Genetic regulation of Kranz anatomy

Jim Fouracre

Worcester College

University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2013

Abstract

The C_4 photosynthetic cycle acts to concentrate CO_2 around the enzyme Rubisco. By doing so, C_4 photosynthesis leads to increased radiation, water and nitrogen use efficiencies. As such, C_4 photosynthesis is the most productive form of photosynthesis known. Because it enables such high levels of productivity there are large international efforts to introduce C_4 photosynthesis into non-C_4 crop species such as rice. Kranz anatomy is a characteristic leaf cellular arrangement of concentric rings of bundle sheath and mesophyll cells around closely spaced veins and is crucial to C_4 photosynthesis in almost all known examples. Despite the fact that Kranz has evolved on over 60 times independently little is known about the genetic regulation of Kranz development, as attempts to elucidate Kranz regulators using conventional mutagenesis screens have provided few insights. However, the advent of next generation DNA sequencing technologies has enabled the interrogation of genetic networks at a previously unprecedented scale. The work in this thesis describes a genome-wide transcriptomic analysis of leaf development in maize, a C_4 species, that develops both Kranz-type and non-Kranz-type leaves. Detailed bioinformatics analyses identified candidate regulators of both Kranz development and additional aspects of maize leaf development. Three of the identified Kranz candidates were functionally characterised in both C_4 and non-C_4 species. Furthermore, expression and phylogenetic analyses of *GOLDEN2-LIKE* (*GLK*) genes, a small transcription factor family previously implicated in C_4 development in maize, were extended to determine the generality of *GLK* function in C_4 evolution.

List of abbreviations

Ac/DS – Activator/Dissociation

ATP – adenosine triphosphate

bHLH – basic helix-loop-helix

BLAST – basic local alignment search tool

Bd – *Brachypodium distachyon*

BS – bundle sheath

BSD – bundle sheath distance

CA – carbonic anhydrase

cDNA – complementary DNA

DC – distinctive cell

DNA – deoxyribonucleic acid

DOF – DNA BINDING WITH ONE ZINC FINGER

DOT5 – DEFECTIVELY ORGANISED TRIBUTARIES 5

EV – empty vector

FA – foliar ascending

FD – foliar descending

FE – foliar expanded

FI – foliar immature

FP – foliar primordia

GCT – GLK C-terminal box

GLK – GOLDEN2-LIKE

GO – gene ontology

GWD – genome-wide duplication

HA – husk ascending

HD – husk descending

HE – husk exposed

HI – husk inner

HMM – hidden Markov model

HP – husk primordia

IVD – intervein distance

LCM – laser capture microdissection

LRR-RLK – leucine rich repeat – receptor-like kinase

M – mesophyll

MDH – malate dehydrogenase

MEME – multiple expectation maximisation for motif elicitation

mRNA – messenger RNA

NAD-ME – nicotinamide adenine dinucleotide dependent – malic enzyme

NADP-ME – NAD phosphate dependent – malic enzyme

Os – *Oryza sativa*

PEP – phosphoenol pyruvate

PEPC – PEP carboxylase

PEPCK – PEP carboxykinase

PIN – PIN-FORMED1

PPDK – pyruvate orthophosphate dikinase

PsCA – photosynthetic carbon assimilation

PsCR – photosynthetic carbon reduction

RNA – ribonucleic acid

RNAi – RNA interference

RNA-seq – RNA sequencing

RPKM – reads per kilobase per million reads

RSEM – RNA-seq by expectation maximisation

Rubisco – ribulose-1,5-bisphosphate carboxylase oxygenase

SCR – SCARECROW

SHR – SHORT-ROOT

Si – *Setaria italica*

SPCH – SPEECHLESS

TF – transcription factor

WT – wild type

Zm – *Zea mays*

Acknowledgements

I would like to express my gratitude to Jane Langdale for all the preserves and delicately flavoured sauces she has provided over the last 4 years, my palate will never be the same again. I could not have asked for better supervision, in either a culinary or academic sense. I would also like to thank my second supervisor, Nick Harberd, for timely and insightful suggestions. In addition I would particularly like to thank Peng Wang and Steve Kelly, without their help this thesis would not have been possible. Peng carried out the dissections that provided the majority of the data for this thesis and held my hand at various molecular junctures; Steve dragged us out of a bioinformatic hole. Some of my fondest memories of Oxford will be those crisp winter afternoons I sat behind you watching you code.

I would also like to thank the rest of the Langdale lab for advice, help and for creating such a supportive working environment. Particularly Mara Schuler for the education in serve volleying and reggae, Eftychis Frangedakis for lessons in Turkish history, Sayuri Ando for the tissue culture assistance and lab coat craft sessions, Mark Waters for gold standard molecular training, Matt Hodges for insights into the importance of a work ethic and Ol' Granny Jules for patient support, both technical and maternal. Thanks also to John Baker for help with plant photography.

I would like to thank the BBSRC for a postgraduate studentship, the Gatsby Charitable Foundation for extremely generous funding and training over many years, Worcester College for support and accommodation, the Genetics

Society and BSCB for conference funding, members of the C_4 rice consortium for advice and sharing data (particularly the Hibberd and Westhoff labs) and IRRI for training and inspiration.

Without the support and encouragement of many friends I may well have published higher and finished quicker, but you cannot put a price on trips to Maxwell's disco diner. I would especially like to thank all the brave brothers (and occasional sister) who pulled on the red and blue or faded green jerseys of Spartak Botley and PSCC with me. Together we learned the joys of solidarity, passion and to accept losing on a regular basis. Special thanks to Matt for all the tandem rides we never went on, Mark for the quiz victories, Luke for the insights into the mind of Kafka, Laith for exemplary mateship, Joe for the South London pride, Zoe for the Friday fishing trips and Madeline for marathon tolerance.

Finally, I would like to thank my family for the mostly constructive criticism and for putting up with me. Thanks to Dorothy and Louise for the insights into possible alternative mechanisms for mammalian evolution (although I am not sure Darwinism has much to worry about) and to my parents for unstinting support: PJ for nurturing my interest in plants as a teenager and Duncs for the lessons in scientific rigour.

Figure legend

Published elements

Elements of this study have been published as research papers in two peer-reviewed journals:

Evolution of *GOLDEN2-LIKE* gene function in C_3 and C_4 plants. Wang, P.,* Fouracre, J.,* Kelly, S., Karki, S., Gowik, U., Aubry, S., Shaw, M.K., Westhoff, P., Slamet-Loedin, I.H., Quick, W.P., Hibberd, J.M., Langdale, J.A. *Planta* (2013) 237, 481-495

The author contributions to the article were as follows: P. Wang carried out rice transformation and characterisation; J. Fouracre isolated sorghum BS and M cells, carried out northern blot analyses and, with S. Kelly, determined gene phylogenies; S. Karki generated rice double mutants; U. Gowik isolated sorghum BS and M cells and carried out Illumina sequencing; S. Aubry isolated *Cleome gynandra* BS and M cells and carried out qPCR analyses; M. Shaw carried out electron microscopy; P. Westhoff, I. Slamet-Loedin, W.P. Quick, J. Hibberd and J. Langdale supervised and funded the research; P.W., J.F., S.Ke and J.A. wrote the paper. P.W. and J.F. contributed equally.

Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C_4 Kranz anatomy. Wang, P.,* Kelly, S.,* Fouracre, J.,* and Langdale, J.A. *Plant Journal* (2013) 75, 656-670

The author contributions to the article were as follows: P. Wang prepared plant material and extracted RNA; S. Kelly designed the bioinformatics methodology and wrote the code; and J. Fouracre validated the quality of sequencing data and performed the profile analyses; J.A. Langdale proposed the original experimental design and obtained funding. All authors designed the data filtration criteria, analysed datasets and wrote the paper. PW, SK and JF contributed equally.

Guide to gene nomenclature

While the distinctions between the genetic nomenclature systems used for different species are recognised, for simplicity and cohesiveness the following system of nomenclature will be used throughout this thesis:

Gene: Italicised upper case, e.g. *GLK*

Protein: Non-italicised upper case, e.g. GLK

Recessive mutant: Italicised lower case, e.g. *glk*

Dominant mutant: Italicised, first letter only upper case, e.g. *Glk*

Contents

Chapter 1: Introduction

1.1 The C_4 photosynthetic pathway

Plants fix CO_2 in one of three pathways: C_3, C_4 or CAM (Crassulacean Acid Metabolism). The majority use the C_3 pathway, whereby atmospheric CO_2 is fixed by the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), to form two molecules of the three-carbon compound 3-phosphoglycerate (3-PGA) (Figure 1.1A). Following the discovery of the C_3 cycle (Bassham et al., 1956) it was assumed that this method of carbon reduction was the first stage of carbon assimilation in all plants. However, while this cycle remains of central importance to the process of photosynthetic carbon assimilation, it was later discovered that in some plant lineages CO_2 is initially fixed into a four carbon compound, instead of 3-PGA (Karpilov, 1960; Kortschak et al., 1965). This reaction was later shown to define a novel extension to the C_3 pathway, in which the four carbon compound that is initially produced is subsequently decarboxylated, releasing CO_2 that enters the C_3 pathway. This novel extension was thus termed the C_4 dicarboxylic acid, or C_4, pathway (Hatch and Slack, 1966).

