchanged in the *Zmnkd* mutant samples. In vitro, *ZmNKD1* and *ZmNKD2* bind short motifs similar to that recognised by ZmID1 (Bryan Gontarek, personal communication). No *ZmNKD1* motifs are found in the 5kb upstream of the start codon of either *ZmSCR1* or *ZmSCR1h*, however, three and five respectively are found between 5kb and 10kb upstream of each gene. A number of putative *ZmNKD2* motifs are present in each promoter region, however, this motif is so short that it is unclear how specifically it predicts *ZmNKD2* binding rather than IDD transcription factor binding more generally. The gene regulatory networks that underpin *ZmSCR1/1h* function in leaves may be different to those in kernel development, and

as such the fact that *ZmSCR1* expression is not reduced in *Zmnkd1-R; Zmnkd2-R* kernels does not necessarily mean that this is also true in leaves. A comparative transcriptome from developing *Zmnkd1-R; Zmnkd2-R* leaves would be necessary to resolve this.

5.3.3. The lack of a developmental alterations in rice lines expressing *ZmNKD1* and *ZmRVN1* is surprising given the failure to recover constitutive expression lines

Attempts to constitutively overexpress both *ZmNKD1* and *ZmRVN1* in rice failed, with no plants recovered from tissue culture. Whilst experimental failure cannot definitively be ruled out, a number of factors point to a biological cause. First, multiple rounds of transformation were attempted in both cases, with no successfully transformed plants recovered from any attempt. In one case a number of hygromycin resistant plants did regenerate, but subsequent genotyping revealed these to only contain the hygromycin resistance gene rather than the *ZmRVN1* transgene. Second, these results have been mirrored for a number of candidate Kranz regulators, with constitutive overexpression of eight out of the twelve high-priority Kranz regulators identified by Fouracre *et al.* (2014) preventing the regeneration of rice callus. Intriguingly, all eight of these genes were originally identified based on their presence in the FA2 cluster of the foliar/husk dataset generated by Wang *et al.* (2013b). This cluster was designed specifically to capture genes with ascending expression from FP1/2 to FP3/4 and maintenance of this expression between FP3/4 and FP5. This expression profile reflects the expected behaviour of positive regulators of Kranz anatomy. Given the predicted importance of these genes in early leaf development, it is perhaps unsurprising that extending the expression domain into all tissues prevents the regeneration of rice callus. However, it is surprising that in the case of *ZmRVN1*, constitutive expression could be induced by Dex in T1 seed without any obvious morphological or leaf patterning perturbation in the resultant

plants. One explanation for this result could be that constitutive expression of *ZmRVN1* (and possibly *ZmNKD1*) in rice callus affects initial meristem organisation (but not later patterning processes), and thus prevents regeneration from undifferentiated callus, whereas for Dex induction of T1 seed meristem formation was already completed prior to *ZmRVN1* expression. This is supported by the observation that no green tissue or arrested shoots were observed during attempts to regenerate constitutive *ZmNKD1* or *ZmRVN1* lines.

Chapter 6: General discussion

6.1. Overview

Over the past 10 years there has been a resurgence of interest in C_4 photosynthesis, in part due to attempts to engineer the C_4 pathway into C_3 crops (Langdale 2011). C_4 photosynthesis is of research interest due to the enhanced productivity, distinctive leaf anatomy and novel biochemistry that are all associated with the pathway (Langdale 2011). Furthermore, C_4 represents one of the most striking examples of convergent evolution, with over 60 independent origins across the angiosperms (Sage et al. 2011a). In the majority of cases (and in all C_4 crops), C_4 photosynthesis is underpinned by characteristic Kranz anatomy. Although Kranz forms are variable, in all cases Kranz is formed by closely spaced veins, with concentric surrounding wreaths of BS and M cells, which in combination act to enable efficient function of the C_4 pathway (reviewed by Edwards and Voznesenskaya 2011). Our understanding of the genetic regulation of Kranz development is extremely limited, and until recently no molecular components had been identified. Over the last few years, systems biology based approaches have identified a number of candidate genes in maize whose expression is consistent with the expected behaviour of a positive Kranz regulator (Li et al. 2010; Wang et al. 2013b). Using knowledge of the orthologous *Arabidopsis* genes, and by integrating insight from other transcriptome datasets, a testable model for Kranz development was recently proposed (Fouracre et al. 2014). In this thesis aspects of this model have been tested and refined accordingly (Chapters 4 and 5). Furthermore, genome wide expression data from the developmental stages at which Kranz develops was used as the basis for the first genome-wide assessment of maize homeolog gene-pair divergence (Chapter 3, Wang et al. 2013b; Hughes et al. 2014).

The combined significance of the results presented in Chapters 3-5 is considered here. First, a revised model of Kranz patterning is proposed that integrates maize duplicate genes and takes into account those genes for which Kranz patterning roles have now been confirmed (Section 6.2). A consideration of the advantages and disadvantages of maize as a model C_4 species is then presented, with a discussion of how future Kranz research should proceed (Section 6.3). Next, an interpretation of these results in the context of C_3 leaf anatomy is provided (Section 6.4), followed by a consideration of the relative merits of rice overexpression as a means of testing maize Kranz candidates (Section 6.5). Finally, the results are discussed in the context of C_4 rice engineering (Section 6.6).

6.2. A revised *SCARECROW* based model for radial patterning in maize leaves

It was recently proposed that the SCR/SHR pathway acts in C_4 leaves to regulate the characteristic radial patterning of BS and M cells associated with Kranz anatomy (Slewinski et al. 2012; Wang et al. 2013b; Slewinski et al. 2014; Fouracre et al. 2014). This hypothesis was based broadly on five pieces of evidence: *1) ZmSCR1, ZmSHR1* and *ZmSHR2* are all expressed early in foliar leaf development, and do not show the same expression profile in maize husk leaves (Wang et al. 2013b). *2)* In separated BS and M cell transcriptomes from mature maize leaves, *ZmSCR1* and *ZmSHR1* (but not *ZmSHR2*) were preferentially expressed in the BS (Li et al. 2010; Chang et al. 2012). *3)* In *Arabidopsis*, AtSCR and AtSHR regulate endodermal and cortical cell patterning in roots (Benfey et al. 1993; Laurenzio et al. 1996), and the root endodermis has long been considered as an analogous cell type to the leaf BS (Esau 1965). *4)* In *Arabidopsis* roots, *SCR* and *SHR* expression is regulated by a suite of IDD transcription factors (Welch et al. 2007; Ogasawara et al. 2011; Long et al. 2015b), and a cohort of maize IDD genes that show cell-preferential expression is upregulated during early foliar leaf development (Wang et al. 2013b). *5)* Single

Zmscr1 and *Zmshr1* mutants were reported to exhibit subtle BS and M cell patterning defects in the maize leaf (Slewinski et al. 2012, 2014).

Based on this evidence, a testable model was proposed for SCR/SHR Kranz patterning (Figure 6.1.A) (Fouracre et al. 2014). In this model, ZmSHR1 was proposed to act as the previously hypothesised mobile signal emanating from the vein that regulates radial BS and M patterning (Langdale and Nelson 1991; Fouracre et al. 2014). Despite the endodermis and BS being considered analogous cell-types (Esau 1965), C_4 leaf patterning involves the patterning of an additional cell layer relative to *Arabidopsis* root patterning (Fouracre et al. 2014). This is because, in addition to the BS cell layer, both C_4 and C_3-like M cells can develop in the leaf (Langdale et al. 1988; Fouracre et al. 2014). For example, in the maize husk leaf, where C_4 photosynthesis is undertaken locally around veins, C_4 BS and M cells develop immediately adjacent to the vein, however, beyond these two cell layers M cells are C_3-like (Langdale et al. 1988; Pengelly et al. 2011). Thus, any plausible model for Kranz patterning must include a mechanism for specifying both BS and C_4-like M cells adjacent to veins, with the assumption that C_3-like M cell development is default. In its current form the *Arabidopsis* SCR/SHR model of root patterning does not do this, as this model only predicts how endodermal cell identity is triggered instead of the default cortical identity (Cui et al. 2007; Wu et al. 2014). The finding that *ZmSHR1* is expressed in both veins and BS cells in mature maize leaves lead to the suggestion that in C_4 leaves a decreasing SHR protein gradient might occur from developing veins through BS cells to C_4-M cells (Li et al. 2010; Chang et al. 2012; Fouracre et al. 2014). In the absence of SHR, which would occur in maize husk leaves in M cells not immediately adjacent to a BS cell, default C_3-like M cells are specified (Fouracre et al. 2014). In both BS and C_4 M cells the model proposes that ZmSHR1 is sequestered to the nucleus by ZmSCR1, where it promotes a positive

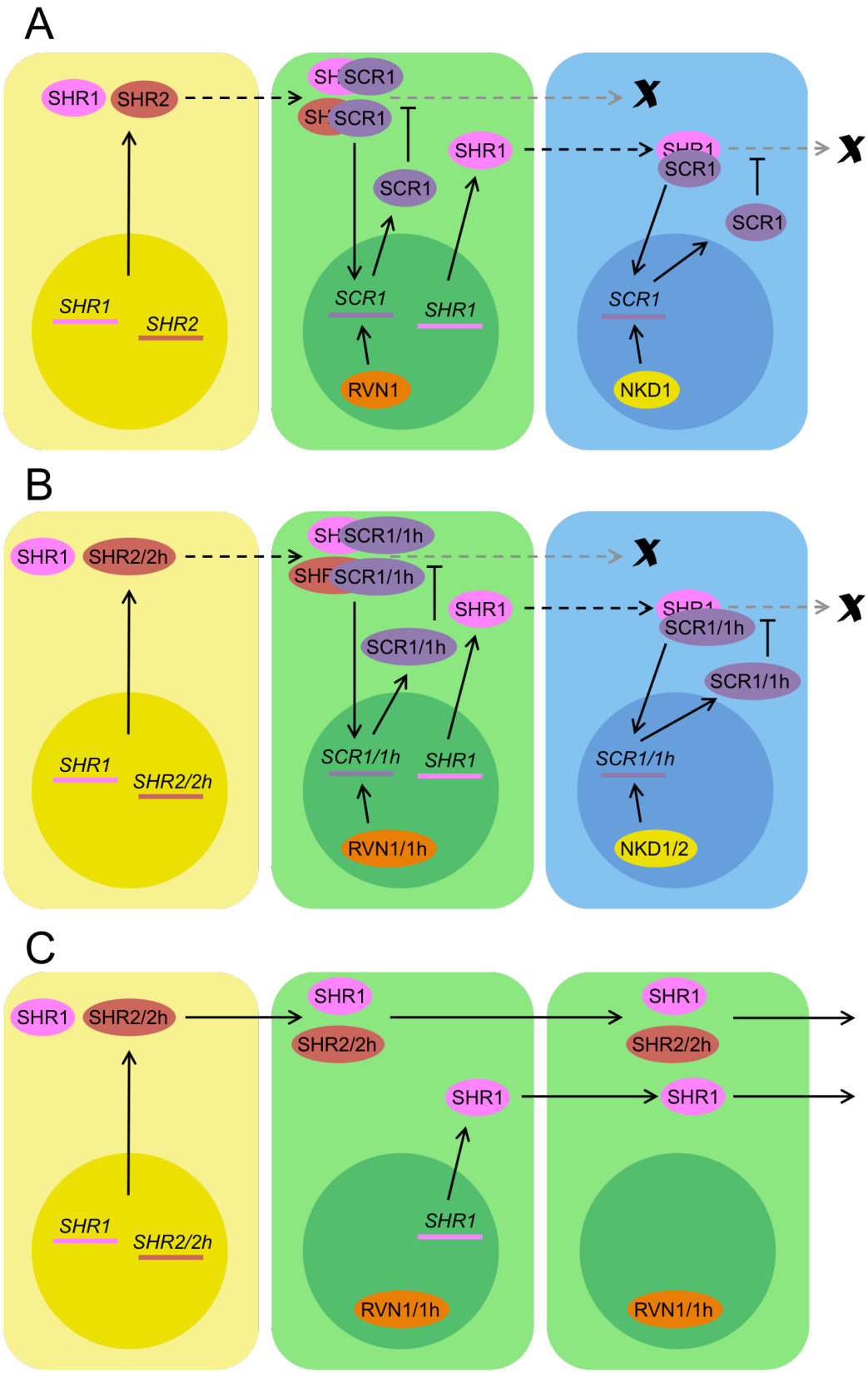


Figure 6.1 Cartoon depictions of Kranz radial patterning models in maize.

Figure legend is on the next page.

Figure 6.1 Cartoon depictions of Kranz radial patterning models in maize. A) Radial patterning of Kranz anatomy as proposed by Fouracre *et al.* (2014). **B)** Revised radial patterning model incorporating maize duplicate genes. Roles for *ZmSCR1/1h* and *ZmNKD1/2* have been confirmed in this study, *ZmSHR1* involvement was confirmed previously (Slewinski et al. 2014). The roles of *ZmSHR2/2h* and *ZmRVN1/1h* remain unconfirmed. **C)** A mechanism for extra BS and veins with no separating M cells formation in *Zmscr1-1; Zmscr1h-1* double mutants, in which the *SHR* protein gradient is proposed to extend into C_4 M cells to a greater extent than in WT. In all diagrams yellow represents vein, green represents BS and blue represents C_4 M. Circles represent the nucleus. Genes are represented by straight lines and italicised gene names. Coloured ovals represent proteins. Solid black arrows within the nucleus or entering the nucleus indicate the promotion of gene expression. Solid black arrows leaving the nucleus indicate transcription and subsequent translation of a protein. Black dashed arrows represent protein movement, grey dashed arrows indicate inhibited protein movement. Lines with flat ends represent the inhibition of protein movement. Duplicated genes are indicated by / within a protein oval, and are not included as separate proteins due to space constraints.

feedback loop in which further *ZmSCR1* expression is triggered (Fouracre et al. 2014). In itself, this would not differentiate between BS and C_4 M cells, however, the finding that different IDD transcription factors are accumulated in distinct cell-types (ZmRVN1 in the BS, ZmNKD1 in C_4 M), might provide a mechanism for BS and M differentiation (Li et al. 2010; Chang et al. 2012; Fouracre et al. 2014).

There are, however, a number of unresolved issues regarding this model. For example, in the model ZmSCR1 is critical in regulating the extent of SHR movement between adjacent cell layers. This predicts that in the absence of ZmSCR1, ZmSHR1 could extend beyond the normal domain and trigger additional BS cell layers (Figure 6.1C). This phenotype was observed in the *Zmscr1* mutant, but only at very low frequency (Slewinski et al. 2012, Chapter 3). In this study, it was hypothesised that this low frequency might be because ZmSCR1 functions redundantly with its previously overlooked homeolog pair ZmSCR1h (Figure 6.1B). This hypothesis was validated, as higher frequencies of extra BS cells and veins with no separating M cells were found in the *Zmscr1-1; Zmscr1h-1* double mutant than in either of the single mutants. Furthermore, the *Zmscr1-1; Zmscr1h-1* double mutant exhibited a suite of phenotypes, including reduced growth, drooping leaves and vein-patterning perturbations, that were not seen in either of the single mutants. However, in terms of Kranz patterning there are still many vascular bundles that differentiate normally in the *Zmscr1-1; Zmscr1h-1* double mutant, with one BS layer and one M cell layer surrounding the vein. This indicates that there must be additional complexity in Kranz radial patterning that is not currently captured in the SCR/SHR model.