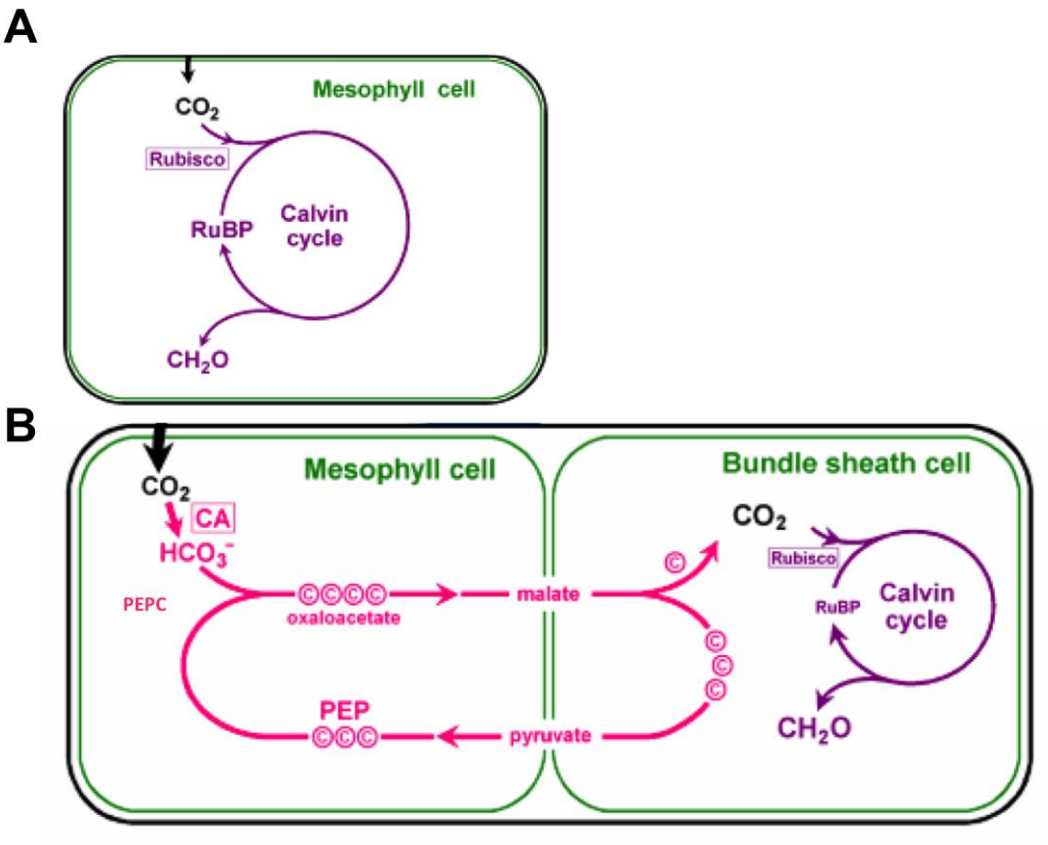


Figure 1.1. Schematic of simplified C_3 and C_4 photosynthetic pathways. A) C_3 pathway B) C_4 pathway. Rubisco – ribulose-1,5-bisphosphate carboxylase/oxygenase, RuBP – Ribulose-1,5-bisphosphate, CA – carbonic anhydrase, PEP – phosphoenol pyruvate, PEPC – PEP carboxylase, © - carbon atom. C_3 pathway is shown in purple, C_4 pathway is shown in pink. Adapted from Langdale (2011).

Figure 1.1B depicts a simplified version of the C_4 pathway as it exists in maize and sugar cane, the latter being the species in which it was originally discovered. In these species, CO_2 is initially fixed into a molecule of HCO_3^- by the enzyme carbonic anhydrase (CA) in mesophyll (M) cells. HCO_3^- is combined with phosphoenol pyruvate (PEP) to form the four carbon acid oxaloacetate (OAA) by the enzyme PEP carboxylase (PEPC), which is then converted into malate by the enzyme malate dehydrogenase (MDH). Malate is subsequently shuttled to a neighbouring bundle sheath (BS) cell where it is decarboxylated by nicotinamide adenine dinucleotide phosphate (NADP) dependent malic enzyme (ME). The CO_2 that is released enters the C_3 cycle via Rubisco, and the byproduct pyruvate is shuttled back to the M cell where it is phosphorylated by pyruvate orthophosphate dikinase (PPDK) to regenerate PEP. This process leads to high CO_2 concentrations localised around Rubisco, which suppresses the competing oxygenation reaction and therefore reduces the need for photorespiration (reviewed in Bauwe, 2011). In C_3 plants photorespiration is required to remove the toxic byproducts of the oxygenase reaction, which comes with energetic costs and the loss of carbon. The C_4 pathway also comes at an energetic cost, 2 molecules of ATP are required to regenerate PEP from pyruvate per molecule of fixed CO_2 (Kanai et al., 1999), but these costs are offset by the increase in photosynthetic efficiency that the C_4 pathway confers, at least in hot and arid growth conditions.

The processes of initial CO_2 fixation by CA and PEPC, and localised rerelease of CO_2 in a compartment containing Rubisco for refixation, are often described as photosynthetic carbon assimilation (PsCA) and photosynthetic

carbon reduction (PsCR) respectively. Both PsCA and PsCR are common to all C_4 species, however, there is variation within other components of the pathway, most notably the type of C_4 compound shuttled between cells and the decarboxylation enzyme used in BS cells. Traditionally, three subtypes of C_4 have been recognised based on the decarboxylation enzyme that is employed: NADP-ME, NAD-ME or PEP-carboxykinse (PEPCK). (reviewed in Drincovich et al., 2011). The observation that certain species use multiple decarboxylation enzymes, however, has cast doubt over the validity of this classification system (Furbank, 2011).

Examples of single celled C_4 photosynthesis exist in aquatic plants, in which PsCA and PsCR are split between the cytoplasm and the chloroplast (Bowes, 2011), and in the Chenopodiaceae, in which they are separated within cytoplasmic compartments (Voznesenskaya et al., 2001; Voznesenskaya et al., 2002; Akhani et al., 2005; Voznesenskaya et al., 2005a) (Figure 1.2). However, in almost all cases, C_4 photosynthesis is dependent on the division of labour of PsCA and PsCR between M and BS cells. This metabolic relationship requires a close association between M and BS cells and is evident in the characteristic C_4 leaf anatomy known as Kranz. Kranz anatomy was first described by Haberlandt in 1882 and is defined by concentric wreaths of M and BS cells around closely spaced veins (Kranz is German for wreath) (Brown, 1975). This arrangement maximises the number of M cells in contact with BS cells and therefore increases the capacity of the C_4 pathway (Figure 1.2). While Kranz is key to C_4 in the vast majority of plants, subtle

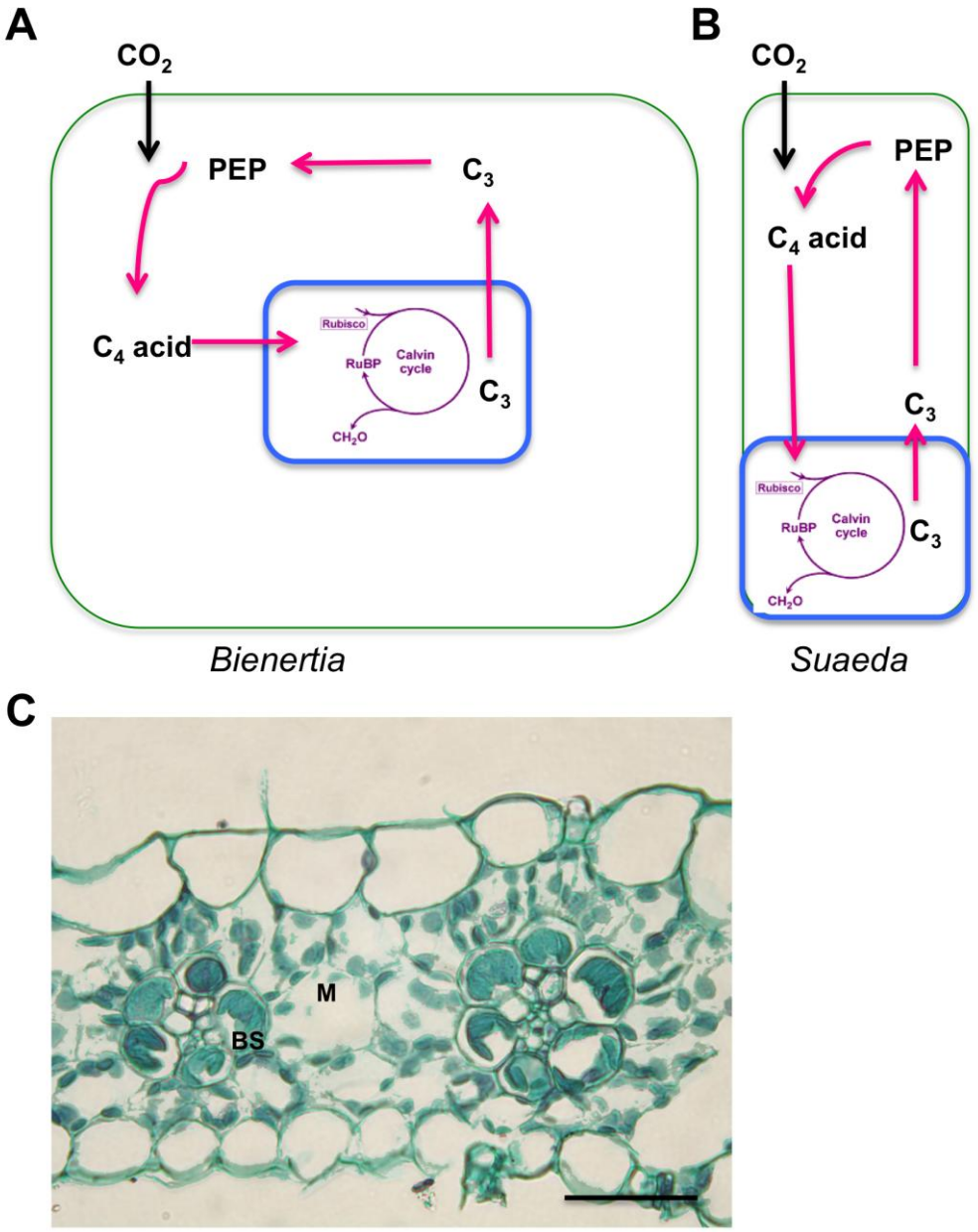


Figure 1.2. Variation in C_4 anatomy. A, B) Schematics of terrestrial single celled C_4 photosynthesis. Pink represents C_4 shuttle, purple represents C_3 pathway, blue boxes represent compartmentalisation of C_4 acid decarboxylation and Rubisco activity. In *Bienertia* (A) this happens within a central cytoplasmic compartment, in *Suaeda* (B) this is localised to a single end of a cell. (A, B) Adapted from Edwards and Voznesenskaya (2011). **C)** Kranz anatomy in a cross section of a sorghum leaf. BS – bundle sheath, M – mesophyll**.** Black scale bar = 50µm

variations in form are also known to exist. This will be discussed in greater detail in section 1.3.1.

1.2 The evolution of C_4

1.2.1 Phylogenetic distribution

In recent years, enhanced sampling strategies (in terms of both the taxa and genes surveyed) have led to progressive increases in estimates for the number of evolutionary origins of C_4. The projected minimum number of origins has swelled from 31 (Kellogg, 1999), to 45 (Sage, 2004), to 62 (Sage et al., 2011a) and most recently to 66 (Sage et al., 2012). As new taxa that exhibit C_4 photosynthesis continue to be discovered these estimates are likely to continue to rise. However, it is already clear that C_4 photosynthesis is one of the most remarkable examples of convergent evolution discovered thus far. Approximately 7500 C_4 species exist of which most (~4600) are grasses, despite the highest number of origins occurring in the eudicots (36 lineages) (Sage et al., 2011a). NADP-ME is the primary decarboxylase in the majority of lineages (~43) and all but four lineages exhibit Kranz anatomy. Single celled C_4 photosynthesis is found in the aquatic lineages *Hydrilla* and *Egeria* and in the Chenopodiaceae genera *Bienertia* and *Suaeda*. The clustering of C_4 evolution within certain orders, most notably the Caryophyllales and Poales, suggests that certain lineages may exhibit traits that 'preadapt' them for

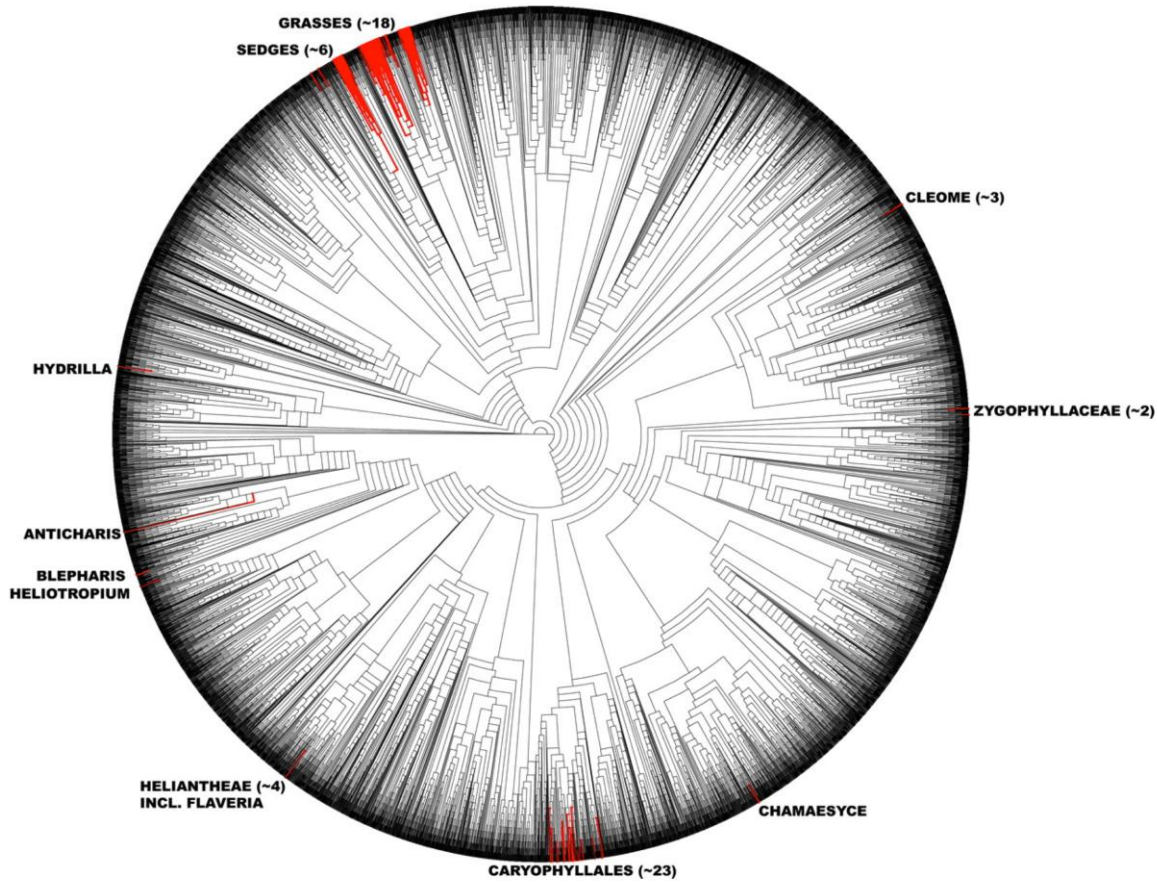


Figure 1.3. C_4 origins are clustered within the angiosperms. 62 origins of C_4 are highlighted in red in an angiosperm phylogeny (Figure taken from Sage et al 2011a).

the evolution of C_4 (Sage, 2001; Christin et al., 2013) (Figure 1.3). C_3-C_4 intermediates species that display biochemical and anatomical adaptations intermediate of C_3 and C_4 plants also exist. These are predominantly found within the eudicots in lineages that also feature full C_4 species, although examples are known in lineages without any C_4 sister species (Sage et al., 2011a). There is some debate as to whether these intermediates represent evolutionary transitions between C_3 and C_4 states, or are evolutionarily stable independent forms (e.g. Monson and Moore, 1989; McKown and Dengler, 2007; Langdale, 2011; Heckmann et al., 2013). Instances of evolutionary reversals from C_4 to C_3 have been suggested, however, these have not been conclusively demonstrated and are negligible in number compared with the transitions from C_3 to C_4 (Christin et al., 2010a).

1.2.2 Ecological drivers of C_4 evolution

The benefits of C_4 photosynthesis arise from enhanced localised CO_2 around Rubisco, which reduces photorespiration and therefore increases photosynthetic efficiency. In addition, by increasing the efficiency of Rubisco, C_4 plants are more water use and nitrogen use efficient because they have a lower requirement for stomatal opening and don't need to synthesise as much Rubisco (Ghannoum et al., 2011). The problem of photorespiration is exacerbated in conditions of high temperature (CO_2 is less soluble than O_2 with increased heat and Rubisco is less effective at discriminating between them), and low CO_2 availability (including as a consequence of reduced stomatal opening due to aridity) (reviewed in Bauwe, 2011). Conversely, due

to the increased energetic cost of the C_4 pathway, C_4 photosynthesis is not beneficial in conditions where photorespiratory rates are minimal (Ehleringer et al., 1997). The exact contribution of different ecological drivers to the evolution of C_4 photosynthesis is unclear. However, a decrease in global atmospheric CO_2 concentration has been shown to correlate with the first evolutionary appearance of C_4 ~30 million years ago (Christin et al., 2008b; Vicentini et al., 2008), suggesting that CO_2 levels were a key selection pressure. The global expansion and ecological dominance of C_4 grasslands did not occur until ~6-8 million years ago (Cerling et al., 1997), however, which implies that other ecological factors may have promoted the spread of C_4 species to become dominant components of ecosystems. Despite the advantages C_4 plants have at higher temperatures, which is demonstrated by the dominance of C_4 grasslands in high temperature regions, temperature was not shown to correlate with C_4 grass expansion (Edwards and Smith, 2010). Instead, a reduction in mean annual rainfall was shown to correlate with 90% of the C_4 grass origins tested (Edwards and Smith, 2010). In addition, mammalian grazing pressures, increased instances of fire and increased seasonality of rainfall have been proposed to affect C_4 evolution, most likely as part of a network of factors that includes climate (Osborne, 2011).

1.2.3 Molecular evolution of C_4

All of the proteins required for the C_4 metabolism are present in C_3 species (Aubry et al., 2011). Therefore, instead of requiring networks of novel

components, C_4 evolution has depended on the co-option of existing metabolic (and likely developmental) genes. The observation that elements of the C_4 pathway function in carbon recycling in the stems of C_3 plants suggests that such co-option would not require many molecular changes (Hibberd and Quick, 2002). Gene duplication followed by neofunctionalisation has been proposed as a mechanism that would facilitate co-option by allowing novel functions to evolve in redundant gene pairs (Monson, 2003). This suggestion is supported by the observation that enzymes in the C_4 pathway are encoded by genes that often exist in multigene families (Hibberd and Covshoff, 2010) and that C_4 isoforms have undergone adaptive evolution (Wang et al., 2009). The discovery that a pyruvate transporter has been co-opted for C_4 function in the eudicots (Furumoto et al., 2011) suggests that this model of evolution extends to multiple components of the C_4 pathway.

Within the multigene families that encode many of the components of the C_4 pathway, phylogenetic analyses have shown that C_4 isoforms have been recruited repeatedly from the same gene lineage. For example, in the grasses C_4 *PEPC* and *NADP-ME* isoforms have been recruited independently from the same lineage within three and four member families respectively (Christin et al., 2007; Christin et al., 2010b). Additionally, presumably as a result of selection for C_4-optimised function (Gowik et al., 2006), parallel evolutionary changes have been identified at the amino acid level for a number of C_4 enzymes. In the grasses, C_4-specific amino acid signatures been identified in isoforms of *PEPC* (Christin et al., 2007), *NADP-ME* (Christin et al., 2010b), *PEPCK* (Christin et al., 2009b) and *RbcL* (Christin et al., 2008a). In the case

of *PEPC*, some of these C_4-specific amino acid signatures have also been identified in C_4 species outside the grasses (Blasing et al., 2000; Gowik et al., 2006; Christin et al., 2011).

In terms of the regulation of C_4 gene expression, few convergent *cis*-regulatory elements have been identified between independent C_4 lineages (Hibberd and Covshoff, 2010). Individual *cis*-elements have been identified that direct BS or M cell specific expression, however, these are not widely distributed in phylogenetic terms and in most cases sufficiency for cell specific expression has yet to be demonstrated. Notable exceptions include an upstream element in the *PPCA* gene of the C_4 species *Flaveria trinervia* that is necessary and sufficient for M cell transcript accumulation in C_3 and C_4 *Flaveria* species (Gowik et al., 2004; Akyildiz et al., 2007). 5' and 3' elements in the untranslated regions of *PPDK* and *CA* genes in the C_4 species *Cleome gynandra* have also been shown to be sufficient for M cell accumulation (Kajala et al., 2012). However, the same regions from orthologues of the C_3 species *Arabidopsis thaliana* also lead to M cell specific expression in *C. gynandra*, suggesting that changes in *trans*-factors have been important in determining cell specificity in *Cleome.* Remarkably, a region of *C. gynandra* and *Arabidopsis NAD-ME* gene coding sequence is sufficient to drive BS cell specific expression in *C. gynandra* but not *Arabidopsis* (Brown et al., 2011). This suggests that *trans*-factor divergence may be a widespread mechanism in C_4 evolution, and that a variety of *cis*-elements conserved between C_3 and C_4 species may exist for additional genes. Studies of the epigenetic regulation of C_4 genes have revealed limited convergence in different lineages. A study

of the promoter regions of *PEPC* and *NADP-ME* showed some convergence of histone modifications between independent C_4 grass lineages (Heimann et al., 2013), however, while other studies have shown epigenetic control of M (*PEPC*) and BS (*NADP-ME*) expression they have been limited to maize (e.g. Langdale et al., 1991; Offermann et al., 2006; Danker et al., 2008; Tolley et al., 2012b).

1.3 Kranz anatomy

1.3.1 Variations in Kranz form

Kranz is a crucial requirement for C_4 photosynthesis in almost all C_4 lineages and despite multiple evolutionary origins certain aspects of Kranz are shared in all cases. Namely, PsCA occurs in the outer layer of M cells and PsCR occurs in the inner layer of BS cells. Because the majority of M cells are in contact with BS cells there is a far lower ratio of M:BS cells in C_4 plants than in C_3, and BS cells have limited contact with intercellular airspaces. In most cases, including C_4 grasses, the concentric layers of Kranz tissue are arranged around closely spaced individual veins. In some eudicot lineages such as the Suaedeae (Chenopodiaceae), however, the concentric rings encircle all of the vascular tissue of the leaf (Kadereit et al., 2003; Schutze et al., 2003).