One gene that could not yet considered in Kranz patterning is *ZmSCR2,* which is the maize orthologs of *AtSCL23*. In *Arabidopsis*, *AtSCL23* is the closest homolog of *AtSCR*, and has been implicated in both root patterning and BS differentiation in leaves (Cui et al. 2014; Long et al. 2015a). Interestingly, a more severe mutant

phenotype was observed in the leaves of an *Atscr; Atscl23* double mutant than in either single mutant (Cui et al. 2014). *ZmSCR2*, which has no homeolog gene-pair in maize, was not identified as a candidate Kranz regulator because its expression is much higher in mature foliar and husk leaves than during early leaf development (Wang et al. 2013b). In mature leaves, *ZmSCR2*, is expressed preferentially in M cells, however, it is not known whether this expression profile is similar during early leaf development (Li et al. 2010; Chang et al. 2012). Given the redundancy between AtSCR and AtSCL23, it remains possible that ZmSCR2 plays a role in Kranz patterning, and could partially compensate in the *Zmscr1-1; Zmscr1h-1* double mutant. A *Zmscr1-1; Zmscr1h-1; Zmscr2-1* triple mutant is currently being generated, but given the severity of the growth and drooping leaf phenotype associated with the *Zmscr1-1; Zmscr1h-1* double mutant, it is possible that generation of this triple mutant will be unsuccessful.

As illustrated by ZmSCR1 and ZmSCR1h, the identification of a number of maize-specific duplicate genes that were not considered by Fouracre *et al.* (2014) complicates the testing of the Kranz model. Given that it is most parsimonious that these duplicates function redundantly, roles for ZmSCR1h, ZmSHR2h, ZmRVN1h and ZmNKD2 must be considered alongside the previously identified partners (Figure 6.1B). In the case of ZmSCR1h and ZmNKD2 this study has provided evidence that both pairs of genes function redundantly during Kranz patterning (Figure 6.1B). Furthermore, preliminary evidence is provided by this study that ZmNKD1 and ZmNKD2 function in the same pathway as ZmSCR1, as indicated by the finding that the *Zmnkd1-R; Zmnkd2-R; Zmscr1-1* triple mutant did not appear to exhibit more severe leaf perturbations than either single *Zmscr1-1* or double *Zmnkd1-R; Zmnkd2-R* mutants. However, the *Zmnkd* leaf phenotype is relatively subtle, and there are therefore likely to be other M-specific factors that contribute to Kranz patterning. It is possible that these factors could be additional IDD genes that

function redundantly with *ZmNKD1/2*, such as *ZmIDD14.* In the case of *ZmRVN1*, no evidence of BS or Kranz defects were observed in either the single *Zmrvn1-4* mutant or in the *Zmnkd1-R; Zmnkd2-R; Zmrvn1-4* triple mutant. The lack of available transposon insertions in *ZmRVN1h* means further testing is necessary to validate a role for the ZmRVN genes in BS patterning. However, *ZmRVN1* and *ZmRVN1h* are the only IDD genes preferentially expressed in the maize BS of mature leaves (Li et al. 2010; Chang et al. 2012), and although it is not known whether this is also true in developing leaf primordia, the ZmRVN transcription factors remain strong candidates for having a role in Kranz BS patterning.

In *Arabidopsis* roots, it is known that the IDD transcription factors AtJKD and AtMGP physically interact with both AtSCR and AtSHR and form part of the AtSCR/AtSHR complex that is sequestered to the nucleus to regulate downstream gene-expression (Welch et al. 2007). In the Kranz radial patterning model proposed by Fouracre *et al.* (2014), both ZmSHR genes and ZmSCR are proposed to accumulate in both BS and C_4 M cells, however, it is not clear what would prevent all BS accumulated SHR being sequestered to the nucleus by SCR. One possibility is that a transcriptional *ZmSCR1/1h* gradient occurs in the opposite direction to the ZmSHR gradient, which could result in ZmSCR1/1h levels in the BS being insufficient to sequester all ZmSHR. This could enable dosage-related control of cell-identity, as the relative amounts of ZmSCR1/1h and ZmSHR would differ between cell-types. Furthermore, it remains possible that ZmSHR1 and ZmSHR2/2h have different transcriptional targets. However, the only other candidate Kranz regulators proposed to accumulate in a highly cell-preferential manner are *ZmNKD1* and *ZmRVN1* (Fouracre et al. 2014). Therefore, testing whether ZmRVN1 or ZmNKD1 can physically interact with ZmSCR and/or ZmSHR genes in a similar manner to that observed between AtJKD/AtMGP/AtSCR/AtSHR could provide a mechanism by which BS and C_4 M cells are specified.

6.3. The strengths and limitations of maize as a model species

Maize is one of the most important global crops, with an annual production only eclipsed by that of sugar cane (FAO 2015). Furthermore, maize has long been established as a critical model organism for a range of research disciplines, including plant domestication, genome evolution, developmental physiology, epigenetics, and C_4 photosynthesis (Strable and Scanlon 2009). This study has highlighted both the advantages that have contributed to the importance of maize to plant research, but has also identified a number of limitations and disadvantages that currently hold back research into the regulation of Kranz patterning.

6.3.1. The maize whole genome duplication adds to potential genetic redundancy in maize

In the grass lineages, at least two whole genome duplications at the base of the clade, and an additional duplication at the base of the monocots, have contributed to the extensive genetic redundancy found in modern grasses (Paterson et al. 2004; Tang et al. 2010; D'Hont et al. 2012; Jiao et al. 2014; McKain et al. 2016). In maize, this situation is further exacerbated by a whole genome duplication that occurred after the divergence with close relative sorghum (Swigonova et al. 2004a). This has resulted in even more potential redundancy in the maize genome, with ~20% of maize gene content consisting of retained homeolog gene-pairs which have single-copy orthologs in the sorghum genome (Schnable et al. 2011; Hughes et al. 2014). Maize has long been one of the main model species for C_4 photosynthesis and Kranz anatomy research (Brown et al. 2005). Recently, our limited understanding of the transcriptional regulation underpinning Kranz anatomy lead to a number of RNA-seq based approaches which sought to identify Kranz regulators (Li et al. 2010; Liu et al. 2013; Wang et al. 2013b; Fouracre et al. 2014; Tausta et al. 2014; Wang et al. 2016). However, the impact of the maize-specific whole genome duplication was not considered in the analysis of these experiments.

To determine the extent to which homeolog gene-pairs have diverged in expression at the relevant stages of Kranz development, this study took advantage of the expression profile assignment of all expressed maize genes in a previous foliar and husk leaf transcriptome comparison (Wang et al. 2013b). This analysis fulfilled two purposes. First, it enabled the first genome-wide assessment of regulatory divergence of maize homeolog gene-pairs (Hughes et al. 2014). Second, it enabled specific information regarding high-priority candidate Kranz regulators and their relevant homeolog pairs to be gathered. In regard to the first aim, the study revealed that regulatory divergence in the form of regulatory neofunctionalisation is a significant phenomenon in maize, with 16% of expressed homeolog gene-pairs showing evidence of divergence. Given that this analysis only considered broad changes in expression at a few developmental stages, it likely represents an underestimate. In addition, homeolog gene-pairs that had diverged in expression were found to be under significantly higher purifying selection with consistent subgenome-1 expression dominance bias. Together, these results highlight the importance of maize as a model species for understanding genome evolution following whole genome duplication.

In relation to the second aim of better understanding how the maize whole genome duplication related to Kranz development, this study revealed that a number of high priority Kranz regulators had previously unidentified homeolog gene-pairs. This conclusion was further supported by phylogenetic tree inference for *ZmSCR1, ZmSHR2* and *ZmRVN1*. Furthermore, the previously identified *ZmNKD1* and *ZmNKD2* were revealed to be maize-specific duplicates, despite not being included in the published homeolog dataset (Schnable et al. 2011; Yi et al. 2015). Within these gene-pairs there were examples of both divergent and non-divergent expression profiles. It was established that the majority of homeolog gene-pairs are under strong purifying selection to maintain the coding sequence of the gene. Whilst

this provides no direct evidence that protein function has remained unaltered, because single amino acid changes could in theory result in protein changes, it does indicate that there is a strong selective pressure to maintain the protein sequence. As such, it is more parsimonious to expect there to be extensive functional redundancy between homeolog protein pairs.

Evidence for functional redundancy of two homeolog gene-pairs was obtained in this study. Both *ZmSCR1/1h* and *ZmNKD1/2* exhibited more severe leaf patterning phenotypes in double mutants than in the respective single mutant backgrounds, and *ZmNKD1/2* are known to function redundantly in aleurone patterning in developing maize kernels (Yi et al. 2015). In the case of *ZmSCR1/1h* this result was predicted, as the gene pair were allocated to the same expression profiles and thus classified as non-regulatory neofunctionalised, and in situ hybridisation has demonstrated that both genes shown similar preferential M expression during early foliar leaf development (Olga Sedelnikova, personal communication). In addition, the protein sequences show high homology and high purifying selection relative to the sorghum *SCR* orthologs, and thus there is unlikely to be significant divergence in terms of protein function. Although not formally classified (as it was not included in the list of homeolog gene-pairs), *ZmNKD1/2* would have been classified as regulatory neofunctionalised in the foliar leaf, as *ZmNKD1* has ascending expression (FA) whereas *ZmNKD2* has neutral expression (FN). Both genes have neutral expression in the husk. However, despite exhibiting a distinct expression profile in the developing foliar leaf, *ZmNKD2* is still expressed in this context, and in addition it is currently not known how *ZmNKD2* expression is altered in the absence of *ZmNKD1*. These results indicate that detailed functional validation is always necessary to fully understand potential redundancy or divergence of a duplicated gene-pair. All of these duplicated genes have now been integrated into the revised model for Kranz patterning (Section 6.2, Figure 6.1B).

6.3.2. Maize transposon insertion resources are a valuable but limited resource for loss-of-function analysis

The ability to conduct loss-of-function analysis is critical for understanding gene function in any model organism. In maize this has been facilitated through the generation of transposon insertion libraries, whereby transposon insertions into genes are identified in a mobilised population, and desirable insertions fixed to facilitate genetic analysis (reviewed by McCarty and Meeley 2009; Strable and Scanlon 2009). This study took advantage of three distinct mutant libraries, the UniformMu and Mu-Illumina collections (Settles et al. 2007; Williams-Carrier et al. 2010), both of which utilise the *Mutator* transposon, and the *Ac/Ds* collection, which utilises the *Ds* transposon (Vollbrecht et al. 2010). These resources have enabled extensive gene function discovery in maize, and have enabled roles for *ZmSCR1, ZmSCR1h*, *ZmSHR1*, *ZmNKD1* and *ZmNKD2* in the development of Kranz anatomy to be confirmed (Chapters 4 and 5, Slewinski et al. 2012, 2014). However, these resources do not provide full coverage of maize genes, and for a number of high-priority Kranz regulators no transposon insertions were available. This was a particular problem in the case of *ZmSHR2h* and *ZmRVN1h*, as in the absence of insertions into each homeolog gene the lack of perturbed phenotype in *Zmshr2-1* or *Zmrvn1-4* mutant lines cannot be taken as conclusive proof that these genes do not function in Kranz development. Both *ZmSHR2h* and *ZmRVN1h* are expressed at far lower levels during early foliar leaf development than their respective partners *ZmSHR2* and *ZmRVN1* (both ~40 fold lower). However, it is not clear whether expression of either of these genes is upregulated in either the *Zmshr2-1* or *Zmrvn1-4* mutant. Previous work in yeast has suggested that compensation by upregulation was rare on a genome scale (Wong and Roth 2005), however, this has not been investigated in plants where the frequency of polyploidy has resulted in extensive potential genetic redundancy.

6.3.3. Transgenic approaches are necessary for higher-throughput analysis of Kranz candidate genes and refinement of the current model

One of the disadvantages of maize as a model organism is that whilst *Agrobacterium* mediated genetic transformation of immature embryos is relatively efficient, the process is slow and requires a large amount of dedicated greenhouse space (Ishida et al. 2007; Strable and Scanlon 2009; Que et al. 2014). Therefore, often the most feasible way to generate transgenic maize is to outsource the transformation to a third-party, which greatly increases both the cost and associated logistical considerations. However, where no transposon insertions are available in a particular gene, genetic transformation is the only way to conduct loss-of-function analysis. Furthermore, genetic transformation greatly improves the speed and flexibility of experiments. RNAi and CRISPR constructs have the potential to target multiple similar genes. For example a construct could be designed that would knock-down expression of *ZmSHR1, ZmSHR2* and *ZmSHR2h* simultaneously. Obtaining this triple mutant (even if insertions were available in all genes) would take at least four generations. Assessing potential movement of ZmSHR proteins both in WT leaves and *Zmscr* mutants could also be carried out by transforming protein constructs tagged with fluorescent marker genes. Thus, further refinement of the SCR/SHR patterning model in maize leaves and validation of proposed roles of other genes requires a genetic transformation approach.

Whilst maize is undoubtedly an important model species for C_4 research, the difficulties associated with its large stature, slow generation time, duplicated genome and expensive genetic transformation, has meant that there have long been discussions about alternative C_4 models (Brown et al. 2005; Langdale 2011). A new population of sorghum mutants might go someway to facilitating further Kranz gene validation, however most of the disadvantages of maize are equally associated with sorghum (Jiao et al. 2016). Recently, the closely related millets *Setaria viridis* and

Setaria italica (also known as foxtail millet) have been proposed as offering an alternative monocot system in which to investigate C_4. The attraction of the species is clear; both are short in stature, have a rapid generation time (~2 months) and can easily be grown in the greenhouse under similar conditions to that of *Arabidopsis* (Brutnell et al. 2010). Furthermore, both species have small genomes, without the complication of a recent whole genome duplication, that have been recently sequenced (Bennetzen et al. 2012). However, despite a number of *Agrobacterium* mediated transformation protocols being published (Yinghui et al. 2007; Brutnell et al. 2010; Li and Brutnell 2011; Martins et al. 2015; Saha and Blumwald 2016), efficient and stable transformation of *Setaria* is still a considerable challenge for most research groups. After much effort in Oxford, no significant improvements in transformation efficiency were observed, and despite antibiotic selection, non-transformed plants that survived selection were so frequent that the approach was rendered unfeasible (data not shown). More recently, however, one research group has demonstrated that they can reliably transform *Setaria*, and are offering a paid service to transform constructs, as well as offering to share their expertise (Joyce Van Eck, personal communication).

Any functional insight that can be provided in *Setaria* will be complementary to that of maize, because *Setaria* is known to have an independent C_4 origin to that of maize and sorghum (Grass Phylogeny Working Group II 2012). As such, *Setaria* could provide crucial insight into the extent to which developmental pathways have been independently recruited to regulate Kranz anatomy and C_4 development. This approach has already successfully identified genes that have been independently duplicated in the maize and *Setaria* C_4 lineages (Emms et al. 2016). The lack of a recent whole genome duplication in the *Setaria* lineage, means that many duplicate maize genes should be single copy in *Setaria*, which could greatly simplify genetic analysis of the underlying regulatory pathways. In summary, at the start of this study

maize insertion mutant libraries represented the best and most reliable means to investigate the function of a number of high-priority candidate Kranz regulators. However a detailed understanding of the mechanisms underpinning Kranz development now requires transgenic approaches. Whilst in some cases the generation of transgenic maize lines will be feasible, the use of *Setaria* species is becoming more and more attractive for understanding C_4 photosynthesis.