Presumably as a consequence of the distinct evolutionary origins in different phylogenetic and leaf contexts, variations to Kranz form exist. This has been recognised for some time (Brown, 1975) and has been extensively reviewed recently (Edwards and Voznesenskaya, 2011). In brief, according to Edwards and Voznesenskaya (2011), the key variable traits in Kranz form are 1) the number of BS layers; 2) the presence or absence of a mestome sheath (a sclerenchymatous cell layer that lacks chloroplasts and encircles vascular tissue (Esau, 1965)) and its position relative to BS layer(s); 3) the presence or absence of suberin lamella in BS cell walls (in grasses); 4) the position of BS chloroplasts (and mitochondria) and 5) the differentiation state of BS and M chloroplasts. Maize, which this thesis is particularly concerned with, exhibits a relatively simple Kranz form in that it possesses a single BS layer and no mestome sheath. BS cell walls are suberized and agranal BS chloroplasts are distributed centrifugally. This simplicity (particularly in terms of vascular sheath number), in addition to well developed genetic resources, makes maize an attractive model for C_4 development. Notable diversions from the maize form include grass and sedge lineages (e.g. *Aristida* and *Eleocharis*) that exhibit an inner Kranz BS cell layer separated from M cells by an additional parenchymatous BS layer (and possibly a mestome sheath). It is possible that the intermediate cell layers limit CO_2 loss following decarboxylation or function in the refixation of photorespired CO_2 (Voznesenskaya et al., 2005b). C_4 species that exhibit files of BS-like distinctive cells (DC), which occupy the position of minor veins, have also been identified in the grass genera *Arundinella*, *Arthraxon* and *Microstegium* (Crookston and Moss, 1973; Ueno, 1995). Despite not being associated with

vascular tissue, DC function as BS cells and act as sites of PsCR (Dengler et al., 1990; Dengler and Dengler, 1990; Wakayama et al., 2002; Wakayama et al., 2003).

Genetic and histological analyses of BS cell layer and DC differentiation have provided insights into the regulation of C_4 development. Depending on the developmental stage and vein type (i.e. sheath number, minor/major vein) analysed, histological studies have interpreted BS cells as derived from either procambial or ground meristem cells, and suggested that BS and M cells may share a clonal origin or arise from distinct cell lineages (Esau, 1943; Dengler et al., 1985; Nelson and Dengler, 1992; Bosabalidis et al., 1994). Given that procambial cells are derived from ground meristem tissue, these differences in interpretation are likely due to the developmental stage at which analyses were carried out. Importantly, a clonal analysis of spontaneous sectors in maize leaves demonstrated that the earliest stages of BS and M cell specification are determined by positional rather than lineage effects (Langdale et al., 1989). This is further supported by the appearance of BS-like DC without associated vascular tissue (Dengler et al., 1997). Later in development, BS and M cell differentiation is cell autonomous, as evidenced by work on the tangled mutant of maize, in which ectopic BS cells extend into the mesophyll (Jankovsky et al., 2001). Continued efforts to describe variation in Kranz form may provide further insights into the generality and modularity of Kranz evolution.

1.3.2 Kranz preconditioning

Recent work linking anatomy and phylogeny in the grasses has provided new insights into a preconditioning event that may have facilitated the repeated evolution of Kranz in certain grass lineages (Christin et al., 2013). Anatomical preconditioning towards Kranz, that is a shift in intervein distance, BS cell size and BS organelle composition, has long been proposed as a key precursor to C_4 evolution (Sage, 2001; McKown and Dengler, 2007; Sage et al., 2012). In the grasses there have been at least 17 independent origins of C_4 (Christin et al., 2008b), all of which have occurred within the PACMAD clade (Figure 1.4). This observation suggests that a preconditioning event may have taken place following the split of the PACMAD lineage from the rest of the grasses. In the eudicots, close vein spacing has been described in C_3 plants that are closely related to C_4 lineages in a number of genera (*Mollugo* (Christin et al., 2011), *Cleome* (Marshall et al., 2007; Voznesenskaya et al., 2007), *Flaveria* (McKown and Dengler, 2007), *Heliotropium* (Muhaidat et al., 2011) and *Euphorbia* (Sage et al., 2011b)) implying that close vein spacing is an important precondition in multiple phylogenetic contexts. Increased vein density has been suggested to confer a wide variety of benefits to plants (Sack and Scoffoni, 2013), particularly in hot and arid conditions (Sage et al., 2012), and therefore possibly represents an evolutionarily advantageous intermediary stage during C_4 evolution.

By measuring leaf anatomical traits for 157 grass species, including a wide variety of C_4 and C_3 species, Christin et al (2013) have shown that the repeated evolution of C_4 within the PACMAD clade is a likely consequence of an anatomical preadaptation related to reduced inter-BS cell distance (BSD).

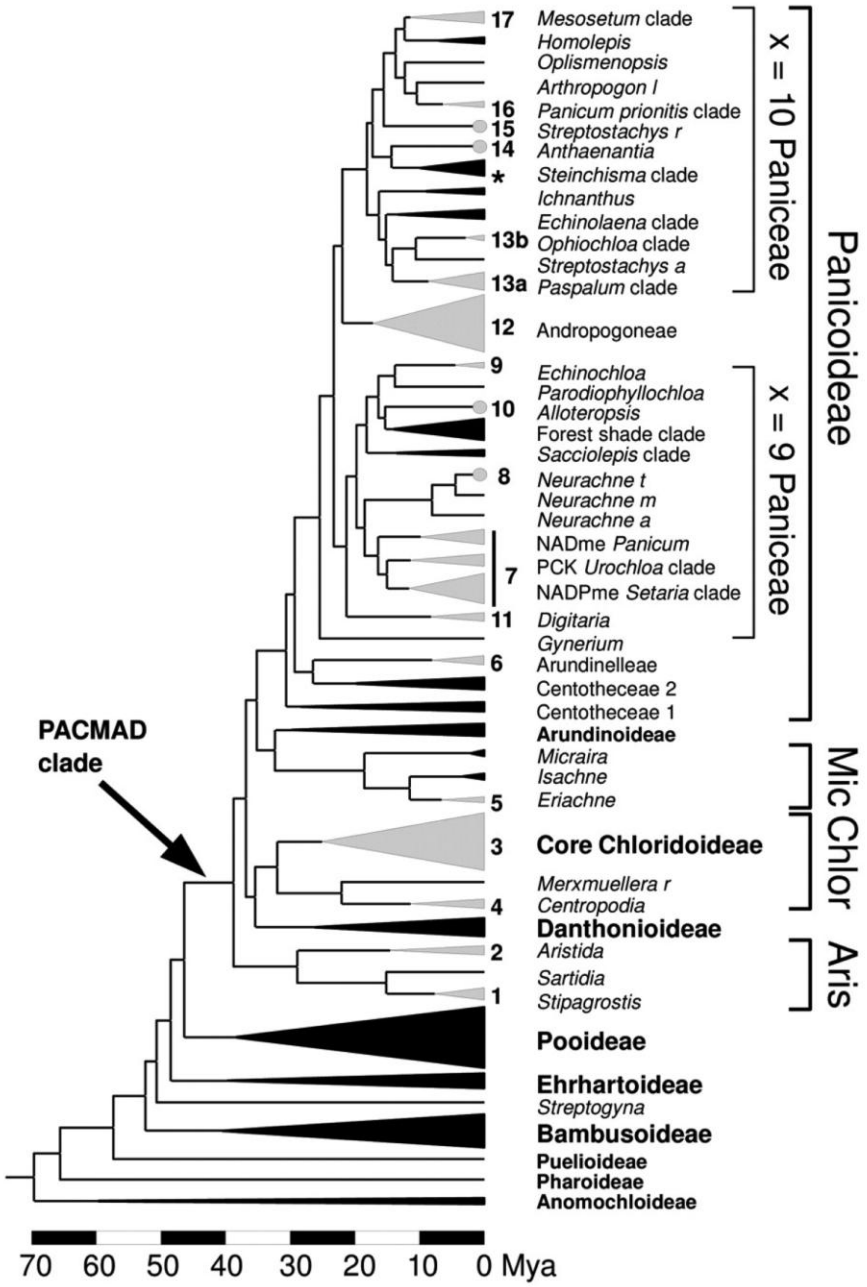


Figure 1.4. C_4 evolution within the grasses. 17 C_4 origins are highlighted in the PACMAD clade of grasses. Clades coloured grey consist solely of C_4 species, clade coloured black consist solely of C_3 species, circles represent C_4 lineages with a single species. Asterix indicates presence of C_3C_4 intermediate species *Steinchisma hians*. X indicates chromosome number in two Panicoideae tribes. Branch lengths are proportional to elapsed time (Mya – million years ago). Figure taken from Christin et al 2009a.

BSD is determined both by the distance between BS cells and by BS cell size. A decrease in BSD was originally found to predate the split between the PACMAD and BEP clade (which does not feature any C_4 species (Figure 1.4)), however, a subsequent reduction in BS cell size was found to occur within the BEP clade. The larger BS cell size that was maintained within the PACMAD clade presumably facilitated the evolution of C_4 when global CO_2 levels decreased following the PACMAD-BEP split (Christin et al., 2013). Within the PACMAD clade, the percentage of leaf area occupied by BS cells was found to be a key determinant in C_4 evolution, with the likelihood of C_4 occurrence associated with BS cell area above a certain threshold.

1.3 Systems biology and Kranz regulation

Experimental, technological and computational advances have ushered in an age of ‘big data’. Through the characterisation of entire networks, rather than individual components, systems approaches have enabled unparalleled insights into biological processes and revealed emergent properties of biological networks (Long et al., 2008). Approaches including transcriptomics, proteomics, interactomics, epigenomics and metabolomics are already widely used in plants (Moreno-Risueno et al., 2010; Bassel et al., 2012) and have been integrated, for example, to define co-expression networks and functional linkage for genes in *Arabidopsis* (Lee et al., 2010) and rice (Lee et al., 2011). In both cases characterisation of candidate genes identified by co-expression networks revealed expected functions.

Systems biology may be particularly useful in helping to identify the regulators of Kranz development, given that conventional means of gene identification have yielded minimal insight (Langdale, 2011). The limited results of forward genetic screens, and the complexity of Kranz development, suggest that a number of genes are involved in the regulation of Kranz development and that there is unlikely to be a Kranz 'master switch' (Westhoff and Gowik, 2010). The insights that systems approaches provide into the co-expression of genes, and therefore putative gene regulatory networks (Usadel et al., 2009), could thus provide the necessary breakthroughs. Recently, a number of systems based analyses have been carried out to identify components of C_4 regulatory networks. Many of these have relevance to the regulation of Kranz development (Table 1.1). These and previous advances will be discussed in the context of three general stages of Kranz development: 1) initiation of procambium, 2) BS and M cell-specification and 3) integration of the C_4 cycle. In addition to the studies shown in Table 1.1, other particularly notable studies of C_4 species using systems approaches have revealed enrichment of proteins in C_4 M cell chloroplast envelopes compared to C_3 (Brautigam et al., 2008), gene co-expression modules in different maize tissues (Downs et al., 2013), genome-wide transcription changes during maize meristem maturation and between meristem types (Takacs et al., 2012) and the transcript and protein content of photosynthetic cells in a single celled C_4 species (Park et al., 2010).