6.4. How is C_3 leaf development regulated, and how does this differ from C_4 species?

A recent analysis identified a cohort of transcription factors enriched for BS expression that were independently duplicated in two independent C_4 origins but not in closely related C_3 species (Emms et al. 2016). The work by Emms *et al.* (2016) demonstrated the potential for duplicated genes to underpin C_4 regulation, and duplication followed by neofunctionalisation has previously been proposed to increase the probability of C_4 evolution (Monson 2003; Wang et al. 2009; Sage et al. 2012). However, all of the known C_4 pathway components are also found in C_3 species, and there is evidence that some C_3 species may use versions of these enzymes to operate a spatially separated C_4 cycle (Hibberd and Quick 2002). Therefore, there has been extensive research into the changes that have occurred in the regulation of these genes to confer C_4-like expression patterns (reviewed by Williams et al. 2012). For the C_4 pathway enzymes, a number of *cis*-elements in gene promoters have been defined that confer C_4-like expression, however, examples of significant regulatory regions in the UTRs and even coding sequence regions of C_4 enzymes have also been found (Brown et al. 2011; Williams et al. 2012). In the absence of confirmed Kranz regulators, this analysis has not yet been possible for the regulation of leaf anatomy. However, the same considerations must apply, because C_3 leaves still form organised veins, BS and M cell layers, albeit at different frequencies with distinct morphologies. Given that all of the candidate Kranz

regulators identified in this study have orthologs in rice, there is an underlying question of how ortholog function differs between Kranz development in maize and C_3 development in rice. Moreover, many of these genes are expressed in maize husk leaves (albeit at much lower levels), indicating that genes identified as candidate Kranz regulators might also function in C_3-like leaf patterning in maize (Wang et al. 2013b).

During early leaf development, the expression profiles of the rice orthologs of a subset of high-priority candidate Kranz regulators are different from those exhibited in maize (Olga Sedelnikova, personal communication). Therefore, it is most likely that these genes acquired novel regulatory sequences that confer a Kranz-specific expression profile. In the case of *ZmSCR1* and *ZmSCR1h* this study provided evidence that the genes have evolved a Kranz-specific role, as maize husk leaves did not exhibit the same perturbations in terms of vein density, vein order or veins with no separating M cells as seen in the foliar leaf. It is not clear what underpins this result, as the expression of *ZmSCR1* or *ZmSCR1h* in husk primordia could not be spatially assessed. To fully understand the extent to which SCR (and other Kranz regulators) function specifically in C_4 leaf development, further functional and bioinformatics analysis is necessary using both C_3 and C_4 monocots.

6.5. Functional validation of candidate Kranz regulators in rice

A major part of the strategy employed in this study to validate the role of candidate Kranz regulators was to express the genes in rice, and screen for alterations in leaf anatomy. This enabled the sufficiency of putative positive regulators of Kranz anatomy in maize to be tested in rice. This has obvious relevance to C_4 rice engineering attempts, and was supported by evidence from rice populations with altered vein spacing, which indicated that rice may be able to tolerate alterations in leaf anatomy traits associated with C_4 photosynthesis (Smillie et al. 2012; Feldman et

al. 2014). Of course there is no guarantee that expression of a maize gene in rice will reveal anything about gene function in the endogenous maize context, and important Kranz regulators might not alter rice leaf patterning in the absence of maize *trans*-acting components. However, constitutive expression of maize candidate genes was employed as an initial test to determine whether a gene has any potential to alter rice leaf development.

In this study, expression of maize genes in rice failed to yield information regarding potential Kranz patterning roles. Rice lines expressing *ZmSCR1* under constitutive, M or BS preferential promoters showed no obvious anatomical difference to WT. In contrast, rice callus transformed with *ZmNKD1* and *ZmRVN1* constitutive expression constructs failed to regenerate. This indicated that these genes affected the regeneration of rice plants from callus, and as such it was proposed that inducible expression or expression from the endogenous maize promoter might reveal altered rice development in T1 lines. Full testing was not possible for *ZmNKD1*, as *ZmNKD1*$_{pro}$*::ZmNKD1* plants failed to set seed and insufficient T1 seed were available for inducible overexpression constructs. However, the lack of perturbed leaf anatomy in T0 *ZmNKD1*$_{pro}$*::ZmNKD1* plants did not support the hypothesis that *ZmNKD1* expression can alter leaf development in rice. Furthermore, neither *ZmRVN1*$_{pro}$*::ZmRVN1* or inducible *ZmRVN1* expression altered rice development in any discernable way. These results contrast with those found for other candidate Kranz regulators for which constitutive expression prevented recovery of transgenic rice plants. For example, expression of a *ZmSHR1*$_{pro}$*::ZmSHR1* construct in rice resulted in extra BS cell formation, under-developed veins and altered stomatal patterning in T0 leaves, and the lines were sterile (Daniela Vlad, personal communication). Furthermore, in rice lines transformed with an inducible *ZmSHR1* construct, where low levels of *ZmSHR1* expression occurred even in the absence of induction, a layer of BS-like cells formed between adjacent veins, with M cells

positioned both ad and abaxially to the layer (Daniela Vlad, personal communication). In the case of *ZmDOT1*, expression driven by the maize endogenous promoter also prevented regeneration of the rice callus, with short green shoots forming and then terminating. These results indicate that the expression of single maize genes can alter rice development and leaf anatomy, however, the genes considered in this study did not.

Interpreting the lack of ZmSCR1 overexpression phenotype in rice is complicated by the fact that it is not yet known whether an endogenous SCR/SHR pathway functions in rice leaves. If an analogous model to that proposed by Fouracre *et al.* (2014) for C_4 patterning operates to pattern C_3 rice leaves, it would be predicted that constitutive ZmSCR1 would prevent OsSHR movement out of developing veins and thus result in perturbations to BS cell formation. However, this was not observed. Overexpression of *OsSCR* in rice also resulted in no obvious leaf defects and only a very subtle root perturbation, and no *Arabidopsis* line with *AtSCR* overexpression has been analysed yet (Christophe Perin, Kim Gallagher, personal communication). Given that ZmSCR1 was able to complement the *Atscr* root phenotype, it is likely that ZmSCR1 is able to interact with AtSHR (Lim et al. 2005). Since maize and rice are more closely related than maize and *Arabidopsis*, it is most parsimonious to suggest that ZmSCR1/1h can interact with OsSHR1/2. If this is the case, however, it is not clear why overexpression of either *OsSCR* or *ZmSCR1* in rice did not result in phenotypic perturbations. These results indicate that overexpression of *ZmSCR1* alone is unlikely to be of use in C_4 rice engineering.

6.6. Prospects for C_4 rice engineering

Whilst previous attempts have been made to engineer a single-celled C_4 cycle into rice, it is now thought that efficient function of an engineered C_4 cycle requires Kranz anatomy (Taniguchi et al. 2008; Miyao et al. 2011). This is a significant challenge,

because until recently no developmental regulators had been shown to play a role in regulating the development of Kranz anatomy in C_4 species. Therefore, attempting to engineer Kranz into rice must take place alongside basic research into how Kranz is regulated in its endogenous context. Transcriptome approaches have facilitated more rapid identification of candidate Kranz regulators, and from this work a number of high-priority candidates have emerged (Wang et al. 2013b; Fouracre et al. 2014). However, in the absence of confirmed functional validation, the majority remain just candidates. Overexpression of candidate genes in rice might have benefits beyond Kranz engineering, as genetic redundancy in maize makes loss-of-function analysis challenging and time-consuming. However, as discussed in the previous section, overexpression of candidates in rice has its own disadvantages. Nevertheless, if C_4 rice is to contribute to future food security, research must continue into the potential of maize Kranz regulators to alter leaf anatomy in rice.

This study has confirmed the role of two redundant gene pairs in the radial patterning of Kranz anatomy. These genes (*ZmSCR1* and *ZmSCR1h*, and *ZmNKD1* and *ZmNKD2*), represent some of the very first confirmed Kranz regulators, and are therefore of considerable interest to C_4 rice engineering. However, given that gene expression in rice failed to confer altered rice leaf anatomy, it appears likely that engineering Kranz anatomy into rice will require a more complex approach than simply modifying the activity of single genes. The advent of Golden Gate cloning technology means that this is now a feasible aim (Weber et al. 2011; Werner et al. 2012), with multi-gene constructs now being quick and easy to clone and transform. However, our knowledge of the underlying genes that regulate radial patterning in maize is still extremely limited. Therefore, there must be a renewed focus on elucidating the regulatory systems that underpin Kranz patterning in C_4 species, before more complex engineering approaches in rice can be undertaken. In particular, the genes considered here must be considered alongside those that are

hypothesised to function in procambium initiation, the increase of which is responsible for the characteristic higher vein density of C_4 species. It might be that expression of proposed procambium initiation regulators, such as *ZmDOT1*, *ZmANT1* and *ZmATL1,* is more likely to yield altered leaf anatomy in rice given the possible overlap between rice and maize SCR/SHR pathways.

In summary, this study has tested a number of aspects of the model put forward by Fouracre *et al.* (2014) concerning the radial patterning of Kranz cell-types. In doing so, some of the very first confirmed regulators of Kranz anatomy have been characterised, with the results used to refine the previous model and inform future experimentation strategies. This work also enabled the first genome-wide assessment of homeolog-gene pair divergence in maize, and this has proved informative both to our understanding of genome evolution following a whole genome duplication, and also to Kranz anatomy development in maize.

Appendices

Appendix 1: List of primers used in this study

Appendix 2: List of maize mutant alleles identified in this study

Appendix 3: All homeolog gene-pairs, and R-NF category gene-lists.

Appendix 4: Enrichment of annotation terms for each R-NF category relative to all annotated maize genes.

Appendix 5: Enrichment of annotation terms for each R-NF category relative to all annotated maize homeolog genes.

Appendix 6: Bayesian model comparison to infer ancestral maize expression profile.

Appendix 7: Homeolog genes identified as being under positive selection.

Appendix 8: Maximum likelihood phylogenetic trees of GRAS and IDD transcription factors.

Appendix 9: Preliminary analysis of constitutive *ZmIDD14* expression in rice

Appendix 10: Representative genotyping of 17282 T1 rice lines

Appendices 2, 4, 6, 7, 9 and 10 are printed below, all appendices are included digitally on an accompanying CD (along with a digital copy of this thesis).

Appendix 2: List of maize mutant alleles identified in this study

Gene	Allele	Insertion type	Insertion in	Known additional insertions (for UniformMu lines)	Insertion verified by PCR?	Phenotyped and crossed?
ZmSCR1	Zmscr1-1	UniformMu	CDS	None	Yes	Yes
ZmSCR1	Zmscr1-2	Mu-Illumina	CDS		Yes	
ZmSCR1	Zmscr1-3	Mu-Illumina	CDS		Yes	
ZmSCR1	Zmscr1-4	Mu-Illumina	CDS		Yes	
ZmSCR1h	Zmscr1h-1	UniformMu	CDS	None	Yes	Yes
ZmSCR1h	Zmscr1h-2	UniformMu	CDS	Yes (4)	Yes	
ZmSCR1h	Zmscr1h-3	UniformMu	5'UTR	Yes (1)	Yes	
ZmSCR1h	Zmscr1h-4	Mu-Illumina	CDS		No	
ZmSHR2	Zmshr2-1	UniformMu	CDS	Yes (5)	Yes	Yes
ZmNKD1	Zmnkd1-R	SNP (suspected) Yi et al. (2015)	CDS		Yes	Yes
ZmNKD1	Zmnkd1-Ds	Ac/Ds	CDS		Yes	Yes
ZmNKD2	Zmnkd2-R	Retrotransposon (suspected) Yi et al. (2015)	5'UTR		Yes	Yes
ZmNKD2	Zmnkd2-Ds	Ac/Ds	5'UTR		Yes	Yes
ZmRVN1	Zmrvn1-1	UniformMu	CDS	Yes (>10)	Yes	
ZmRVN1	Zmrvn1-2	UniformMu	5'UTR	Yes (>10)	Yes	
ZmRVN1	Zmrvn1-3	Mu-Illumina	CDS		Yes	
ZmRVN1	Zmrvn1-4	Mu-Illumina	CDS		Yes	Yes

Appendix 4: Enrichment of annotation terms for each R-NF category relative to all annotated maize genes.

A) Non-R-NF:

Term	Description	p-value	Total genes with annotation	Non-R-NF genes with annotation
GO:0003735	structural constituent of ribosome	3.29E-17	887	157
GO:0005840	ribosome	5.28E-17	873	154
GO:0005622	intracellular	1.79E-16	1856	306
GO:0006412	translation	1.23E-13	1071	172
GO:0019787	small conjugating protein ligase activity	6.29E-11	81	34
GO:0043687	post-translational protein modification	6.29E-11	81	34
GO:0051246	regulation of protein metabolic process	6.29E-11	81	34
GO:0005634	nucleus	6.29E-11	2042	305
GO:0008270	zinc ion binding	2.00E-10	2880	342
GO:0006355	regulation of transcription, DNA-dependent	8.18E-08	1527	235
GO:0004298	threonine-type endopeptidase activity	5.50E-07	39	18
GO:0005839	proteasome core complex	5.50E-07	40	18
GO:0051603	proteolysis involved in cellular protein catabolic process	5.50E-07	40	18
GO:0004175	endopeptidase activity	5.50E-07	36	17
GO:0000159	protein phosphatase type 2A complex	3.23E-06	34	19
GO:0008601	protein phosphatase type 2A regulator activity	3.23E-06	34	19
GO:0003677	DNA binding	6.09E-06	3081	333
GO:0042309	homoiothermy	2.37E-05	1403	178
GO:0050825	ice binding	2.37E-05	1403	178
GO:0050826	response to freezing	2.37E-05	1403	178
GO:0006357	regulation of transcription from RNA polymerase II promoter	2.45E-05	20	11
GO:0042254	ribosome biogenesis	4.82E-05	15	10
GO:0006886	intracellular protein transport	0.00021746	309	64
GO:0008299	isoprenoid biosynthetic process	0.00021746	41	16
GO:0003779	actin binding	0.00025202	109	28
GO:0005741	mitochondrial outer membrane	0.00034582	17	10
GO:0015031	protein transport	0.00059976	239	47
GO:0016592	mediator complex	0.00069731	16	8
GO:0005525	GTP binding	0.00099827	464	76
GO:0030163	protein catabolic process	0.00099827	28	14
GO:0019773	proteasome core complex, alpha-subunit complex	0.00107006	20	9
GO:0005198	structural molecule activity	0.00126829	352	56
GO:0005746	mitochondrial respiratory chain	0.00133542	5	4
GO:0015035	protein disulfide oxidoreductase activity	0.00133542	82	20
GO:0005875	microtubule associated complex	0.00133542	10	6
GO:0005515	protein binding	0.00135339	2087	315
GO:0003723	RNA binding	0.00146736	603	91
GO:0006511	ubiquitin-dependent protein catabolic process	0.00214939	138	34
GO:0008308	voltage-gated anion channel activity	0.00230267	14	8
GO:0043565	sequence-specific DNA binding	0.00322088	654	105
GO:0008121	ubiquinol-cytochrome-c reductase activity	0.00381259	17	8
GO:0015934	large ribosomal subunit	0.00415964	47	13
GO:0016021	integral to membrane	0.006998	1536	220
GO:0000786	nucleosome	0.00731524	212	34
GO:0007165	signal transduction	0.00906311	240	52
GO:0006334	nucleosome assembly	0.00967252	213	34
GO:0045454	cell redox homeostasis	0.01128127	169	32
GO:0005787	signal peptidase complex	0.01264828	5	4
GO:0006333	chromatin assembly or disassembly	0.01264828	41	8
GO:0016192	vesicle-mediated transport	0.0128852	167	36
GO:0006820	anion transport	0.0128852	23	10
GO:0046961	proton-transporting ATPase activity, rotational mechanism	0.01405845	51	13
GO:0030117	membrane coat	0.01469482	38	14
GO:0000502	proteasome complex	0.0212739	6	5
GO:0030234	enzyme regulator activity	0.0212739	6	5
GO:0042176	regulation of protein catabolic process	0.0212739	6	5
GO:0001503	ossification	0.02159245	19	8
GO:0005956	protein kinase CK2 complex	0.02212297	5	4
GO:0019887	protein kinase regulator activity	0.02212297	5	4
GO:0006626	protein targeting to mitochondrion	0.02212297	7	4
GO:0042719	mitochondrial intermembrane space protein transporter complex	0.02212297	7	4
GO:0045039	protein import into mitochondrial inner membrane	0.02212297	7	4
GO:0003924	GTPase activity	0.02305167	213	41
GO:0016597	amino acid binding	0.02407427	69	18
GO:0009289	pilus	0.02809824	10	6
GO:0003678	DNA helicase activity	0.02870415	5	4
GO:0003711	transcription elongation regulator activity	0.02932969	7	5
GO:0003700	sequence-specific DNA binding transcription factor activity	0.02932969	967	132
GO:0005801	cis-Golgi network	0.02932969	5	4

GO:0016757	transferase activity, transferring glycosyl groups	0.03128772	95	26
GO:0007264	small GTPase mediated signal transduction	0.03865744	167	30
GO:0004879	ligand-activated sequence-specific DNA binding RNA polymerase II transcription factor activity	0.0401309	356	50
GO:0016226	iron-sulfur cluster assembly	0.04376471	23	7
GO:0007010	cytoskeleton organization	0.04926769	37	9
GO:0016651	oxidoreductase activity, acting on NADH or NADPH	0.04926769	25	8
29	protein	1.37E-32	4242	795
29.2.1.2	protein.synthesis.ribosomal protein.eukaryotic	2.36E-26	325	98
29.2.1.2.2	protein.synthesis.ribosomal protein.eukaryotic.60S subunit	1.47E-21	182	65
29.2.1	protein.synthesis.ribosomal protein	1.64E-17	534	115
29.2	protein.synthesis	4.74E-16	717	147
27	RNA	1.13E-15	3961	658
29.5.11	protein.degradation.ubiquitin	3.27E-14	1186	241
27.3	RNA.regulation of transcription	2.32E-13	3208	534
29.3	protein.targeting	3.78E-07	344	80
29.5	protein.degradation	5.78E-07	1732	295
9	mitochondrial electron transport / ATP synthesis	1.52E-06	169	42
29.2.1.2.1	protein.synthesis.ribosomal protein.eukaryotic.40S subunit	2.84E-06	143	33
21	redox	8.23E-06	253	53
29.3.4	protein.targeting.secretory pathway	4.21E-05	168	42
29.5.11.4	protein.degradation.ubiquitin.E3	0.00054158	928	157
9.1	mitochondrial electron transport / ATP synthesis.NADH-DH	0.00191341	53	15
16.1.1	secondary metabolism.isoprenoids.non-mevalonate pathway	0.00253237	22	10
31	cell	0.00309425	1009	172
4.1	glycolysis.cytosolic branch	0.00653798	65	19
34.1	transport.p- and v-ATPases	0.01221757	32	10
PWY-5121	superpathway of geranylgeranyldiphosphate biosynthesis II (via MEP)	4.73E-05	54	22
PWY-3781	aerobic respiration -- electron donor II	5.50E-05	192	46
NONMEVIPP	methylerythritol phosphate pathway	0.00049182	23	12
PWY-5910	superpathway of geranylgeranyldiphosphate biosynthesis I (via mevalonate)	0.00975872	68	21
PWY-5670	epoxysqualene biosynthesis	0.04688252	4	4
PWY-1061	homogalacturonan biosynthesis	0.04688252	28	12
PF00179.21	UQ_con	5.40E-09	79	34
PF01423.17	LSM	1.48E-05	30	14
PF00227.21	Proteasome	0.00059432	39	16
PF03151.11	TPT	0.00065328	66	24
PF13639.1	zf-RING_2	0.00149238	329	62
PF14369.1	zf-RING_3	0.00149238	19	11
PF13658.1	UCH37_bd	0.00149238	8	7
PF04852.7	DUF640	0.00386779	20	10
PF01501.15	Glyco_transf_8	0.0049192	54	21
PF13920.1	zf-C3HC4_3	0.00628489	63	22
PF01849.13	NAC	0.00896866	8	6
PF04758.9	Ribosomal_S30	0.01319007	4	4
PF13837.1	Myb_DNA-bind_4	0.01319007	55	21
PF01459.17	Porin_3	0.01319007	16	8
PF00498.21	FHA	0.01482784	18	10
PF01529.15	zf-DHHC	0.01482784	39	16
PF13266.1	DUF4057	0.01869206	8	7
PF06886.6	TPX2	0.01997799	20	10
PF05002.10	SGS	0.01997799	4	4
PF08879.5	WRC	0.02353731	28	12
PF03248.8	Rer1	0.02353731	4	4
PF00804.20	Syntaxin	0.02353731	16	8
PF01080.12	Presenilin	0.02666528	4	4
PF04099.7	Sybindin	0.02684679	4	4
PF02681.9	DUF212	0.03309509	11	6
PF01221.13	Dynein_light	0.03309509	10	5
PF03798.11	TRAM_LAG1_CLN8	0.03309509	16	8
PF04548.11	AIG1	0.03309509	11	7
PF06547.7	DUF1117	0.03309509	4	4
PF02212.13	GED	0.03311413	13	8
PF13865.1	FoP_duplication	0.04244522	4	4
PF00293.23	NUDIX	0.04244522	31	11

B) R-NF-Foliar:

Term	Description	p-value	Total genes with annotation	R-NF-Foliar genes with annotation
GO:0005634	nucleus	1.74E-07	2042	52
GO:0006355	regulation of transcription, DNA-dependent	6.36E-06	1527	41
GO:0030001	metal ion transport	2.30E-05	156	12
GO:0030528	transcription regulator activity	0.00021628	481	19
GO:0015979	photosynthesis	0.00021628	91	8
GO:0000155	two-component sensor activity	0.0003655	23	6
GO:0046872	metal ion binding	0.00059859	314	14
GO:0005992	trehalose biosynthetic process	0.00088711	35	6
GO:0003700	sequence-specific DNA binding transcription factor activity	0.00295774	967	25
GO:0000156	two-component response regulator activity	0.00663776	55	6
GO:0043565	sequence-specific DNA binding	0.01139901	654	19
GO:0000160	two-component signal transduction system (phosphorelay)	0.01139901	64	6
GO:0048529	magnesium-protoporphyrin IX monomethyl ester (oxidative) cyclase activity	0.01536242	2	2
GO:0019419	sulfate reduction	0.02020626	2	2
GO:0003677	DNA binding	0.02020626	3081	43
GO:0004612	phosphoenolpyruvate carboxykinase (ATP) activity	0.02769627	3	2
1.1	PS.lightreaction	6.04E-07	221	15
1	PS	4.22E-06	318	17
27.3	RNA.regulation of transcription	4.22E-06	3208	71
14	S-assimilation	0.00013233	8	4
1.1.1	PS.lightreaction.photosystem II	0.00020257	76	7
3.2	minor CHO metabolism.trehalose	0.00020257	32	6
15	metal handling	0.00020257	67	7
27	RNA	0.00053356	3961	75
1.1.2	PS.lightreaction.photosystem I	0.00285128	32	4
10.1	cell wall.precursor synthesis	0.00445453	66	6
17	hormone metabolism	0.00445453	571	17
33	development	0.01531231	673	18
29.2.1	protein.synthesis.ribosomal protein	0.02813159	534	12
3	minor CHO metabolism	0.02813159	174	8
29.2.1.99.1	protein.synthesis.ribosomal protein.unknown.small subunit	0.02813159	9	2
17.4	hormone metabolism.cytokinin	0.03033561	40	4
29.2.1.1.1.2	protein.synthesis.ribosomal protein.prokaryotic.chloroplast.50S subunit	0.04586318	75	4
10.1.1	cell wall.precursor synthesis.NDP sugar pyrophosphorylase	0.04586318	6	2
SULFMETII-PWY	sulfate reduction II (assimilatory)	0.00271499	10	4
PF00403.21	HMA	2.26E-07	85	14
PF00512.20	HisKA	0.00054466	19	6
PF04873.8	EIN3	0.00162831	9	4
PF02518.21	HATPase_c	0.00162831	26	6
PF02358.11	Trehalose_PPase	0.00207575	29	6
PF03924.8	CHASE	0.00289861	8	4
PF00072.19	Response_reg	0.01565418	46	7
PF01241.13	PSI_PSAK	0.03375754	3	2
PF00924.13	MS_channel	0.03375754	7	3
PF02427.12	PSI_PsaE	0.03375754	2	2
PF13326.1	PSII_Pbs27	0.03545981	3	2
PF02915.12	Rubrerythrin	0.03545981	2	2
PF01698.11	FLO_LFY	0.03545981	2	2
PF00101.15	RuBisCO_small	0.03545981	2	2
PF12338.3	RbcS	0.03545981	2	2
PF00046.24	Homeobox	0.03981691	96	7
PF02309.11	AUX_IAA	0.03981691	69	6
PF00504.16	Chloroa_b-bind	0.03981691	31	4
PF07014.7	Hs1pro-1_C	0.03981691	2	2
PF07231.7	Hs1pro-1_N	0.03981691	2	2
PF03195.9	DUF260	0.04564464	44	4

C) R-NF-Husk:

Term	Description	p-values	Total genes with annotation	R-NF-Husk genes with annotation
GO:0006950	response to stress	0.031697274	88	6
GO:0004990	oxytocin receptor activity	0.031697274	11	3
GO:0005886	plasma membrane	0.031697274	14	3
GO:0004657	proline dehydrogenase activity	0.031697274	2	2
GO:0006562	proline catabolic process	0.031697274	2	2
GO:0003677	DNA binding	0.033498679	3081	32
PF03634.8	TCP	0.023002852	44	5

D) R-NF-Foliar&Husk:

Term	Description	p-values	Total genes with annotation	R-NF-Foliar&Husk genes with annotation
GO:0030528	transcription regulator activity	2.41E-06	481	14
GO:0006807	nitrogen compound metabolic process	0.007071845	39	4
GO:0004107	chorismate synthase activity	0.007071845	2	2
GO:0016788	hydrolase activity, acting on ester bonds	0.007071845	144	6
GO:0045181	glutamate synthase activity, NADH or NADPH as acceptor	0.043164999	3	2
12	N-metabolism	5.70E-06	38	6
12.2	N-metabolism.ammonia metabolism	5.52E-05	13	4
27.3	RNA.regulation of transcription	0.000524199	3208	30
34	transport	0.014216089	1364	16
27	RNA	0.014216089	3961	30
10.5	cell wall.cell wall proteins	0.014216089	84	4
34.2	transport.sugars	0.018095608	7	2
20.2	stress.abiotic	0.02587074	518	8
13.2.5	amino acid metabolism.degradation.serine-glycine-cysteine group	0.02587074	11	2
PWY-3282	ammonia assimilation cycle II	0.000101624	12	4
PWY-381	nitrate reduction II (assimilatory)	0.000101624	14	4
GLUGLNSYN-	glutamate biosynthesis IV	0.039985373	4	2
PF00010.21	HLH	4.68E-05	161	9
PF00657.17	Lipase_GDSL	0.012139834	116	6
PF01264.16	Chorismate_synt	0.024275215	2	2

Appendix 6: Ancestral profile inference by Bayesian model comparison

Steve Kelly formulated the following Bayesian model comparison to formalise the logic summarised in section 3.2.2.1 regarding the inference of maize ancestral gene expression profiles:

To determine whether the homeolog from subgenome-1 or subgenome-2 has been neofunctionalised it is necessary to know the ancestral gene expression profile. As this data is unobtainable, ancestral profiles must be inferred. Assuming that the expression profiles of non-neofunctionalised homeolog gene-pairs are an un-biased sample of the global distribution of gene expression profiles, the probability of observing an ancestral gene expression profile can be inferred using:

$$P(A_i) = \frac{N_i}{\sum_j N_j} \qquad [1]$$

where A_i is the ancestral gene expression profile in question and N_i is the number of times that the same profile is observed in all of the homeolog genes present in this analysis. The probability of observing an expression profile is inversely proportional to the number of changes between the observed and ancestral gene expression profile such that:

$$P(E_i \mid A_j) = \frac{1}{1 + d(i,j)} \qquad [2]$$

where $P(E_i|A_j)$ is the probability of observing expression profile E_i given ancestral profile A_j and $d(i,j)$ is the Hamming distance (number of changes) between the observed and ancestral profile. We can thus compute the posterior probability of each ancestral profile for any observed profile using Bayes rule and the law of total probability

$$P(A_j \mid E_i) = \frac{P(E_i \mid A_j)P(A_j)}{\sum_k P(E_i \mid A_k)P(A_k)} \qquad [3].$$

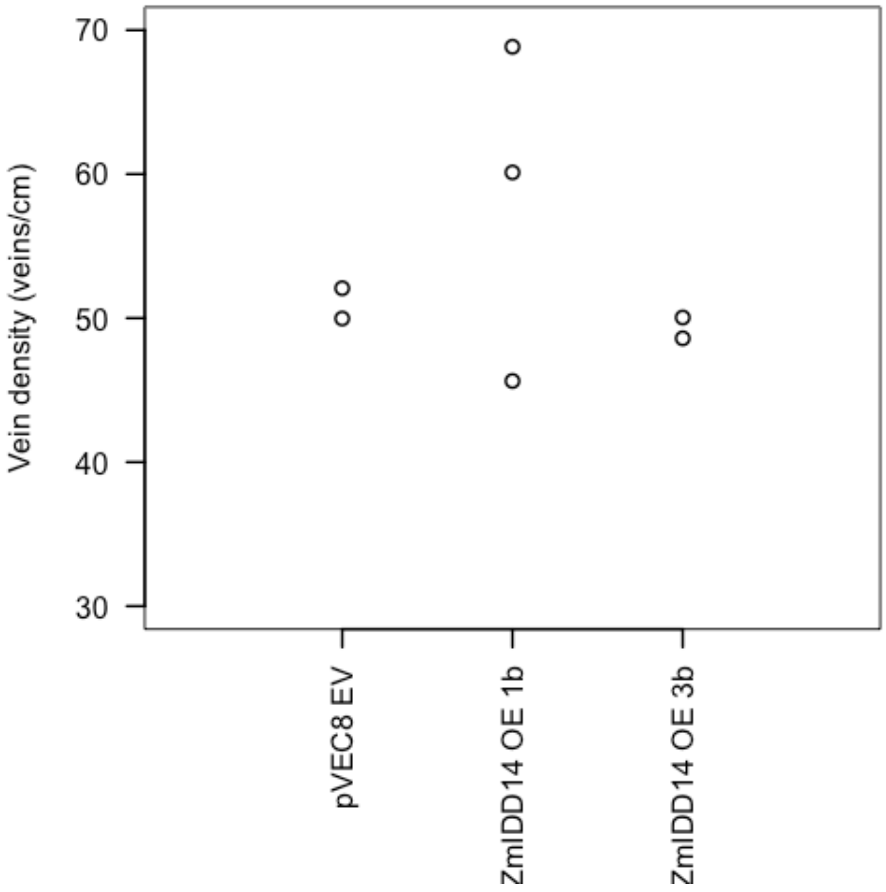
For the purposes of all downstream analyses we selected the maximum likelihood ancestral profile.