Author	Species	Tissue	Biological system	Methodology	Level of Kranz regulation
Liu et al (2013)	Maize	Pooled embryonic leaves	Transcriptome	Illumina sequencing	Procambium initiation BS and M cell specification
Li et al (2010)	Maize	Developmental gradient of an expanding foliar leaf (including BS and M cell isolates)	Transcriptome	Illumina sequencing	BS and M cell specification C_4 cycle integration
Pick et al (2011)	Maize	Developmental gradient of an expanding foliar leaf	Transcriptome Metabolome	Microarray GC-MS, LC-MS	BS and M cell specification C_4 cycle integration
Chang et al (2012)	Maize	BS and M cells from mid-section of an expanded leaf	Transcriptome	Illumina sequencing	BS and M cell specification C_4 cycle integration
Majeran et al (2010)	Maize	Developmental gradient of an expanding foliar leaf (including BS cell isolates)	Proteome	LC-MS, MS-MS	BS and M cell specification C_4 cycle integration
Friso et al (2010)	Maize	Leaf tip chloroplasts specific to BS and M cells	Proteome	LC-MS, MS-MS	C_4 cycle integration
Brautigam et al (2011)	*Cleome* (C_3 and C_4)	Fully mature leaves	Transcriptome	454 sequencing	C_4 cycle integration
Gowik et al (2011)	*Flaveria* (C_3, C_3-C_4, C_4)	Fully mature leaves	Transcriptome	454 sequencing	C_4 cycle integration
de Oliveira Dal'Molin et al (2010)	Maize Sugarcane Sorghum	Metabolic model of BS and M cells	Metabolome Proteome	Computational modelling	C_4 cycle integration
Heckmann et al (2013)	*Flaveria* (C_3, C_3-C_4, C_4) *Moricandia* (C_3-C_4) *Panicum* (C_3-C_4, C_4)	Metabolic model of C_4 evolution	Metabolome	Computational modelling	C_4 cycle integration

Table 1.1. Summary of previous systems based analyses of Kranz regulation.

1.3.1 Procambium initiation

Procambium initiation refers to the induction of procambial initials, which will give rise to vasculature, from ground meristem cells. As formation precedes the specification of BS and M cells, it can be thought of as the first stage of Kranz development. Leaf vasculature can be subdivided into hierarchical orders defined by the order of vein formation, i.e. lower order veins are first to form. In both monocots and eudicots the midvein forms first and extends acropetally into the leaf primordia (Nelson and Dengler, 1997). In most cases, midvein formation is followed by the acropetal extension of independent lateral veins (monocots) and secondary vein formation from the primary vein (eudicots). Higher order veins extend basipetally from the leaf tip (monocots) or in a basipetal reticulate wave (eudicots). As the increased vein density observed in both C_4 monocots and eudicots has been shown to be a consequence of increases in higher order vein numbers (Ueno et al., 2006; Mckown and Dengler, 2009), higher rates of procambial initiation following midvein and second order vein development are key to Kranz development. Searches for genetic regulators of higher order vein formation through mutagenesis screens have revealed very little, with the only mutant thus far identified (in the C_4 grass *Panicum maximum*) having been subsequently lost (Fladung, 1994). The addition of individual maize chromosomes to the C_3 species oat led to an increase in vein density, but the underlying genetic cause is unknown (Tolley et al., 2012a). As such, alternate methods of candidate identification may be necessary to determine the genetic regulators of procambial initiation.

Procambial initiation for all vein orders is known to follow gradients of auxin distribution, which is controlled by the localised expression of the auxin efflux PIN1 proteins (Scarpella et al., 2010). Polarised intracellular distribution of PIN1 proteins to cell membranes directs the path of auxin flow, leading to local maxima which precedes procambial specification. The direct regulation of the preprocambial marker *ATHB8* (Kang and Dengler, 2004; Scarpella et al., 2004) by the auxin response factor *MONOPTEROS* (*MP*), via an auxin response element in the *ATHB8* promoter, is further evidence for a role of auxin in vascular patterning (Donner et al., 2009). In addition to *ATHB8*, the expression domains of the GRAS transcription factor *SHORT ROOT* (*SHR*) and *DNA BINDING WITH ONE ZINC FINGER* (*DOF*) genes have also been shown to predict patterns of procambial initiation in *Arabidopsis* (Gardiner et al., 2010; Gardiner et al., 2011).

As the earliest stages of vascular patterning take place within young leaf primordia, the isolation of appropriate samples for analyses of gene expression is technically difficult. This is reflected in the paucity of systems data that provides insight into the processes of procambial initiation (Table 1.1). A transcriptome analysis of germinating maize seed at different time points over 72 hours includes leaf primordia that are undergoing Kranz differentiation, and suggests a time point at which vascular differentiation is occurring based on the expression of maize orthologues of *MP* and other *Arabidopsis* genes involved in vascular patterning (Liu et al., 2013). However, embryonic leaves of different states of development were pooled in this

analysis, making it difficult to determine the relative contribution of distinct leaf developmental states to the identified transcriptomic signatures.

1.3.2 BS and M cell-specification

As with procambial specification, mutagenesis screens have revealed little about BS and M cell-specification, with insights limited to chloroplast development in specific cell types. Two mutations, *bundle sheath defective2* (*bsd2*) and *high chlorophyll fluorescence136* (*hcf136*), revealed the role of a chaperone protein (BSD2) in Rubisco stabilisation in BS cells (Brutnell et al., 1999) and a thylakoid localised protein (HCF136) in the stabilisation of photosystem II in M cells (Covshoff et al., 2008). A third mutation identified the *golden2* gene, which encodes a transcription factor that specifically regulates BS chloroplast formation in maize. Intriguingly, transcripts of the paralogue *GOLDEN2-LIKE 1* (*GLK1*) accumulate in M cells, where *ZmGLK1* is proposed to have a role in M chloroplast regulation (Hall et al., 1998; Rossini et al., 2001). In C_3 species, *GLK* gene pairs have been shown to function redundantly in photosynthetic development in all M cells (Rossini et al., 2001; Fitter et al., 2002; Yasumura et al., 2005). Compartmentalisation of *GLK* gene expression, therefore, may be an important determinant of C_4 evolution.

As maize leaves exhibit a developmental gradient from tip (most mature, photosynthetic source) to base (least mature, photosynthetic sink), systems analyses of different stages along this gradient have provided insights into the processes that regulate a wide variety of maturation programmes (Nelson,

2011). In terms of BS and M cell-specification, anatomical analyses have shown that rudimentary Kranz anatomy has already formed at the base of the third leaf of nine day old maize seedlings (Majeran et al., 2010). However, at this stage BS and M cells remain relatively small and undifferentiated. Therefore, while genetic regulators involved in the earliest stages of cell type specification may no longer be expressed at the leaf base, it is likely that both transcripts and proteins involved in BS and M cell elaboration will be active in this region. Analyses of transcriptomes (Li et al., 2010) and proteomes (Majeran et al., 2010) sampled along this gradient, therefore, may reveal important components of BS and M cell elaboration, particularly in terms of cell wall deposition and plastid biogenesis/maturation. The appearance of *G2* transcripts and proteins after plastid biogenesis has ceased suggests that chloroplast specialisation is subsequent to biogenesis (Li et al., 2010; Majeran et al., 2010). A transcriptome study along a similar leaf gradient is also likely to be informative in this regard (Pick et al., 2011). However, in this study although the third leaf was also used, it was sampled when slightly older (~4cm larger). A lack of anatomical data from the leaf base sample means there is no information about the differentiation state of BS and M cells, although the similarity in transcriptomic trends between the different leaf types (Li et al., 2010; Pick et al., 2011) suggests similar differentiation states across the leaves.

The isolation of BS specific proteins from distinct stages along the leaf gradient provides an additional layer of information about cell specific processes during BS differentiation (Majeran et al., 2010). In particular,

proteins involved in chloroplast biogenesis were found to be both under and over-represented in BS samples, suggesting cell-specific pathways of plastid formation. BS and M cell transcriptomes for cells isolated from the tip of the leaf gradient (Li et al., 2010) and from the middle of an expanded second leaf on a nine day old plant (Chang et al., 2012) have also been described. The age of these samples suggest they may be less informative about cellular differentiation. Transcripts involved in BS and M cell-specification will be present in the gene expression data collected on pooled embryonic leaves (Liu et al., 2013), however, the pooled nature of these samples means that they will be difficult to identify.

1.3.3 C_4 cycle integration

The C_4 pathway is known to be dependent on both environmental and developmental signals. In maize and the eudicot amaranth, light and/or developmental cues are required to limit Rubisco accumulation to BS cells (Langdale et al., 1988b; Wang et al., 1993). Observations in maize have led to the proposal that cells respond to a C_4 inducing signal that emanates from veins, and that cells need to be within a two cell radius of vascular tissue for enzymes of the C_4 pathway to accumulate correctly (Langdale and Nelson, 1991). Such a signal is yet to be identified, and few *trans*-factors are known to induce cell-specific gene expression. A notable example is the maize DOF1 protein that regulates the cell-specific expression of *PEPC* (Yanagisawa and Sheen, 1998), however, it has also been shown to act more generally in the transcriptional regulation of a wide range of genes (Yanagisawa, 2000).

To date systems approaches may be most useful in identifying putative regulators of the C_4 pathway and unknown pathway components. Transcriptomic (Li et al., 2010; Pick et al., 2011), proteomic (Majeran et al., 2010) and metabolomics analyses (Pick et al., 2011) along a maize leaf developmental gradient have all shown that the C_4 pathway does not function at the base of the leaves used for analysis, and that the pathway is initiated distally along the lamina. The appearance of transcripts and proteins of genes that function in the C_4 pathway at similar points along the leaf (Majeran et al., 2010; Pick et al., 2011) suggest that coexpression analyses will reveal novel pathway regulators and components. BS and M cell transcriptomes from photosynthetic source tissue are likely to include genes involved in the active maintenance of the C_4 pathway within the two cell types (Li et al., 2010; Chang et al., 2012). In this regard, BS/M transcriptomes and proteomic data from BS and M chloroplast membranes have both identified putative novel chloroplast transporters (Friso et al., 2010; Chang et al., 2012). Similarly, comparisons of mature leaf transcriptomes from C_4 and C_3 *Cleome* (Brautigam et al., 2011) and *Flaveria* (Gowik et al., 2011) species both identified a sodium symporter that is upregulated in C_4 species. This has subsequently been shown to be important for pyruvate transport in the C_4 pathway (Furumoto et al., 2011), highlighting the power of transcriptomic analyses for candidate gene identification. Both *Cleome* and *Flaveria* analyses have also shown that plastidic sigma factors and individual *GLK* genes are upregulated in C_4 species (Brautigam et al., 2011; Gowik et al., 2011).