Appendix 7: Homeolog genes identified as being under positive selection

R-NF?	Homeolog under positive selection	Subgenome	KaKs	Homeolog-pair	Subgenome	KaKs
Non-R-NF	GRMZM2G005229	1	1.05643	GRMZM2G072690	2	0.0630727
Non-R-NF	GRMZM2G032244	1	1.05052	GRMZM2G063729	2	0.132259
Non-R-NF	GRMZM2G100333	2	1.01085	GRMZM2G469409	1	0.606234
Non-R-NF	GRMZM2G113053	1	1.36636	GRMZM5G816015	2	0.239668
Non-R-NF	GRMZM2G128902	1	1.08988	GRMZM2G138632	2	0.643236
Non-R-NF	GRMZM2G130854	2	1.0111	GRMZM2G398506	1	1.04133
Non-R-NF	GRMZM2G163159	2	1.00525	GRMZM2G027856	1	0.71873
Non-R-NF	GRMZM2G386088	2	1.00375	GRMZM2G332560	1	0.486829
Non-R-NF	GRMZM2G398506	1	1.04133	GRMZM2G130854	2	1.0111
R-NF (H)	GRMZM2G162781	2	1.0399	GRMZM2G081857	1	0.896319

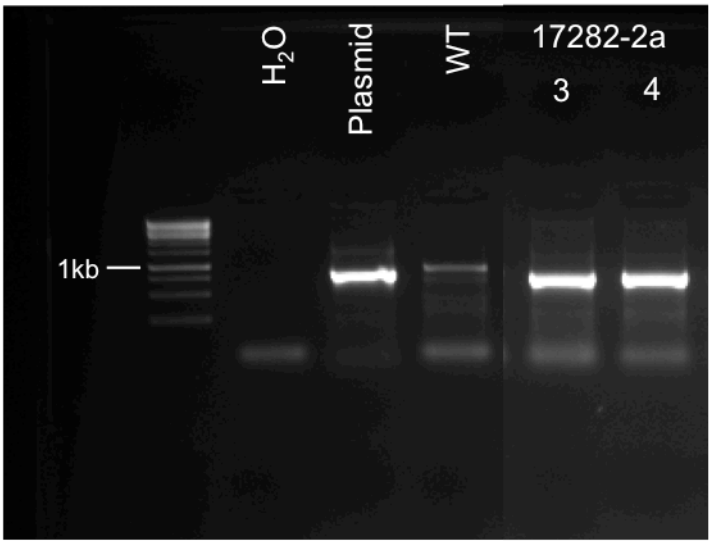
Appendix 9: Preliminary analysis of constitutive *ZmIDD14* expression in rice

pVEC8 EV = Empty vector control line.



Appendix 10: Representative genotyping of 17282 T1 rice lines

Rice lines transformed with construct 17282 (*ZmRVN1*$_{pro}$*::ZmRVN1 with 35S*$_{pro}$*::Hyg*^{R}) were PCR genotyped using primers that amplify from the *ZmRVN1* promoter to the *ZmRVN1* CDS. An example is shown below. A bright amplicon of the correct size (792bp) is observed in the plasmid control and in the two individual gDNA samples from line 2a, individuals 3 and 4. A fainter slightly larger band is observed in the Kitaake WT control. No amplicon is observed in the H_2O control. gDNA was extracted from leaf 6.



References

Adams KL, Wendel JF. 2013. Dynamics of Duplicated Gene Expression in Polyploid Cotton. In *Polyploid and Hybrid Genomics* (eds. Z.J. Chen and J.A. Birchler), John Wiley & Sons, Inc., Oxford, UK.

Adams KL, Wendel JF. 2005. Polyploidy and genome evolution in plants. *Curr Opin Plant Biol* **8**: 135–41.

Agarwal P, Arora R, Ray S, Singh AK, Singh VP, Takatsuji H, Kapoor S, Tyagi AK. 2007. Genome-wide identification of C2H2 zinc-finger gene family in rice and their phylogeny and expression analysis. *Plant Mol Biol* **65**: 467–85.

Aloni R, Schwalm K, Langhans M, Ullrich CI. 2003. Gradual shifts in sites of free-auxin production during leaf-primordium development and their role in vascular differentiation and leaf morphogenesis in *Arabidopsis*. *Planta* **216**: 841–53.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Arnozis PA, Nelemans JA, Findenegg GR. 1988. Phosphoenolpyruvate Carboxylase Activity in Plants Grown with either NO^-_3 or NH^+_4 as Inorganic Nitrogen Source. *J Plant Physiol* **132**: 23–27.

Bainbridge G, Madgwick P, Parmar S, Mitchell R, Paul M, Pitts J, Keys AJ, Parry MAJ. 1995. Engineering Rubisco to change its catalytic properties. *J Exp Bot* **46**: 1269–1276.

Bassham JA, Barker SA, Calvin M, Quarck UC. 1956. Intermediates in the Photosynthetic Cycle. *Biochim Biophys Acta* **21**: 376–377.

Battraw MJ, Hall TC. 1990. Histochemical analysis of CaMV 35S promoter-β-glucuronidase gene expression in transgenic rice plants. *Plant Mol Biol* **15**: 527–538.

Benfey P, Linstead P, Roberts K, Schiefelbein J, Hauser M, Aeschbacher R. 1993. Root development in *Arabidopsis*: four mutants with dramatically altered root morphogenesis. *Development* **119**: 57–70.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Mulitple Testing. *J R Stat Soc* **57**: 289–300.

Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J, et al. 2012. Reference genome sequence of the model plant Setaria. *Nat Biotechnol* **30**: 555–561.

Bisalputra T, Downton WJS, Tregunna EB. 1969. The distribution and ultrastructure of chloroplasts in leaves differing in photosynthetic carbon metabolism. I. Wheat, Sorghum , and Aristida (Gramineae). *Can J Bot* **47**: 15–21.

Blanc G, Wolfe KH. 2004a. Functional Divergence of Duplicated Genes Formed by Polyploidy during Arabidopsis Evolution. *Plant Cell* **16**: 1679–1691.

Blanc G, Wolfe KH. 2004b. Widespread Paleopolyploidy in Model Plant Species Inferred from Age Distributions of Duplicate Genes. *Plant Cell* **16**: 1667–1678.

Bosabalidis AM, Evert RF, Russin WA. 1994. Ontogeny of the Vascular Bundles and Contiguous Tissues in the Maize Leaf Blade. *Am J Bot* **81**: 745–752.

Bowes G. 2011. Single-cell C_4 photosynthesis in aquatic plants. In *C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms* (eds. A. Raghavendra and R. Sage), pp. 63–80, Springer Netherlands, Dordrecht.

Bowes G, Ogren WL. 1972. Oxygen Inhibition and Other Properties of Soybean Ribulose 1,5-Diphosphate Carboxylase. *J Biol Chem* **247**: 2171–2176.

Bowes G, Salvucci ME. 1984. Hydrilla: Inducible C_4-type Photosynthesis without Kranz Anatomy. In *Advances in Photosynthesis Research*, pp. 829–832, Springer Netherlands, Dordrecht.

Brandt R, Salla-Martret M, Bou-Torrent J, Musielak T, Stahl M, Lanz C, Ott F, Schmid M, Greb T, Schwarz M, et al. 2012. Genome-wide binding-site analysis of REVOLUTA reveals a link between leaf patterning and light-mediated growth responses. *Plant J* **72**: 31–42.

Brooks A, Farquhar GD. 1985. Effect of temperature on the CO_2/O_2 specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase and the rate of respiration in the light. *Planta* **165**: 397–406.

Brown NJ, Newell CA, Stanley S, Chen JE, Perrin AJ, Kajala K, Hibberd JM, Bowes G, Ogren WL, Hageman RH, et al. 2011. Independent and parallel recruitment of preexisting mechanisms underlying C_4 photosynthesis. *Science* **331**: 1436–1439.

Brown NJ, Parsley K, Hibberd JM. 2005. The future of C_4 research- maize, *Flaveria* or *Cleome*? *Trends Plant Sci* **10**: 215–221.

Brown R. 1999. Agronomic implications of C_4 photosynthesis. In *C_4 Plant Biology* (eds. R. Sage and R. Monson), pp. 473–507, Academic Press, San Diego.

Brown RH. 1978. A Difference in N Use Efficiency in C_3 and C_4 Plants and its Implications in Adaptation and Evolution. *Crop Sci* **18**: 93–98.

Brown W. 1960. A cytological difference between the Eupanicoideae and the Chloridoideae (Gramineae). *Southwest Nat* **5**: 7–11.

Brown WV. 1958. Leaf Anatomy in Grass Systematics. *Bot Gaz* **119**: 170–178.

Brutnell TP, Wang L, Swartwood K, Goldschmidt A, Jackson D, Zhu X-G, Kellogg E, Van Eck J. 2010. *Setaria viridis*: a model for C_4 photosynthesis. *Plant Cell* **22**: 2537–2544.

Buggs RJA, Chamala S, Wu W, Gao L, May GD, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB. 2010. Characterization of duplicate gene evolution in the recent natural allopolyploid Tragopogon miscellus by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. *Mol Ecol* **19**: 132–146.

von Caemmerer S, Furbank R. 2003. The C_4 pathway: an efficient CO_2 pump. *Photosynth Res* **77**: 191–207.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–3.

Chang Y-M, Liu W-Y, Shih AC-C, Shen M-N, Lu C-H, Lu M-YJ, Yang H-W, Wang T-Y, Chen SC-C, Chen SM, et al. 2012. Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. *Plant Physiol* **160**: 165–177.

Cheng F, Sun C, Wu J, Schnable J, Woodhouse MR, Liang J, Cai C, Freeling M, Wang X. 2016. Epigenetic regulation of subgenome dominance following whole genome triplication in *Brassica rapa*. *New Phytol* **211**: 288–299.

Cheng F, Wu J, Fang L, Sun S, Liu B, Lin K, Bonnema G, Wang X. 2012. Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLoS One* **7**: e36442.

Christin P-A, Besnard G. 2009. Two independent C_4 origins in Aristidoideae (Poaceae) revealed by the recruitment of distinct phosphoenolpyruvate carboxylase genes. *Am J Bot* **96**: 2234–9.

Christin P-A, Besnard G, Samaritani E, Duvall MR, Hodkinson TR, Savolainen V, Salamin N. 2008. Oligocene CO_2 Decline Promoted C_4 Photosynthesis in Grasses. *Curr Biol* **18**: 37–43.

Christin P-A, Osborne CP. 2014. The evolutionary ecology of C_4 plants. *New Phytol* **204**: 765–781.

Christin P-A, Osborne CP, Chatelet DS, Columbus JT, Besnard G, Hodkinson TR, Garrison LM, Vorontsova MS, Edwards EJ. 2013. Anatomical enablers and the evolution of C_4 photosynthesis in grasses. *Proc Natl Acad Sci U S A* **110**: 1381–1386.

Christin P-A, Sage TL, Edwards EJ, Ogburn RM, Khoshravesh R, Sage RF. 2011. Complex evolutionary transitions and the significance of C_3-C_4 intermediate forms of photosynthesis in Molluginaceae. *Evolution (N Y)* **65**: 643–660.

Ciftci-Yilmaz S, Mittler R. 2008. The zinc finger network of plants. *Cell Mol Life Sci* **65**: 1150–1160.

Clark LH, Harris WH. 1981. Observations on the Root Anatomy of Rice (*Oryza sativa* L.). *Am J Bot* **68**: 154.

Colasanti J, Tremblay R, Wong AYM, Coneva V, Kozaki A, Mable BK. 2006. The maize *INDETERMINATE1* flowering time regulator defines a highly conserved zinc finger protein family in higher plants. *BMC Genomics* **7**: 158.

Collingridge PW, Kelly S. 2012. MergeAlign: improving multiple sequence alignment performance by dynamic reconstruction of consensus multiple sequence alignments. *BMC Bioinformatics* **13**: 117.

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet* **9**: 938–50.

Cornic G, Briantais J-M. 1991. Partitioning of photosynthetic electron flow between CO_2 and O_2 reduction in a C_3 leaf (Phaseolus vulgaris L.) at different CO_2 concentrations and during drought stress. *Planta* **183**: 178–184.

Craft J, Samalova M, Baroux C, Townley H, Martinez A, Jepson I, Tsiantis M, Moore I. 2005. New pOp/LhG4 vectors for stringent glucocorticoid-dependent transgene expression in Arabidopsis. *Plant J* **41**: 899–918.

Crookston RK, Moss DN. 1973. A Variation of C_4 Leaf Anatomy in Arundinella hirta (Gramineae). *Plant Physiol* **52**: 397–402.

Cui D, Zhao J, Jing Y, Fan M, Liu J, Wang Z, Xin W, Hu Y, Zhao Y, Petrášek J, et al. 2013. The Arabidopsis IDD14, IDD15, and IDD16 Cooperatively Regulate Lateral Organ Morphogenesis and Gravitropism by Promoting Auxin Biosynthesis and Transport ed. H. Yu. *PLoS Genet* **9**: e1003759.

Cui H, Hao Y, Kong D. 2012. SCARECROW has a SHORT-ROOT-independent role in modulating the sugar response. *Plant Physiol* **158**: 1769–78.

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*. *Plant J* **78**: 319–27.

Cui H, Levesque MP, Vernoux T, Jung JW, Paquette AJ, Gallagher KL, Wang JY, Blilou I, Scheres B, Benfey PN. 2007. An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. *Science* **316**: 421–5.

D'Hont A, Denoeud F, Aury J-M, Baurens F-C, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M, et al. 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* **488**: 213–217.

Dannenhoffer J, Jr WE, Evert R. 1990. Leaf vasculature in barley, *Hordeum vulgare* (Poaceae). *Am J Bot* **77**: 636–652.

Dengler NG, Dengler RE, Hattersley PW. 1985. Differing Ontogenetic Origins of PCR ("Kranz") Sheaths in Leaf Blades of C4 Grasses (Poaceae). *Am J Bot* **72**: 284–302.

Deyholos M, Cordner G, Beebe D, Sieburth L. 2000. The *SCARFACE* gene is required for cotyledon and leaf vein patterning. *Development* **127**: 3205–3213.

Dhondt S, Coppens F, De Winter F, Swarup K, Merks RMH, Inzé D, Bennett MJ, Beemster GTS. 2010. SHORT-ROOT and SCARECROW regulate leaf growth in Arabidopsis by stimulating S-phase progression of the cell cycle. *Plant Physiol* **154**: 1183–95.

Dolan L, Janmaat K, Willemsen V, Linstead P, Poethig S, Roberts K, Scheres B. 1993. Cellular organisation of the *Arabidopsis thaliana* root. *Development* **119**: 71–84.

Drincovich M, Lara M. 2011. C4 decarboxylases: Different solutions for the same biochemical problem, the provision of CO_2 to RuBisCO in the bundle sheath cells. In *C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms*, pp. 277–300.

Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, Ma H, Altman N, dePamphilis CW. 2006. Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis. *Mol Biol Evol* **23**: 469–78.

Duvall M, Davis J, Clark L, Noll J, Goldman D, Sánchez-Ken G. 2007. Phylogeny of the Grasses (Poaceae) Revisited. *Aliso* **23**: 237–247.

Edwards EJ, Ogburn RM. 2012. Angiosperm Responses to a Low-CO_2 World: CAM and C_4 Photosynthesis as Parallel Evolutionary Trajectories. *Int J Plant Sci* **173**: 724–733.

Edwards EJ, Smith SA. 2010. Phylogenetic analyses reveal the shady history of C_4 grasses. *Proc Natl Acad Sci* **107**: 2532–2537.

Edwards GE, Franceschi VR, Voznesenskaya E V. 2004. Single-cell C_4 photosynthesis versus the dual-cell (Kranz) paradigm. *Annu Rev Plant Biol* **55**: 173–96.

Edwards GE, 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. A.S. Raghavendra and R.F. Sage), Vol. 32 of *Advances in Photosynthesis and Respiration*, pp. 29–61, Springer Netherlands, Dordrecht.

Ehleringer JR, Sage RF, Flanagan LB, Pearcy RW. 1991. Climate change and the evolution of C_4 photosynthesis. *Trends Ecol Evol* **6**: 95–99.

Emery JF, Floyd SK, Alvarez J, Eshed Y, Hawker NP, Izhaki A, Baum SF, Bowman JL. 2003. Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. *Curr Biol* **13**: 1768–1774.

Emms DM, Covshoff S, Hibberd JM, 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. *Mol Biol Evol* **33**: 1796–1806.

Englbrecht CC, Schoof H, Böhm S. 2004. Conservation, diversification and expansion of C2H2 zinc finger proteins in the Arabidopsis thaliana genome. *BMC Genomics* **5**: 39.

Erdmann R, Gramzow L, Melzer R, Theissen G, Becker A. 2010. *GORDITA (AGL63)* is a young paralog of the *Arabidopsis thaliana* B_{sister} MADS box gene *ABS (TT16)* that has undergone neofunctionalization. *Plant J* **63**: 914–24.

Erlinghaeuser M, Hagenau L, Wimmer D, Offermann S. 2016. Development, subcellular positioning and selective protein accumulation in the dimorphic chloroplasts of single-cell C_4 species. *Curr Opin Plant Biol* **31**: 76–82.

Esau K. 1943. Ontogeny of the Vascular Bundle in *Zea Mays*. *Hilgardia* **15**: 325–368.

Esau K. 1965. *Plant Anatomy*. John Wiley & Sons, Inc., New York.

FAO. 2015. *FAO Statistical Pocketbook World Food and Agriculture 2015*. Rome.

Feldman AB, Murchie EH, Leung H, Baraoidan M, Coe R, Yu S-M, Lo S-F, Quick WP, Sage T, Sage R, et al. 2014. Increasing Leaf Vein Density by Mutagenesis: Laying the Foundations for C_4 Rice ed. K. Wu. *PLoS One* **9**: e94947.

Feurtado JA, Huang D, Wicki-Stordeur L, Hemstock LE, Potentier MS, Tsang EWT, Cutler AJ. 2011. The Arabidopsis C2H2 Zinc Finger INDETERMINATE DOMAIN1/ENHYDROUS Promotes the Transition to Germination by Regulating Light and Hormonal Signaling during Seed Maturation. *Plant Cell* **23**: 1772–1794.

Fladung M. 1994. Genetic Variants of *Panicum maximum* (Jacq.) in C_4 Photosynthetic Traits. *J Plant Physiol* **143**: 165–172.

Flagel LE, Wendel JF. 2010. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytol* **186**: 184–93.

Flagel LE, Wendel JF. 2009. Gene duplication and evolutionary novelty in plants. *New Phytol* **183**: 557–64.

Fouracre JP, Ando S, Langdale JA. 2014. Cracking the Kranz enigma with systems biology. *J Exp Bot* **65**: 3327–39.

Freeling M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu Rev Plant Biol* **60**: 433–53.

Freeling M. 2008. The evolutionary position of subfunctionalization, downgraded. *Genome Dyn* **4**: 25–40.

Freeling M, Scanlon MJ, Fowler JE. 2015. Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. *Curr Opin Genet Dev* **35**: 110–118.

Freitag H, Stichler W. 2000. A Remarkable New Leaf Type With Unusual Photosynthetic Tissue in a Central Asiatic Genus of Chenopodiaceae. *Plant Biol* **2**: 154–160.

Furbank RT. 2011. Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? *J Exp Bot* **62**: 3103–8.

Furbank RT. 2016. Walking the C_4 pathway: past, present, and future. *J Exp Bot* **67**: 4057–4066.

Gallagher KL, Benfey PN. 2009. Both the conserved GRAS domain and nuclear localization are required for SHORT-ROOT movement. *Plant J* **57**: 785–97.

Garsmeur O, Schnable JC, Almeida A, Jourda C, D'Hont A, Freeling M. 2014. Two Evolutionarily Distinct Classes of Paleopolyploidy. *Mol Biol Evol* **31**: 448–454.

Gaut BS, Doebley JF. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc Natl Acad Sci* **94**: 6809–14.

Gout J-F, Lynch M. 2015. Maintenance and Loss of Duplicated Genes by Dosage Subfunctionalization. *Mol Biol Evol* **32**: 2141–2148.

Grass Phylogeny Working Group II. 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytol* **193**: 304–12.

Griffiths H, Weller G, Toy LFM, Dennis RJ. 2013. You're so vein: bundle sheath physiology, phylogeny and evolution in C_3 and C_4 plants. *Plant Cell Environ* **36**: 249–61.

Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF. 2012. Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol* **196**: 966–971.

Guo H, Lee T-H, Wang X, Paterson AH. 2013. Function Relaxation Followed by Diversifying Selection after Whole Genome Duplication in Flowering Plants. *Plant Physiol* **162**: 769–78.

Gutierrez M, Gracen VE, Edwards GE. 1974. Biochemical and cytological relationships in C_4 plants. *Planta* **119**: 279–300.

Haberlandt G. 1896. *Physiologische Pflanzenanatomie*. ed. E. Wilhelm. Leipzig.

Hall LN. 1998. GOLDEN 2: A Novel Transcriptional Regulator of Cellular Differentiation in the Maize Leaf. *Plant Cell* **10**: 925–936.

Hall LN, Langdale JA. 1996. Molecular genetics of cellular differentiation in leaves. *New Phytol* **132**: 533–553.

Harikrishnan L, Pucholt P, Berlin S. 2015. Sequence and gene expression evolution of paralogous genes in willows. *Nature Scientific Reports* **5:** 18662. doi: 10.1038/srep18662

Hatch M, Slack C. 1966. Photosynthesis by sugar-cane leaves: A new carboxylation reaction and the pathway of sugar formation. *Biochem J* **101**: 103–111.

Hatch MD. 2002. C_4 photosynthesis: discovery and resolution. *Photosynth Res* **73**: 251–6.

Hattersley PW, Browning AJ. 1981. Occurrence of the suberized lamella in leaves of grasses of different photosynthetic types. I. In parenchymatous bundle sheaths and PCR ("Kranz") sheaths. *Protoplasma* **109**: 371–401.

Hattersley PW, Wong S-C, Perry S, Roksandic Z. 1986. Comparative ultrastructure and gas exchange characteristics of the C_3-C_4 intermediate Neurachne minor S. T. Blake (Poaceae). *Plant Cell Environ* **9**: 217–233.

Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber APM, Lercher MJ. 2013. Predicting C_4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell* **153**: 1579–88.

Helariutta Y, Fukaki H, Wysocka-Diller J, Nakajima K, Jung J, Sena G, Hauser MT, Benfey PN. 2000. The SHORT-ROOT gene controls radial patterning of the Arabidopsis root through radial signaling. *Cell* **101**: 555–67.

Hibberd JM, Quick WP. 2002. Characteristics of C_4 photosynthesis in stems and petioles of C_3 flowering plants. *Nature* **415**: 451–4.

Hibberd JM, Sheehy JE, Langdale JA. 2008. Using C_4 photosynthesis to increase the yield of rice-rationale and feasibility. *Curr Opin Plant Biol* **11**: 228–31.

Hoffmann RD, Palmgren M, Blanc G, Hokamp K, Wolfe K, Friedman R, Hughes A, McLysaght A, Hokamp K, Wolfe K, et al. 2016. Purifying selection acts on coding and non-coding sequences of paralogous genes in *Arabidopsis thaliana*. *BMC Genomics* **17**: 456.

Hughes TE, Langdale JA, Kelly S. 2014. The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. *Genome Res* **24**: 1348–55.

Iltis HH. 2000. Homeotic sexual translocations and the origin of maize (*Zea Mays*, Poaceae): A new look at an old problem. *Econ Bot* **54**: 7–42.

Ingvarsson PK. 2006. Gene Expression and Protein Length Influence Codon Usage and Rates of Sequence Evolution in *Populus tremula*. *Mol Biol Evol* **24**: 836–844.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* **11**: 97–108.

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

Ishida Y, Hiei Y, Komari T. 2007. Agrobacterium-mediated transformation of maize. *Nat Protoc* **2**: 1614–1621.

Jacobs W. 1952. The role of auxin in differentiation of xylem around a wound. *Am J Bot* **5**: 301–309.

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. *Plant Cell* **28**: 1551–62.

Jiao Y, Li J, Tang H, Paterson AH. 2014. Integrated Syntenic and Phylogenomic Analyses Reveal an Ancient Genome Duplication in Monocots. *Plant Cell* **26**: 2792–2802.

Jordan DB, Ogren WL. 1981. Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase. *Nature* **291**: 513–515.

Jordan DB, Ogren WL. 1984. The CO_2/O_2 specificity of ribulose 1,5-bisphosphate carboxylase/oxygenase. *Planta* **161**: 308–313.

Kadereit G, Ackerly D, Pirie MD, Bauwe H, Pagani M, Zachos JC, Freeman KH, Tipple B, Bohaty S, Beerling DJ, et al. 2012. A broader model for C_4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae s.s.). *Proc R Soc Biol Sci* **279**: 3304–11.

Kamiya N, Itoh J-I, Morikami A, Nagato Y, Matsuoka M. 2003. The *SCARECROW* gene's role in asymmetric cell divisions in rice plants. *Plant J* **36**: 45–54.

Kaplan A, Reinhold L. 1999. CO_2 concentrating mechanisms in photosynthetic microorganisms. *Annu Rev Plant Physiol Plant Mol Biol* **50**: 539–570.

Kapralov M V, Kubien DS, Andersson I, Filatov DA. 2011. Changes in Rubisco kinetics during the evolution of C_4 photosynthesis in Flaveria (Asteraceae) are associated with positive selection on genes encoding the enzyme. *Mol Biol Evol* **28**: 1491–503.

Karpilov Y. 1960. The distribution of radioactive carbon 14 amongst the products of photosynthesis of maize. *Tr Kazan Sel'shokoz Inst* **41**: 15–24.

Kebeish R, Niessen M, Thiruveedhi K, Bari R, Hirsch H-J, Rosenkranz R, Stäbler N, Schönfeld B, Kreuzaler F, Peterhänsel C. 2007. Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in *Arabidopsis thaliana*. *Nat Biotechnol* **25**: 593–599.

Kim SL, Choi M, Jung K-H, An G. 2013. Analysis of the early-flowering mechanisms and generation of T-DNA tagging lines in Kitaake, a model rice cultivar. *J Exp Bot* **64**: 4169–82.

Kirchanski SJ. 1975. The Ultrastructural Development of the Dimorphic Plastids of *Zea mays* L. *Am J Bot* **62**: 695–705.

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.

Kondrashov FA, Rogozin IB, Wolf YI, Koonin E V. 2002. Selection in the evolution of gene duplications. *Genome Biol* **3**: RESEARCH0008.

Koroleva OA, Tomos AD, Farrar J, Roberts P, Pollock CJ. 2000. Tissue distribution of primary metabolism between epidermal, mesophyll and parenchymatous bundle sheath cells in barley leaves. *Funct Plant Biol* **27**: 747.

Kortschak HP, Hartt CE, Burr GO. 1965. Carbon Dioxide Fixation in Sugarcane Leaves. *Plant Physiol* **40**: 209–13.

Kozaki A, Hake S, Colasanti J. 2004. The maize ID1 flowering time regulator is a zinc finger protein with novel DNA binding properties. *Nucleic Acids Res* **32**: 1710–20.

Krecek P, Skupa P, Libus J, Naramoto S, Tejos R, Friml J, Zazímalová E. 2009. The PIN-FORMED (PIN) protein family of auxin transporters. *Genome Biol* **10**: 249.

Ku SB, Edwards GE. 1977a. Oxygen Inhibition of Photosynthesis: I. Temperature Dependence and Relation to O_2/CO_2 Solubility Ratio. *Plant Physiol* **59**: 986–90.

Ku SB, Edwards GE. 1977b. Oxygen Inhibition of Photosynthesis: II. Kinetic Characteristics as Affected by Temperature. *Plant Physiol* **59**: 991–9.

Laetsch W, Price I. 1969. Development of the dimorphic chloroplasts of sugar cane. *Am J Bot* **56**: 77–87.

Langdale J, Kidner C. 1994. Bundle sheath defective, a mutation that disrupts cellular differentiation in maize leaves. *Development* **120**: 673–681.

Langdale JA. 2011. C_4 cycles: past, present, and future research on C_4 photosynthesis. *Plant Cell* **23**: 3879–92.

Langdale JA, Lane B, Freeling M, Nelson T. 1989. Cell lineage analysis of maize bundle sheath and mesophyll cells. *Dev Biol* **133**: 128–39.

Langdale JA, Metzler MC, Nelson T. 1987. The argentia mutation delays normal development of photosynthetic cell-types in *Zea mays*. *Dev Biol* **122**: 243–255.

Langdale JA, Nelson T. 1991. Spatial regulation of photosynthetic development in C_4 plants. *Trends Genet* **7**: 191–196.

Langdale JA, Zelitch I, Miller E, Nelson T. 1988. Cell position and light influence C_4 versus C_3 patterns of photosynthetic gene expression in maize. *EMBO J* **7**: 3643–51.

Laurenzio L Di, Wysocka-Diller J, Malamy J. 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.

Lawrence CJ, Dong Q, Polacco ML, Seigfried TE, Brendel V. 2004. MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res* **32**: D393–7.

Lawrence CJ, Harper LC, Schaeffer ML, Sen TZ, Seigfried TE, Campbell DA. 2008. MaizeGDB: The maize model organism database for basic, translational, and applied research. *Int J Plant Genomics* **2008**: 496957.

Lee M-H, Kim B, Song S-K, Heo J-O, Yu N-I, Lee SA, Kim M, Kim DG, Sohn SO, Lim CE, et al. 2008. Large-scale analysis of the GRAS gene family in *Arabidopsis thaliana*. *Plant Mol Biol* **67**: 659–70.

Levesque MP, Vernoux T, Busch W, Cui H, Wang JY, Blilou I, Hassan H, Nakajima K, Matsumoto N, Lohmann JU, et al. 2006. Whole-genome analysis of the SHORT-ROOT developmental pathway in Arabidopsis. ed. D. Preuss. *PLoS Biol* **4**: e143.

Li A, Liu D, Wu J, Zhao X, Hao M, Geng S, Yan J, Jiang X, Zhang L, Wu J, et al. 2014. mRNA and Small RNA Transcriptomes Reveal Insights into Dynamic

Homoeolog Regulation of Allopolyploid Heterosis in Nascent Hexaploid Wheat. *Plant Cell* **26**: 1878–1900.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.

Li L, Wang X, Stolc V, Li X, Zhang D, Su N, Tongprasit W, Li S, Cheng Z, Wang J, et al. 2006. Genome-wide transcription analyses in rice using tiling microarrays. *Nat Genet* **38**: 124–9.

Li P, Brutnell TP. 2011. *Setaria* viridis and *Setaria italica*, model genetic systems for the Panicoid grasses. *J Exp Bot* **62**: 3031–7.

Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, et al. 2010. The developmental dynamics of the maize leaf transcriptome. *Nat Genet* **42**: 1060–7.

Li L, Briskine R, Schaefer R, Schnable PS, Myers CL, Flagel LE, Springer NM, Muehlbauer GJ. 2016. Co-expression network analysis of duplicate genes in maize (Zea mays L.) reveals no subgenome bias. *BMC Genomics* **17:**875 doi: 10.1186/s12864-016-3194-0

Lim J, Helariutta Y, Specht CD, Jung J, Sims L, Bruce WB, Diehn S, Benfey PN. 2000. Molecular analysis of the SCARECROW gene in maize reveals a common basis for radial patterning in diverse meristems. *Plant Cell* **12**: 1307–18.

Lim J, Jung JW, Lim CE, Lee M-H, Kim BJ, Kim M, Bruce WB, Benfey PN. 2005. Conservation and diversification of SCARECROW in maize. *Plant Mol Biol* **59**: 619–30.

Lin MT, Occhialini A, Andralojc PJ, Parry MAJ, Hanson MR. 2014. A faster Rubisco with potential to increase photosynthesis in crops. *Nature* **513**: 547–550.

Liu S-L, Baute GJ, Adams KL. 2011. Organ and cell type-specific complementary expression patterns and regulatory neofunctionalization between duplicated genes in *Arabidopsis thaliana*. *Genome Biol Evol* **3**: 1419–36.

Liu W-Y, Chang Y-M, Chen SC-C, Lu C-H, Wu Y-H, Lu M-YJ, Chen D-R, Shih AC-C, Sheue C-R, Huang H-C, et al. 2013. Anatomical and transcriptional dynamics of maize embryonic leaves during seed germination. *Proc Natl Acad Sci* **110**: 3979–84.