In terms of the evolutionary changes required for C_4 development, 603 transcripts were found to be differentially expressed between C_3 and C_4 species in *Cleome* (Brautigam et al., 2011), while an upper limit of 3582 changes in gene expression has been proposed for the evolution of C_4 in *Flaveria* (Gowik et al., 2011). 1213 expression changes were identified as shared between two C_4 and a C_3-C_4 intermediate *Flaveria* species, but not two C_3 species, although a greater number of expression changes were C_4 species specific (Gowik et al., 2011). This suggests a gradual trajectory of C_4 evolution, which is consistent with the continuous pattern of transcriptome changes observed along a maize leaf gradient (Pick et al., 2011). *In silico* modelling of C_4 evolution, validated by biochemical data from *Flaveria, Morcandia* and *Panicum* species, has suggested that intermediate steps in C_4 photosynthesis, as typified by *Flaveria* C_3-C_4 intermediates, are adaptive in their own right (Heckmann et al., 2013). This would imply that C_3-C_4 intermediates are true intermediates, and not on distinct evolutionary trajectories. This study integrated a previous genome wide model of C_4 grass metabolism that for the first time allows predictions to be made about changes in metabolic flux between BS and M cells (de Oliveira Dal'Molin et al., 2010). As systems guided experiments provide more data on the regulation of the C_4 pathway in different contexts, models such as this will be invaluable in interpreting the results at the level of leaf metabolism.

1.4 The *SHORT-ROOT* pathway

A model has been proposed whereby C_4 BS and M cell type specification is regulated by an inducing signal that emanates from veins (Langdale and Nelson, 1991). This model proposes that, through interaction with cell autonomous specification factors, high levels of the signal induce BS type C_4 development and reduced levels induce M type C_4 development. Since the proposal of this model, however, no C_4 inducing signal has been identified.

A system where a developmental signal is expressed in the vasculature and migrates outwards to determine cell types is known to exist in angiosperm roots. In *Arabidopsis* the GRAS transcription factor *SHORT-ROOT* (*SHR*) is expressed in the root stele, however, SHR protein migrates to adjacent endodermal tissue where it is bound by *SCARECROW* (*SCR*), another GRAS transcription factor. SCR sequesters SHR in the nuclei of the endodermis, which prevents further SHR migration, and leads to upregulation of *SCR* expression, reinforcing the cycle (Helariutta et al., 2000; Nakajima et al., 2001; Heidstra et al., 2004; Cui et al., 2007). Loss of function *shr* and *scr* mutants display a reduced number of root ground tissue layers (Benfey et al., 1993; Scheres et al., 1995; Di Laurenzio et al., 1996; Helariutta et al., 2000), while ectopic *SHR* expression in the root leads to supernumerary cell layers with endodermal identity (Nakajima et al., 2001), suggesting that *SHR* is necessary for both cell division and endodermal specification in the root. Following the specification of endodermal and cortical layers of ground tissue, the function of SHR in the regulation of endodermal cell division to form a

middle cortex later in development has since been shown to be dose dependent, with moderate decreases in SHR leading to periclinal cell divisions (Koizumi et al., 2012). A role for *SHR* in the regulation of vascular patterning has also been shown through autonomous stele activity (Levesque et al., 2006) and through regulation of a non-cell autonomous microRNA pathway in the endodermis (Carlsbecker et al., 2010). The *SHR*/*SCR* regulatory network is also comprised of *INDETERMINATE 1 (ID1)* (Colasanti et al., 1998) like genes such as *MAGPIE* (*MGP*), *NUTCRACKER* (*NUT*) and *JACKDAW* (*JKD*). *MGP* and *NUT* have been identified as direct targets of *SHR* and *SCR* activity (Levesque et al., 2006; Cui et al., 2007) while *JKD* is required for the correct regulation of *SCR* expression and loss of function *jkd* mutants display ectopic cell layers (Welch et al., 2007). A schematic of the roles of *SHR*, *SCR* and *JKD* in root radial patterning is shown in Figure 1.5.

The first suggestion that the *SHR*/*SCR* regulatory module may also regulate development in the above ground organs of plants came from the observation that *Arabidopsis shr* and *scr* mutants lack the endodermis layer, or starch sheath, that surrounds the stem (Fukaki et al., 1998). A role in leaf development was subsequently demonstrated by the observation that *SCR* is expressed in *Arabidopsis* BS cells, and that the petioles of *scr* cotyledons lack a layer of ground tissue (Wysocka-Diller et al., 2000). *SHR* was also shown to be expressed in a pattern that predicts future vein formation in *Arabidopsis* leaves (Gardiner et al., 2011) while the maize *SCR* orthologue *ZmSCR* was shown to be expressed in leaf vasculature (Lim et al., 2005). Notably, it has been suggested that the root endodermis, the starch sheath and BS cell layer

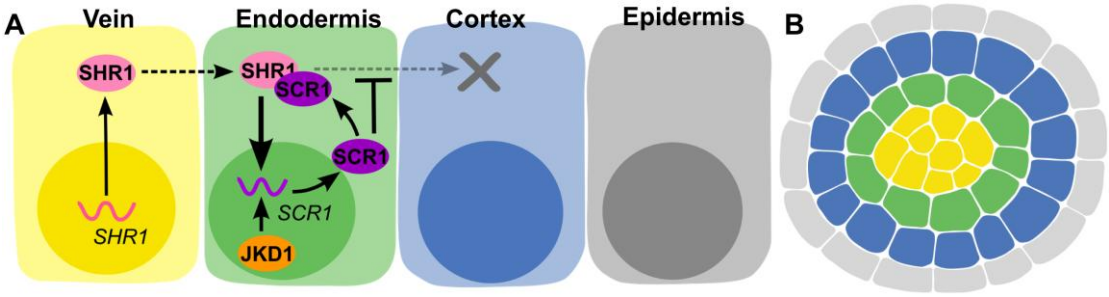


Figure 1.5. Schematic of SHORT-ROOT regulation of root radial patterning. **A)** SHR regulatory networks. Arrows imply promotion of gene expression, lines with 'T' ends imply inhibition of protein movement, dashed lines imply protein movement, waved lines represent mRNA, coloured ovals represent proteins. **B)** Transverse section of root schematic showing cell layers at in (A). Yellow depicts vasculature, green depicts endodermis, blue depicts cortex and grey depicts epidermis. Figure produced by Sayuri Ando.

are equivalent tissues (Esau, 1965) and as such, the *SHR*/*SCR* regulatory module may function in the developmental patterning of all three tissues (Slewinski, 2013). The observation that *Zmscr* mutants exhibit perturbed Kranz patterning further suggests that the *SHR/SCR* network functions in C_4 development (Slewinski et al., 2012). Considering that elements of the C_4 pathway have been shown to operate in the starch sheath of C_3 species (Hibberd and Quick, 2002), a common regulatory module that could explain development in the root, BS and starch sheath has intriguing consequences for the evolution of C_4 photosynthesis.

1.5 C_4 rice development

As a consequence of the increased photosynthetic, water use and nitrogen use efficiencies that C_4 species exhibit (Zhu et al., 2008; Ghannoum et al., 2011), there are currently large international efforts to convert C_3 crop species, such as rice, to C_4 (Hibberd et al., 2008). Rice is the principal source of calories in the developing world but, with increases in yield plateauing, there is likely to be a shortage of rice as the global population continues to increase (Dawe, 2007). A C_4 rice variety could help to meet the nutritional needs of a growing population, however, it will require the introduction of a suite of changes. These would include cell-specific expression of C_4 pathway enzymes, altered enzyme kinetics, reduced vein spacing, increased BS cell size, reduced BS CO_2 conductivity, increased plasmodesmata number, increased transporters and possibly chloroplast localisation (Sheehy et al.,

2007). Alternatively, a single celled model of C_4 could be introduced, which would require fewer anatomical changes (Miyao et al., 2011). However, as the most productive C_4 species exhibit Kranz, and this has already evolved multiple times within the grasses, Kranz-type C_4 is perhaps a preferable paradigm for rice (Kajala et al., 2011). If this is to be successful, a fuller understanding of the genetic regulation of Kranz development will be essential. The paucity of knowledge of Kranz development, and the extent of the changes required for C_4 rice development, make this a daunting project. However, the observations that C_4 has evolved on over 60 occasions independently (Sage et al., 2011a), and that the rice genus display certain Kranz preadaptations (Christin et al., 2013), suggest that the project is feasible.

1.6 Thesis aims

Considering the argument for C_4 rice development outlined above, there are strong practical reasons for working on Kranz development. In addition, insights into Kranz development will elucidate the mechanisms behind one of the most widely dispersed but poorly understood examples of convergent evolution. As such, the work described in this thesis aims to identify novel components of Kranz regulation and to test the generality of *GLK* gene function in C_4 evolution. To identify novel components of Kranz development, a transcriptomic dataset of maize leaf development was analysed in detail. Chapter 3 describes work carried out to validate the RNA-sequencing data used for this transcriptomic analysis. Following the validation of the

sequencing data, Chapter 4 details bioinformatics analyses that were designed to identify candidate Kranz regulators from within this dataset. Chapter 5 describes the functional characterisation of three of these candidates. Chapter 6 describes phylogenetic and gene expression analyses to determine whether *GLK* genes may function in C_4 development in lineages aside from maize. Finally, in Chapter 7, using insights from the literature and transcriptome data from the current and an additional published study, a genetic model for Kranz patterning is proposed.

Chapter 2: Materials and Methods

2.1 General materials

Solutions were made using de-ionised water that was purified using MilliQ (Millipore, www.millipore.com) cartridges. Where necessary sterilisation was carried out by autoclaving at 120^{o}C for 15 min (at 100kPa) or through microfiltration using a 0.2μm syringe filter (VWR, www.uk.vwr.com). Unless otherwise mentioned, chemicals were purchased from Sigma-Aldrich (www.sigmaaldrich.com), Fisher Scientific (www.fisher.co.uk) or Melford (www.melford.com). Synthetic oligonucleotides were purchased from Eurogentec (www.eurogentec.com) and Sigma-Aldrich. Radioactive compounds were purchased from Perkin-Elmer (www.perkinelmer.com).