Lloyd J, Farquhar GD. 1994. ^{13}C discrimination during CO_2 assimilation by the terrestrial biosphere. *Oecologia* **99**: 201–215.

Long SP, Zhu X-G, Naidu SL, Ort DR. 2006. Can improvement in photosynthesis increase crop yields? *Plant Cell Environ* **29**: 315–30.

Long Y, Goedhart J, Schneijderberg M, Terpstra I, Shimotohno A, Bouchet BP, Akhmanova A, Gadella TWJ, Heidstra R, Scheres B, et al. 2015a. SCARECROW-LIKE23 and SCARECROW jointly specify endodermal cell fate but distinctly control SHORT-ROOT movement. *Plant J* **84**: 773–784.

Long Y, Smet W, Cruz-Ramírez A, Castelijns B, de Jonge W, Mähönen AP, Bouchet BP, Perez GS, Akhmanova A, Scheres B, et al. 2015b. Arabidopsis BIRD Zinc Finger Proteins Jointly Stabilize Tissue Boundaries by Confining the Cell Fate Regulator SHORT-ROOT and Contributing to Fate Specification. *Plant Cell* **27**: 1185–99.

Lorimer GH. 1981. The Carboxylation and Oxygenation of Ribulose 1,5-Bisphosphate: The Primary Events in Photosynthesis and Photorespiration. *Annu Rev Plant Physiol* **32**: 349–382.

Lorimer GH, Andrews TJ. 1973. Plant Photorespiration—An Inevitable Consequence of the Existence of Atmospheric Oxygen. *Nature* **243**: 359–360.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–5.

Maier A, Fahnenstich H, von Caemmerer S, Engqvist MKM, Weber APM, Flügge U-I, Maurino VG. 2012. Transgenic Introduction of a Glycolate Oxidative Cycle into *A. thaliana* Chloroplasts Leads to Growth Improvement. *Front Plant Sci* **3**: 38.

Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U, Anderson L, Aubry S, Brown N, et al. 2014. The role of photorespiration during the evolution of C_4 photosynthesis in the genus Flaveria. *Elife* **3**: e02478.

Marshall DM, Muhaidat R, Brown NJ, Liu Z, Stanley S, Griffiths H, Sage RF, Hibberd JM. 2007. Cleome, a genus closely related to Arabidopsis, contains species

spanning a developmental progression from C_3 to C_4 photosynthesis. *Plant J* **51**: 886–96.

Martins PK, Ribeiro AP, Cunha BADB da, Kobayashi AK, Molinari HBC. 2015. A simple and highly efficient *Agrobacterium*-mediated transformation protocol for *Setaria viridis*. *Biotechnol Reports* **6**: 41–44.

Matsuoka M, Kyozuka J, Shimamoto K, Kano-Murakami Y. 1994. The promoters of two carboxylases in a C_4 plant (maize) direct cell-specific, light-regulated expression in a C_3 plant (rice). *Plant J* **6**: 311–319.

Mattsson J, Ckurshumova W, Berleth T. 2003. Auxin signaling in Arabidopsis leaf vascular development. *Plant Physiol* **131**: 1327–39.

Mayfield S, Taylor W. 1984. The appearance of photosynthetic proteins in developing maize leaves. *Planta* **161**: 481–486.

McCarty DR, Meeley RB. 2009. Transposon Resources for Forward and Reverse Genetics in Maize. In *Handbook of Maize*, pp. 561–584, Springer New York, New York, NY.

McGrath C, Lynch M. 2012. Evolutionary significance of whole-genome duplication. In *Polyploidy and genome evolution* (eds. P.S. Soltis and D.E. Soltis), pp. 1–20, Springer Berlin Heidelberg, Berlin, Heidelberg.

McKain MR, Tang H, McNeal JR, Ayyampalayam S, Davis JI, dePamphilis CW, Givnish TJ, Pires JC, Stevenson DW, Leebens-Mack JH. 2016. A Phylogenomic Assessment of Ancient Polyploidy and Genome Evolution across the Poales. *Genome Biol Evol* **8**: 1150–64.

McKown AD, Dengler NG. 2007. Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in Flaveria (Asteraceae). *Am J Bot* **94**: 382–99.

McKown AD, Moncalvo J-M, Dengler NG. 2005. Phylogeny of Flaveria (Asteraceae) and inference of C_4 photosynthesis evolution. *Am J Bot* **92**: 1911–28.

Miller J, McLachlan AD, Klug A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J* **4**: 1609–14.

Miyao M, Masumoto C, Miyazawa S-I, Fukayama H. 2011. Lessons from engineering a single-cell C_4 photosynthetic pathway into rice. *J Exp Bot* **62**: 3021–9.

Monson RK. 2003. Gene Duplication, Neofunctionalization, and the Evolution of C_4 Photosynthesis. *Int J Plant Sci* **164**: S43–S54.

Moore RC, Purugganan MD. 2005. The evolutionary dynamics of plant duplicate genes. *Curr Opin Plant Biol* **8**: 122–8.

Morita MT, Sakaguchi K, Kiyose S, Taira K, Kato T, Nakamura M, Tasaka M. 2006. A C2H2-type zinc finger protein, SGR5, is involved in early events of gravitropism in Arabidopsis inflorescence stems. *Plant J* **47**: 619–628.

Muhaidat R, Sage TL, Frohlich MW, Dengler NG, Sage RF. 2011. Characterization of C_3-C_4 intermediate species in the genus Heliotropium L. (Boraginaceae): anatomy, ultrastructure and enzyme activity. *Plant Cell Environ* **34**: 1723–1736.

Nagasawa N, Miyoshi M, Sano Y, Satoh H, Hirano H, Sakai H, Nagato Y. 2003. *SUPERWOMAN1* and *DROOPING LEAF* genes control floral organ identity in rice. *Development* **130**: 705–18.

Nakajima K, Sena G, Nawy T, Benfey PN. 2001. Intercellular movement of the putative transcription factor SHR in root patterning. *Nature* **413**: 307–11.

Naseer S, Lee Y, Lapierre C, Franke R, Nawrath C, Geldner N. 2012. Casparian strip diffusion barrier in Arabidopsis is made of a lignin polymer without suberin. *Proc Natl Acad Sci* **109**: 10101–6.

Nelson T. 2011. The grass leaf developmental gradient as a platform for a systems understanding of the anatomical specialization of C_4 leaves. *J Exp Bot* **62**: 3039–48.

Nelson T, Dengler N. 1997. Leaf Vascular Pattern Formation. *Plant Cell* **9**: 1121–1135.

Nelson T, Dengler N. 1992. Photosynthetic tissue differentiation in C_4 plants. *Int J Plant Sci* **153**: S93–S105.

Nguyen D-Q, Webber C, Hehir-Kwa J, Pfundt R, Veltman J, Ponting CP. 2008. Reduced purifying selection prevails over positive selection in human copy number variant evolution. *Genome Res* **18**: 1711–23.

Ni J, Shen Y, Zhang Y, Wu P. 2014. Definition and stabilisation of the quiescent centre in rice roots. *Plant Biol* **16**: 1014–9.

Nomura M, Higuchi T, Ishida Y, Ohta S, Komari T, Imaizumi N, Miyao-Tokutomi M, Matsuoka M, Tajima S. 2005. Differential expression pattern of C_4 bundle sheath expression genes in rice, a C_3 plant. *Plant Cell Physiol* **46**: 754–61.

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 Mol Biol* **77**: 489–99.

Ogren WL. 2003. Affixing the O to Rubisco: discovering the source of photorespiratory glycolate and its regulation. *Photosynth Res* **76**: 53–63.

Ohno S. 1970. *Evolution by Gene Duplication*. Springer.

Osborne CP, Sack L, Hohmann-Marriott MF, Blankenship RE, Crayn DM, Winter K, Smith JAC, Edwards EJ, Osborne CP, Strömberg CAE, et al. 2012. Evolution of C_4 plants: a new hypothesis for an interaction of CO_2 and water relations mediated by plant hydraulics. *Philos Trans R Soc Lond B Biol Sci* **367**: 583–600.

Pagani M, Zachos JC, Freeman KH, Tipple B, Bohaty S. 2005. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* **309**: 600–603.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, et al. 2009. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**: 551–556.

Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci* **101**: 9903–9908.

Pengelly JJL, Kwasny S, Bala S, Evans JR, Voznesenskaya E V, Koteyeva NK, Edwards GE, Furbank RT, von Caemmerer S. 2011. Functional analysis of corn husk photosynthesis. *Plant Physiol* **156**: 503–13.

Peterhansel C, Blume C, Offermann S. 2013. Photorespiratory bypasses: how can they work? *J Exp Bot* **64**: 709–15.

Peterhansel C, Horst I, Niessen M, Blume C, Kebeish R, Kürkcüoglu S, Kreuzaler F. 2010. Photorespiration. In *The Arabidopsis Book*, Vol. 8 of, p. e0130, American Society of Plant Biologists.

Petricka JJ, Clay NK, Nelson TM. 2008. Vein patterning screens and the defectively organized tributaries mutants in *Arabidopsis thaliana*. *Plant J* **56**: 251–63.

Pick TR, Bräutigam A, Schlüter U, Denton AK, Colmsee C, Scholz U, Fahnenstich H, Pieruschka R, Rascher U, Sonnewald U, et al. 2011. Systems analysis of a maize leaf developmental gradient redefines the current C_4 model and provides candidates for regulation. *Plant Cell* **23**: 4208–20.

Pophaly SD, Tellier A. 2015. Population Level Purifying Selection and Gene Expression Shape Subgenome Evolution in Maize. *Mol Biol Evol* **32**: 3226–3235.

Prigge MJ, Otsuga D, Alonso JM, Ecker JR, Drews GN, Clark SE. 2005. Class III Homeodomain-Leucine Zipper Gene Family Members Have Overlapping, Antagonistic, and Distinct Roles in Arabidopsis Development. *Plant Cell* **17**: 61–76.

Qi J, Qian Q, Bu Q, Li S, Chen Q, Sun J, Liang W, Zhou Y, Chu C, Li X, et al. 2008. Mutation of the rice *Narrow leaf1* gene, which encodes a novel protein, affects vein patterning and polar auxin transport. *Plant Physiol* **147**: 1947–59.

Que Q, Elumalai S, Li X, Zhong H, Nalapalli S, Schweiner M, Fei X, Nuccio M, Kelliher T, Gu W, et al. 2014. Maize transformation technology development for commercial event generation. *Front Plant Sci* **5**: 379.

Rapp RA, Udall JA, Wendel JF. 2009. Genomic expression dominance in allopolyploids. *BMC Biol* **7**: 18.

Ray DK, Ramankutty N, Mueller ND, West PC, Foley JA. 2012. Recent patterns of crop yield growth and stagnation. *Nat Commun* **3**: 1293.

Reinhart BJ, Liu T, Newell NR, Magnani E, Huang T, Kerstetter R, Michaels S, Barton MK. 2013. Establishing a Framework for the Ad/Abaxial Regulatory Network of Arabidopsis: Ascertaining Targets of Class III HOMEODOMAIN LEUCINE ZIPPER and KANADI Regulation. *Plant Cell* **25**: 3228–3249.

Reiskind JB, Madsen T V., Van Ginkel LC, Bowes G. 1997. Evidence that inducible C_4-type photosynthesis is a chloroplastic CO_2-concentrating mechanism in Hydrilla, a submersed monocot. *Plant, Cell Environ* **20**: 211–220.

Rhoades M, Carvalho A. 1944. The function and structure of the parenchyma sheath plastids of the maize leaf. *Bull Torrey Bot Club* **71**: 335–346.

Rim Y, Huang L, Chu H, Han X, Cho WK, Jeon CO, Kim HJ, Hong J-C, Lucas WJ, Kim J-Y. 2011. Analysis of Arabidopsis transcription factor families revealed extensive capacity for cell-to-cell movement as well as discrete trafficking patterns. *Mol Cells* **32**: 519–26.

Rizal G, Thakur V, Dionora J, Karki S, Wanchana S, Acebron K, Larazo N, Garcia R, Mabilangan A, Montecillo F, et al. 2015. Two forward genetic screens for vein density mutants in sorghum converge on a cytochrome P450 gene in the brassinosteroid pathway. *Plant J* **84**: 257–266.

Rondeau P, Rouch C, Besnard G. 2005. NADP-malate dehydrogenase gene evolution in Andropogoneae (Poaceae): gene duplication followed by sub-functionalization. *Ann Bot* **96**: 1307–1314.

Rossini L, Cribb L, Martin DJ, Langdale JA. 2001. The maize golden2 gene defines a novel class of transcriptional regulators in plants. *Plant Cell* **13**: 1231–44.

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

Russell SH, Evert RF. 1985. Leaf vasculature in *Zea mays* L. *Planta* **164**: 448–458.

Sabatini S, Heidstra R, Wildwater M, Scheres B. 2003. SCARECROW is involved in positioning the stem cell niche in the Arabidopsis root meristem. *Genes Dev* **17**: 354–8.

Sage RF. 2001. Environmental and Evolutionary Preconditions for the Origin and Diversification of the C_4 Photosynthetic Syndrome. *Plant Biol* **3**: 202–213.

Sage RF. 2004. The evolution of C_4 photosynthesis. *New Phytol* **161**: 341–370.

Sage RF, Christin P-A, Edwards EJ. 2011a. The C_4 plant lineages of planet Earth. *J Exp Bot* **62**: 3155–69.

Sage RF, Pearcy RW. 1987. The Nitrogen Use Efficiency of C3 and C4 Plants: II. Leaf Nitrogen Effects on the Gas Exchange Characteristics of *Chenopodium album* (L.) and *Amaranthus retroflexus* (L.). *Plant Physiol* **84**: 959–963.

Sage RF, Sage TL, Kocacinar F. 2012. Photorespiration and the Evolution of C_4 Photosynthesis. *Annu Rev Plant Biol* **63**: 19–47.

Sage RF, Zhu X-G. 2011. Exploiting the engine of C_4 photosynthesis. *J Exp Bot* **62**: 2989–3000.

Sage TL, Busch FA, Johnson DC, Friesen PC, Stinson CR, Stata M, Sultmanis S, Rahman BA, Rawsthorne S, Sage RF. 2013. Initial events during the evolution of C_4 photosynthesis in C_3 species of Flaveria. *Plant Physiol* **163**: 1266–76.

Sage TL, Sage RF, Vogan PJ, Rahman B, Johnson DC, Oakley JC, Heckel MA. 2011b. The occurrence of C_2 photosynthesis in *Euphorbia* subgenus *Chamaesyce* (Euphorbiaceae). *J Exp Bot* **62**: 3183–95.

Saha P, Blumwald E. 2016. Spike-dip transformation of *Setaria viridis*. *Plant J* **86**: 89–101.

Samalova M, Brzobohaty B, Moore I. 2005. pOp6/LhGR: a stringently regulated and highly responsive dexamethasone-inducible gene expression system for tobacco. *Plant J* **41**: 919–935.

Scarpella E, Barkoulas M, Tsiantis M. 2010. Control of leaf and vein development by auxin. *Cold Spring Harb Perspect Biol* **2**: a001511.

Scarpella E, Marcos D, Friml J, Berleth T. 2006. Control of leaf vascular patterning by polar auxin transport. *Genes Dev* **20**: 1015–27.

Scarpella E, Rueb S, Meijer AH. 2003. The *RADICLELESS1* gene is required for vascular pattern formation in rice. *Development* **130**: 645–658.

Scheres B, Di Laurenzio L, Willemsen V, Hauser MT, Janmaat K, Weisbeek P, Benfey PN. 1995. Mutations affecting the radial organisation of the Arabidopsis root display specific defects throughout the embryonic axis. *Development* **121**: 53–62.

Schnable JC, Springer NM, Freeling M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc Natl Acad Sci* **108**: 4069–4074.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**: 1112–1115.

Schulze S, Mallmann J, Burscheidt J, Koczor M, Streubel M, Bauwe H, Gowik U, Westhoff P. 2013. Evolution of C_4 photosynthesis in the genus flaveria: establishment of a photorespiratory CO_2 pump. *Plant Cell* **25**: 2522–2535.

Seo PJ, Ryu J, Kang SK, Park C-M. 2011. Modulation of sugar metabolism by an INDETERMINATE DOMAIN transcription factor contributes to photoperiodic flowering in Arabidopsis. *Plant J* **65**: 418–429.

Settles AM, Holding DR, Tan BC, Latshaw SP, Liu J, Suzuki M, Li L, O'Brien BA, Fajardo DS, Wroclawska E, et al. 2007. Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. *BMC Genomics* **8**: 116.

Shan H, Zahn L, Guindon S, Wall PK, Kong H, Ma H, DePamphilis CW, Leebens-Mack J. 2009. Evolution of plant MADS box transcription factors: evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications. *Mol Biol Evol* **26**: 2229–44.

Sharkey TD. 1988. Estimating the rate of photorespiration in leaves. *Physiol Plant* **73**: 147–152.

Sharman B. 1942. Developmental anatomy of the shoot of *Zea mays* L. *Ann Bot* **6**: 245–281.

Sheehy J, Ferrer A, Mitchell P, Elmido-Mabilangan A, Pablico P, Dionora M. 2007. How the rice crop works and why it needs a new engine. In *Charting New Pathways to C4 Rice* (eds. J. Sheehy, P. Mitchell, and B. Hardy), pp. 3–26, Int. Rice Res. Inst., Los Banos (Philippines).

Sheen JY, Bogorad L. 1988. Differential Expression in Bundle Sheath and Mesophyll Cells of Maize of Genes for Photosystem II Components Encoded by the Plastid Genome. *Plant Physiol* **86**: 1020–6.

Shiu S-H, Byrnes JK, Pan R, Zhang P, Li W-H. 2006. Role of positive selection in the retention of duplicate genes in mammalian genomes. *Proc Natl Acad Sci* **103**: 2232–6.

Slewinski TL, Anderson AA, Price S, Withee JR, Gallagher K, Turgeon R. 2014. Short-Root1 Plays a Role in the Development of Vascular Tissue and Kranz Anatomy in Maize Leaves. *Mol Plant* **7**: 1388–1392.

Slewinski TL, Anderson AA, Zhang C, Turgeon R. 2012. Scarecrow plays a role in establishing Kranz anatomy in maize leaves. *Plant Cell Physiol* **53**: 2030–7.

Smillie IRA, Pyke KA, Murchie EH. 2012. Variation in vein density and mesophyll cell architecture in a rice deletion mutant population. *J Exp Bot* **63**: 4563–4570.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. *Am J Bot* **96**: 336–48.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–3.

Steige KA, Slotte T. 2016. Genomic legacies of the progenitors and the evolutionary consequences of allopolyploidy. *Curr Opin Plant Biol* **30**: 88–93.

Strable J, Scanlon MJ. 2009. Maize (*Zea mays*): A Model Organism for Basic and Applied Research in Plant Biology. *Cold Spring Harb Protoc* **2009**: pdb.emo132–pdb.emo132.

Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J. 2004a. Close split of sorghum and maize genome progenitors. *Genome Res* **14**: 1916–1923.

Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J. 2004b. On the tetraploid origin of the maize genome. *Comp Funct Genomics* **5**: 281–284.

Takabayashi A, Kishine M, Asada K, Endo T, Sato F. 2005. Differential use of two cyclic electron flows around photosystem I for driving CO_2-concentration mechanism in C_4 photosynthesis. *Proc Natl Acad Sci* **102**: 16898–16903.

Talbert PB, Adler HT, Parks DW, Comai L. 1995. The *REVOLUTA* gene is necessary for apical meristem development and for limiting cell divisions in the leaves and stems of *Arabidopsis thaliana*. *Development* **121**: 2723–2735.

Tang H, Bowers JE, Wang X, Paterson AH. 2010. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc Natl Acad Sci* **107**: 472–477.

Taniguchi Y, Ohkawa H, Masumoto C, Fukuda T, Tamai T, Lee K, Sudoh S, Tsuchida H, Sasaki H, Fukayama H, et al. 2008. Overproduction of C_4 photosynthetic enzymes in transgenic rice plants: an approach to introduce the C_4-like photosynthetic pathway into rice. *J Exp Bot* **59**: 1799–809.

Tanimoto M, Tremblay R, Colasanti J. 2008. Altered gravitropic response, amyloplast sedimentation and circumnutation in the Arabidopsis shoot gravitropism 5 mutant are associated with reduced starch levels. *Plant Mol Biol* **67**: 57–69.

Tateoka T. 1958. Notes on Some Grasses. VIII. On Leaf Structure of Arundinella and Garnotia. *Bot Gaz* **120**: 101–109.

Tausta SL, Li P, Si Y, Gandotra N, Liu P, Sun Q, Brutnell TP, Nelson T. 2014. Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C_4-related processes. *J Exp Bot* **65**: 3543–55.

Tcherkez GGB, Farquhar GD, Andrews TJ. 2006. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc Natl Acad Sci* **103**: 7246–51.

Thomas BC, Pedersen B, Freeling M. 2006. Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res* **16**: 934–46.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178–192.

Tsiantis M, Brown MI, Skibinski G, Langdale JA. 1999. Disruption of auxin transport is associated with aberrant leaf development in maize. *Plant Physiol* **121**: 1163–8.

Ueno O. 1995. Occurrence of Distinctive Cells in Leaves of C_4 Species in Arthraxon and Microstegium (Andropogoneae-Poaceae) and the Structural and Immunocytochemical Characterization of these Cells. *Int J Plant Sci* **156**: 270–289.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**: e115.

Vats SK, Kumar S, Ahuja PS. 2011. CO_2 sequestration in plants: lesson from divergent strategies. *Photosynthetica* **49**: 481–496.

Vollbrecht E, Duvick J, Schares JP, Ahern KR, Deewatthanawong P, Xu L, Conrad LJ, Kikuchi K, Kubinec TA, Hall BD, et al. 2010. Genome-wide distribution of transposed Dissociation elements in maize. *Plant Cell* **22**: 1667–85.

Voznesenskaya E V, Franceschi VR, Kiirats O, Artyusheva EG, Freitag H, Edwards GE. 2002. Proof of C_4 photosynthesis without Kranz anatomy in *Bienertia cycloptera* (Chenopodiaceae). *Plant J* **31**: 649–62.

Voznesenskaya E V, Franceschi VR, Kiirats O, Freitag H, Edwards GE. 2001. Kranz anatomy is not essential for terrestrial C_4 plant photosynthesis. *Nature* **414**: 543–6.

Voznesenskaya E V., Chuong SDX, Koteyeva NK, Edwards GE, Franceschi VR. 2005. Functional compartmentation of C_4 photosynthesis in the triple layered chlorenchyma of Aristida (Poaceae). *Funct Plant Biol* **32**: 67–77.

Voznesenskaya E V., Edwards GE, Kiirats O, Artyusheva EG, Franceschi VR. 2003. Development of biochemical specialization and organelle partitioning in the single-cell C_4 system in leaves of *Borszczowia aralocaspica* (Chenopodiaceae). *Am J Bot* **90**: 1669–1680.

Voznesenskaya E V., Koteyeva NK, Chuong SDX, Ivanova AN, Barroca J, Craven LA, Edwards GE. 2007. Physiological, anatomical and biochemical characterisation of photosynthetic types in genus Cleome (Cleomaceae). *Funct Plant Biol* **34**: 247.

Wakayama M, Ohnishi J, Ueno O. 2006. Structure and enzyme expression in photosynthetic organs of the atypical C_4 grass *Arundinella hirta*. *Planta* **223**: 1243–55.

Wakayama M, Ueno O, Ohnishi J-I. 2003. Photosynthetic enzyme accumulation during leaf development of Arundinella hirta, a C_4 grass having Kranz cells not associated with veins. *Plant Cell Physiol* **44**: 1330–40.

Wang D, Liu F, Wang L, Huang S, Yu J. 2011. Nonsynonymous substitution rate (*Ka*) is a relatively consistent parameter for defining fast-evolving and slow-evolving protein-coding genes. *Biol Direct* **6**: 13.

Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* **8**: 77–80.

Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, et al. 2006. Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. *Genetics* **172**: 507–517.

Wang L, Czedik-Eysenberg A, Mertz RA, Si Y, Tohge T, Nunes-Nesi A, Arrivault S, Dedow LK, Bryant DW, Zhou W, et al. 2014a. Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice. *Nat Biotechnol* **32**: 1158–65.

Wang MB, Li ZY, Matthews PR, Upadhyaya NM, Waterhouse PM. 1998. Improved vectors for *Agrobacterium* tumefaciens-mediated transformation of monocot plants. *Acta Hortic* **461**: 401–407.

Wang P, Fouracre J, Kelly S, Karki S, Gowik U, Aubry S, Shaw MK, Westhoff P, Slamet-Loedin IH, Quick WP, et al. 2013a. Evolution of *GOLDEN2-LIKE* gene function in C_3 and C_4 plants. *Planta* **237**: 481–95.

Wang P, Kelly S, Fouracre JP, Langdale JA. 2013b. Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C_4 Kranz anatomy. *Plant J* **75**: 656–670.

Wang P, Vlad D, Langdale JA. 2016. Finding the genes to build C_4 rice. *Curr Opin Plant Biol* **31**: 44–50.

Wang X, Gowik U, Tang H, Bowers JE, Westhoff P, Paterson AH, Hatch M, Slack C, Seemann J, Sharkey T, et al. 2009. Comparative genomic analysis of C_4 photosynthetic pathway evolution in grasses. *Genome Biol* **10**: R68.

Wang Y, Bräutigam A, Weber APM, Zhu X-G. 2014b. Three distinct biochemical subtypes of C_4 photosynthesis? A modelling analysis. *J Exp Bot* **65**: 3567–78.

Waszczak C, Kerchev PI, Mühlenbock P, Hoeberichts FA, Van Der Kelen K, Mhamdi A, Willems P, Denecker J, Kumpf RP, Noctor G, et al. 2016. SHORT-ROOT Deficiency Alleviates the Cell Death Phenotype of the Arabidopsis catalase2 Mutant under Photorespiration-Promoting Conditions. *Plant Cell* **28**: 1844–1859.

Watt M, Schneebeli K, Dong P, Wilson IW. 2009. The shoot and root growth of Brachypodium and its potential as a model for wheat and other cereal crops. *Funct Plant Biol* **36**: 960.

Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. 2011. Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. ed. J. Peccoud. *PLoS One* **6**: e16765.

Wei F, Coe E, Nelson W, Bharti AK, Engler F, Butler E, Kim H, Goicoechea JL, Chen M, Lee S, et al. 2007. Physical and genetic structure of the maize genome reflects its complex evolutionary history. *PLoS Genet* **3**: e123.

Wei K, Pan S, Li Y. 2015. Functional Characterization of Maize C2H2 Zinc-Finger Gene Family. *Plant Mol Biol Report* **34**: 761–776.

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 Dev* **21**: 2196–204.

Werner S, Engler C, Weber E, Gruetzner R, Marillonnet S. 2012. Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioeng Bugs* **3**: 38–43.

Westhoff P, Gowik U. 2010. Evolution of C_4 Photosynthesis--Looking for the Master Switch. *Plant Physiol* **154**: 598–601.

Williams-Carrier R, Stiffler N, Belcher S, Kroeger T, Stern DB, 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. *Plant J* **63**: 167–77.

Williams BP, Aubry S, Hibberd JM. 2012. Molecular evolution of genes recruited into C_4 photosynthesis. *Trends Plant Sci* **17**: 213–20.

Williams BP, Johnston IG, Covshoff S, Hibberd JM, Abramoff M, Magalhaes P, Ram S, Adams C, Leung F, Sun S, et al. 2013. Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis. *Elife* **2**: e00961.

Wisniewska J, Xu J, Seifertová D, Brewer PB, Ruzicka K, Blilou I, Rouquié D, Benková E, Scheres B, Friml J. 2006. Polar PIN localization directs auxin flow in plants. *Science* **312**: 883.

Wong SL, Roth FP. 2005. Transcriptional compensation for gene loss plays a minor role in maintaining genetic robustness in *Saccharomyces cerevisiae*. *Genetics* **171**: 829–833.

Woo KC. 1970. Deficient Photosystem II in Agranal Bundle Sheath Chloroplasts of C_4 Plants. *Proc Natl Acad Sci* **67**: 18–25.

Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. *PLoS Biol* **8**: e1000409.

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection on the maize genome. *Science* **308**: 1310–4.

Wright SI, Yau CBK, Looseley M, Meyers BC. 2004. Effects of Gene Expression on Molecular Evolution in *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Mol Biol Evol* **21**: 1719–1726.

Wu S, Lee C-M, Hayashi T, Price S, Divol F, Henry S, Pauluzzi G, Perin C, Gallagher KL. 2014. A plausible mechanism, based upon Short-Root movement, for regulating the number of cortex cell layers in roots. *Proc Natl Acad Sci* **111**: 16184–9.

Wu X, Tang D, Li M, Wang K, Cheng Z. 2013. Loose Plant Architecture1, an INDETERMINATE DOMAIN protein involved in shoot gravitropism, regulates plant architecture in rice. *Plant Physiol* **161**: 317–329.

Wysocka-Diller JW, Helariutta Y, Fukaki H, Malamy JE, Benfey PN. 2000. Molecular analysis of SCARECROW function reveals a radial patterning mechanism common to root and shoot. *Development* **127**: 595–603.

Xu H, Lu Y, Tong S, Song F. 2011. Lipid peroxidation , antioxidant enzyme activity and osmotic adjustment changes in husk leaves of maize in black soils region of Northeast China. **6**: 3098–3102.

Yamaguchi T, Nagasawa N, Kawasaki S, Matsuoka M, Nagato Y, Hirano H-Y. 2004. The YABBY gene *DROOPING LEAF* regulates carpel specification and midrib development in *Oryza sativa*. *Plant Cell* **16**: 500–509.

Yi G, Neelakandan AK, Gontarek BC, Vollbrecht E, Becraft PW. 2015. The naked endosperm genes encode duplicate INDETERMINATE domain transcription factors required for maize endosperm cell patterning and differentiation. *Plant Physiol* **167**: 443–56.

Yinghui L, Jingjuan Y, Guangming A, Qian Z. 2007. Factors influencing *Agrobacterium*-mediated transformation of foxtail millet (*Setaria italica*). *Chinese J Biochem Mol Biol* **23**: 531–536.

Yoo M-J, Szadkowski E, Wendel JF. 2013. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **110**: 171–80.

Yoo SY, Bomblies K, Yoo SK, Yang JW, Choi MS, Lee JS, Weigel D, Ahn JH. 2005. The 35S promoter used in a selectable marker gene of a plant transformation vector affects the expression of the transgene. *Planta* **221**: 523–530.

Yoon EK, Dhar S, Lee M-H, Song JH, Lee SA, Kim G, Jang S, Choi JW, Choe J-E, Kim JH, et al. 2016. Conservation and Diversification of the SHR-SCR-SCL23 Regulatory Network in the Development of the Functional Endodermis in Arabidopsis Shoots. *Mol Plant* **9**: 1197–1209.

Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* **11**: R14.

Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol* **18**: 292–298.

Zhang J, Rosenberg HF, Nei M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci* **95**: 3708–13.

Zhang L, Li W-H. 2004. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Mol Biol Evol* **21**: 236–9.

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 Rep* **26**: 1195–1203.

Zhu X-G, Portis AR, Long SP. 2004. Would transformation of C_3 crop plants with foreign Rubisco increase productivity? A computational analysis extrapolating from kinetic properties to canopy photosynthesis. *Plant, Cell Environ* **27**: 155–165.