2.2 Plant material and growth conditions

2.2.1 *Sorghum bicolor*

Sorghum bicolor L. Moench inbred line BTx623 (USDA, www.ars-grin.gov) was used to test *GOLDEN2-LIKE* (*GLK*) gene expression in bundle sheath (BS) and mesophyll (M) cells. Sorghum plants were grown in soil (potting compost number 2 or 3, John Innes Manufacturers Association,

www.johninnes.info) in a greenhouse, with the natural diurnal light period in Oxford, and were supplemented with 500μmol photon m^{-2} s^{-1} when necessary, and up to 14 h in winter. The average daytime temperature was 28°C and the average nighttime temperature was 20°C.

2.2.2 *Zea mays*

Zea mays inbred line B73 (USDA, www.ars-grin.gov) was used as a source of RNA for the transcriptomic study analysed in this thesis as well as a source of genomic DNA for molecular cloning and transformation. Maize lines that were identified with *Ds* insertions in genes of interest were derived from W22-inbred lines and obtained from Dr Tom Brutnell's lab at Cornell University (NY, USA). All maize lines were grown as described for *Sorghum bicolor*.

2.2.3 *Oryza sativa*

Oryza sativa cv. Kitaake seed were obtained from the International Rice Research Institute (IRRI, Philippines, www.irri.org) and used for the generation of transgenic rice lines. Rice plants were grown in saturated soil conditions (John Innes number 2 or 3) and under the same light and temperature regime as *Sorghum bicolor*.

2.2.4 *Setaria viridis*

Setaria viridis (accession A10) seed were obtained from Dr Tom Brutnell's lab at Cornell University (Ithaca, NY, USA). Following seed set, subsequent generations of seed required storage for at least 1 month in desiccating conditions prior to germination. Plants were grown as described for *Sorghum bicolor*.

2.2.5 *Arabidopsis thaliana*

Arabidopsis thaliana lines identified with T-DNA insertions in homologues of candidate maize Kranz regulators were obtained from the European Arabidopsis Stock Centre (NASC, UK, www.arabidopsis.info). The lines obtained form part of the GABI-Kat, SALK and SAIL T-DNA insertion collections, as described on the NASC website, which were all generated in the Columbia ecotype. Arabidopsis plants were grown in peat plugs (Jiffy Products International, www.jiffygroup.com) in a greenhouse, with the natural diurnal light period in Oxford, and were supplemented with 500μmol photon m^{-2} s^{-1} when necessary, and up to 14 h in winter. The average daytime temperature was 25 °C and the average nighttime temperature was 20 °C.

2.2.6 Bacterial growth conditions

Escherichia coli strains XL1-Blue MRF', DB3.1 and DH5α were used for plasmid manipulation. *Agrobacterium tumefaciens* strains EHA105 and AGL1 were used for transformation of rice and *Setaria* respectively. Cultures were

grown on Luria Bertrani (LB) medium (1% tryptone, 0.5% yeast, 0.5% NaCl (plus 1.5% agar for solid media)). *E. coli* was cultured at 37^{o}C overnight while *A. tumefaciens* was grown at 28^{o}C for 2 days, either shaking at 200rpm in liquid cultures or in a static incubator for cultures grown on plates. Where appropriate, bacterial cultures were screened with antibiotics at the following concentrations:

Ampicillin	100μg ml^{-1}
Gentamycin	12.5μg ml^{-1}
Kanamycin	50μg ml^{-1}
Rifampycin	50μg ml^{-1}
Spectinomycin	50μg ml^{-1}

2.3 Molecular cloning

2.3.1 Polymerase Chain Reaction (PCR)

Primers 18-26 nucleotides in length were designed using Primer3 software (Untergasser et al., 2012). The sequence information for all primers can be found in Appendix 2.1. For standard reactions, GoTaq Polymerase (Promega, www.promega.com) was used to amplify DNA. Reactions were assembled on ice and consisted of 1X GoTaq Flexi Buffer, 1-4mM $MgCl_2$, 0.2mM of each dNTP, 0.1-1μM of each primer, 1.25 unit/50μl enzyme and 0.05-0.5μg/50μl

DNA. Where necessary PCR enhancers such as DMSO (4-8%) or betaine (1-1.6M) were added to denature secondary structures. Amplification programmes were carried out in thermal cyclers (Applied Biosystems, www.appliedbiosystems.com) as below:

Denaturation	1 cycle	95°C for 2 min	
Denaturation	1 cycle	95°C for 2 min	
Annealing	1 cycle	50°C-68°C for 30 sec	} x25-40 cycles
Extension	1 cycle	72°C for 0.5-3 min	
Extension	1 cycle	72°C for 5 min	

The annealing temperature was set to 3°C lower than the primer Tm and extension carried out at a rate of one minute per kb. Fragments used for transformation were amplified using Phusion polymerase (Thermo Scientific, www.thermoscientific.com) and the necessary adjustments to the amplification programme made as described in the manufacturer's guidelines. PCR products were separated by gel electrophoresis (tanks and power supplies – Bio-Rad, www.bio-rad.com) across 0.8-2% agarose gels and visualised using 0.1% ethidium bromide and a Biospectrum Image System with Vision Works software (UVP, www.uvp.com).

2.3.2 Plasmid recombination

DNA fragments amplified by PCR were purified using the Wizard SV Gel Clean Up kit (Promega, www.promega.com) as per the manufacturer's instructions. If a fragment was designed to be expressed *in planta* it was cloned using Gateway technology (Invitrogen, www.invitrogen.com) in accordance with the manufacturer's guidelines. pDONR201 and pDONR207 were used to generate entry clones for fragments designed for overexpression in rice and RNAi in *Setaria* respectively. A Gateway adapted version of pVEC8 (Wang et al., 1998) (supplied by Dr. Chulmin Kim, University of Oxford) was used as the destination vector for overexpression in rice and pANDA (Miki and Shimamoto, 2004) as a destination vector for RNAi in *Setaria*.

Fragments used to generate probes for Southern or northern blot analyses were cloned using the CloneJET cloning kit (Thermo Scientific, www.thermoscientific.com) as per the manufacturer's instructions.

2.3.3 DNA sequencing

To confirm PCR fragment identity, sequencing reactions were carried out using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, www.appliedbiosystems.com) as described in the manufacturer's guidelines. The DNA was then precipitated using 2.5x volumes of ethanol and 0.08M NaOAc on ice before being analysed using an ABI prism 3100 capillary sequencer (Applied Biosystems).

2.3.4 Bacterial transformation (*E.coli* and A. *tumefaciens*)

Constructs were electroporated into *E. coli* XL1-Blue MRF' and *A. tumefaciencs* strains, and transformed through heat shock into *E. coli* strains DB3.1 and DH5α. For electroporation, 1-5μl of binary vector was added to recently thawed 40μl electrocompetent cells on ice. Cuvettes of DNA/cell mixtures were then subjected to a pulse of electricity using a Gene Pulser II (Bio-Rad, www.bio-rad.com) set at 25μF capacitance, 2.5kV voltage and 200Ω (*E. coli*) or 600Ω (*A. tumefaciens*) resistance. 1ml of SOC medium (2% tryptone, 0.5% yeast, 10mM NaCl, 0.36% glucose, 2.5mM KCl and 10mM $MgCl_2$) for *E. coli*, or 1ml LB medium for *A. tumefaciens*, was immediately added to the cuvettes following the electric pulse and the cuvettes left to recover at 37°C (*E. coli*) or 28°C (*A. tumefaciens*) shaking at 200rpm for 1-4 hours. The cell suspension was then plated onto LB media with the appropriate antibiotic(s).

Heat shock transformation was carried out by adding 20-100ng of DNA to 250μl thawed, heat shock competent cells on ice. After 30 minutes the DNA/cell mix was placed in a heat block at 42°C for 40 seconds before being returned to ice for 30 seconds. 1ml SOC medium was then added and the tubes were left to recover at 37°C without shaking for 1 hour, before being plated on LB selection media.

Following transformation, plasmid purification was carried out using the QIAprep spin miniprep kit (Qiagen, www.qiagen.com).

2.4 Plant transformation

2.4.1 Tissue culture media

R1: 4.3g/L Murashige and Skoog (MS) salts and vitamins (Duchefa Biochemie, www.duchefa-biochemie.nl), 30g/L sucrose, 0.5g/L MES, 300mg/L Casamino acid, 2mg/L 2,4-D, 4g/L Phytagel, pH 5.8

S1: 4.3g/L MS salts and vitamins, 30g/L sucrose, 0.5g/L MES, 2mg/L 2,4-D, 0.2mg/L kinetin, 0.6mg/L $CuSO_4$, 4g/L Phytagel, pH 5.8

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 5.2

RS2: 4.3g/L MS salts and vitamins, 30g/L sucrose, 0.5g/L MES, 10g/L glucose, 300mg/L Casamino acid, 2mg/L 2,4-D, 4g/L Phytagel, pH 5.2 – 20mg/L acetosyringone was added after autoclaving

R3: 4.3g/L MS salts and vitamins, 30g/L sucrose, 0.5g/L MES, 300mg/L Casamino acid, 2mg/L 2,4-D, 4g/L Phytagel, pH 5.8 – 200mg/L timentin and 30mg/L hygromycin was added after autoclaving

S3: 4.3g/L MS salts and vitamins, 30g/L sucrose, 0.5g/L MES, 2mg/L 2,4-D, 0.2mg/L kinetin, 0.6mg/L $CuSO_4$, 4g/L Phytagel, pH 5.8 – 200mg/L timentin and 40mg/L hygromycin was added 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 – 200mg/L timentin and 30mg/L hygromycin was added after autoclaving

S4: 4.3g/L MS salts and vitamins, 30g/L sucrose, 0.5g/L MES, 2mg/L kinetin, 2g/L Phytagel, pH 5.8 – 200mg/L timentin and 40mg/L hygromycin was added after autoclaving

RS5: 2.15g/L MS salts and vitamins, 15g/L sucrose, 0.5g/L MES, 2g/L Phytagel, pH 5.8 – 200mg/L timentin was added after autoclaving

2.4.2 Rice and *Setaria* callus induction

Seed coats were removed from Kitaake rice and *Setaria viridis* seed before being washed in 70% ethanol for 2 minutes followed by either a 15 minute wash in 5% NaClO (rice) or a 10 minute wash in 0.625% NaClO (*Setaria*). Seeds were then washed five times in sterile distilled water before being plated on R1 (rice) or S1 (*Setaria*) medium. Plates were kept in the dark at 25^{o}C and dividing calli separated from seeds after 2 weeks (rice) or 4 weeks (*Setaria*). Calli were subcultured on fresh R1 (rice) or S1 (*Setaria*) media

every 2 weeks until being used for transformation (within 3 months following callus induction).

2.4.3 *Agrobacterium* mediated transformation

Prior to transformation, *Agrobacterium* stocks containing the appropriate binary vector were plated on LB media and grown for 2 days at 28^{o}C. 3ml of sterile AAM medium that had been supplemented with 40mg/L acetosyringone was inoculated from the cultures grown on LB plates to an OD600 of approximately 0.5. Calli were immersed in these *Agrobacterium* suspensions for approximately 5 minutes before being blotted dry on sterilised filter paper. Dried calli were then placed on additional sterillised filter paper on RS2 medium and left to recover for 3 days at 25^{o}C.

2.4.4 Transformed callus selection

After 3 days on RS2 media, calli were transferred to R3 (rice) or S3 (*Setaria*) plates for selection and kept at 25^{o}C in the dark for 2 weeks. After this time, proliferating calli were transferred to fresh R3 or S3 media and dead calli were removed. After another 2 weeks (4 weeks total on R3 or S3 plates), living calli were transferred to R4 (rice) or S4 (*Setaria*) plates and grown in cabinets under a light intensity of 700μmol photon m^{-2} s^{-1} at 28^{o}C for regeneration.

2.4.5 Transformed callus regeneration

Proliferating calli grown on R4 or S4 plates were transferred to fresh R4 or S4 plates every two weeks until shootlets arose. Following this they were then transferred to RS5 plates or tubes to encourage root growth, still at 28^{o}C in the light. After root emergence (>3cm main root growth) plants were taken to the greenhouse and exposed to the air for acclimatisation for 3-5 days before being transferred to soil and grown as described in section 2.2.

2.5 Plant DNA extraction and screening

2.5.1 Tissue harvesting

Leaf tissue was harvested using forceps and stored at -80^{o}C. Frozen tissue was homogenised either by manual grinding using a pestle and mortar in liquid nitrogen or by being ground mechanically in 2ml tubes with ball bearings in a Retsch Mixer Mill 300 (Retsch, www.retsch.com).

2.5.2 CTAB method for DNA extraction

650μl of 1.5x CTAB extraction buffer (1.5% CTAB, 75mM Tris-HCl pH 8.0, 15mM EDTA pH 8.0, 1.05M NaCl) was added to ground tissue and tubes were incubated at 65^{o}C for 20-30 minutes with occasional vortexing. 650μl of

Chloroform:Isoamyl Alcohol (24:1) was added and, after vortexing, tubes were spun at 12,000rpm for 5 minutes. Supernatants were subsequently transferred to fresh tubes and DNA precipitated by adding 1ml 100% ethanol. After another centrifugation step of 12,000rpm for 5 minutes, the resultant pellets were washed in 70% ethanol and again spun at 12,000rpm for 5 minutes. The ethanol was then poured off and after being air dried, pellets were resuspended in 50-100μl TE buffer (10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0). The concentration of resuspended DNA was quantified using a ND-1000 NanoDrop spectrophotometer (Thermo Scientific, www.thermoscientific.com).

2.5.3 PCR

To initially verify the transgenic status of plants, PCR was carried out as described in section 2.3 to amplify gene specific components of the binary construct, the hygromycin resistance marker, construct promoter or terminator elements or a combination of these elements.

2.5.4 Southern blot analysis

10μg of sample DNA was normally digested with 3μl of the appropriate enzyme in a 30μl reaction mix for a minimum of 4 hours. Digested DNA was then electrophoresed across a 0.8-2% agarose gel at 15v overnight and photographed as described in section 2.3. The DNA was subsequently

denatured by treating the gel in denaturing buffer (1.5M NaCl, 0.5M NaOH) for 1 hour and then neutralised by treating in neutralising buffer (3M NaCl, 0.5M Tris-HCl pH 7.5) for 1 hour. DNA was then transferred to a Nytran nylon membrane (Flowgen Bioscience, www.flowgen.co.uk) using a standard Southern blotting technique, with 20x SSC (3M NaCl, 0.3M tri-sodium citrate, pH 7.6) as the blotting agent. After blotting overnight, the membrane was rinsed in 2x SSC (0.3M NaCl, 30mM tri-sodium citrate, pH 7.6) for 5 minutes and the DNA UV cross-linked to the membrane using a Stratalinker 1800 (Stratagene, www.genomics.agilent.com).

DNA probes were radiolabelled as described previously (Feinberg and Vogelstein, 1983) (25μl radiolabeling reaction: 40ng DNA, 0.05M Tris-HCl pH 8.0, 5mM $MgCl_2$, 0.2mM dATP, 0.2mM dTTP, 0.2mM dGTP, 0.0007% β-mercaptoethnanol, 0.2M hydroxyethylpiperazine ethanesulphonic acid (HEPES), 0.01x hexadeoxyribonucleatides (in TE at 90 OD units ml^{-1}), 1μl 10mg ml^{-1} BSA, 1μl Klenow polymerase, 2.5μCi ^{32}P-dCTP) and blots hybridised also as described previously (Langdale et al., 1991). Blots were then exposed to a Fuji imaging plate (Fuji Photo Film, www.fujifilm.com) before being imaged using a Bio-Rad FX Molecular Imager and Quantity One Software (Bio-Rad, www.bio-rad.com).

2.6 Manipulation of RNA

2.6.1 RNA extraction

For the protocol below, all solutions were made with 0.1% diethylpyrocarbonate (DEPC) dH_2O and, unless specified otherwise, the samples were kept on ice. Filter pipette tips were also used throughout.

Tissue used for RNA extraction was first ground as described in section 2.5. Ground tissue was added to eppendorf tubes containing 500μl TRIzol reagent (38% phenol, 0.8M guanidine thiocyanate, 0.4M ammonium thiocyanate, 0.1M sodium acetate pH 5.0, 5% glycerol) and spun at 12,000rpm for 5 minutes. The supernatants were transferred to new tubes and left at room temperature for 5 minutes. 200μl $CHCl_3$ was added and, after vortexing, again left at room temperature for 5 minutes. Samples were subsequently spun at 12,000rpm for 15 minutes at 4^{o}C and the supernatants added to 500μl isopropanol. After being left for 10 minutes at room temperature, tubes were inverted and then spun at 12,000rpm for 10 minutes at 4^{o}C. Supernatants were poured away and the pellets washed in 1ml 75% ethanol. After being spun at 7,500rpm for 5 minutes at 4^{o}C the pellets were air dried and resuspended in 50-100μl DEPC dH_2O.

2.6.2 Reverse transcription-PCR

cDNA was produced using either SuperScript II or SuperScript III Reverse Transcriptases (both Invitrogen, www.invitrogen.com) as per the

manufacturers guidelines (oligo(dT) primers were used rather than random or gene specific primers)).

2.6.3 RNA formaldehyde gels

Prior to use for RNA gels, electrophoresis tanks and trays were treated with 1M NaOH for 30 minutes to remove any RNAses and rinsed with dH_2O. A 1.5% agarose formaldehyde gel was made by combining 17ml formaldehyde (37% solution) with 33ml dH_2O and adding this to 1.5g agarose fully dissolved in heated 10ml 10x gel buffer (200mM MOPS pH 8.0, 50mM NaOAc, 10mM EDTA) and 40ml dH_2O. When set, the gel was placed in the tank and just covered in 1x running buffer (20mM MOPS pH 7.0, 5mM NaOAc, 10mM EDTA). 10ng of RNA was prepared for electrophoresis by adding 7μl RNA (+ DEPC dH_2O) to 23μl RNA sample buffer (300μl 10x gel buffer, 1500μl formamide, 14μl EtBr 10mg ml^{-1}, 486μl formaldehyde). The RNA sample was heated to 65^{o}C for 5 minutes prior to use. Glass plates were placed on top of the gel to prevent formaldehyde diffusion and the gel was run at 25v overnight. A gel photo was taken the next day as described in section 2.3.

2.6.4 Northern blotting

After electrophoresis the gel was soaked in 20x SSC for 15 minutes. RNA was then transferred to a Nytran nylon membrane (Flowgen Bioscience, www.flowgen.couk) using a standard northern blotting technique, with 20x

SSC as the blotting agent. After blotting overnight, the membrane was immediately UV cross-linked using a Stratalinker 1800 (Stratagene, www.genomics.agilent.com).

2.6.5 DNA radiolabelling and RNA blot hybridisation

Labelling of DNA and hybridisation was carried out as described in section 2.5 except that prehybridisation was carried out in 5x SSC prehybridisation buffer and hybridisation in 5x SSC hybridisation buffer.

2.7 Quantitative PCR

2.7.1 Primer selection

Primers used for qPCR validation of RNA-seq data were designed by an automated programme written by Dr Steve Kelly (University of Oxford). Primer pairs were designed to amplify 70-250bp regions of exons predicted in the gene models of version 5b.60 of the maize genome (www.maizesequence.org), with annealing temperatures of between 59-63^{o}C (with no more than 1^{o}C between each pair member). Alongside exon models and primer positions, the programme also depicted where and how many reads were mapped to individual genes as measured by RNA-seq. This allowed manual selection from the list of possible primer pairs for each gene

for those that were a) likely to amplify the expressed components of a gene model in a qPCR analysis and b) would give meaningful and quantifiable comparisons with the RNA-seq data. Primers used for qPCR analysis of heterologous gene expression were designed using Primer3 software (www.primer3plus.com) with default options selected.

2.7.2 Standard curve and amplification efficiency calculation

Using 5x 2 fold dilution series (360ng/μl as the most concentrated sample ≡ ~100,000 gene copies, 180ng/μl as the next most concentrated etc) of genomic DNA, standard curves were calculated for 97 primer pairs designed to amplify genes expressed during early leaf development for RNA-seq validation. 5 primer pairs were also designed for housekeeping genes (based on consistency in expression across the samples tested in this study and from the literature), and standard curves calculated for these. PCR amplification was carried out using GoTaq Hot Start polymerase (Promega, www.promega.com) and amplification detected with 1/60,000 SYBR Green II (Sigma-Aldrich, www.sigmaaldrich.com) and the Mx3000P QPCR System (Agilent, www.genomics.agilent.com). The thermal profile ran as follows: 95^{o}C, 10 minutes; (95^{o}C, 15 seconds, 57^{o}C, 30 seconds, 72^{o}C, 30 seconds) x 45 cycles. This was followed by a gradual ramp from 55^{o}C to 95^{o}C to determine a dissociation curve for each reaction. Baseline fluorescence levels were set manually at the bottom of the log-linear phase of amplification and dissociation curves checked for primer dimer formation. Amplification efficiencies and R-sq values were calculated manually for each pair and 20

primer pairs selected with appropriate amplification efficiencies (90-110%) and R-sq values (>98.5) across three replicates. For heterologous expression analyses primer pair amplification efficiencies and dissociation curves were checked individually. Thermal profiles were altered as necessary depending on primer Tm values.

2.7.3 Expression quantification

For qPCR validation of the RNA-seq data, 1µg of total RNA was retained from samples FP1, HP1 and FI1. RNA was treated with RQ1 DNAase (Promega, www.promega.com) and cDNA was synthesised using SuperScript II Reverse Transcriptase (Invitrogen, www.invitrogen.com) following the manufacturer's instructions. In addition, a no reverse transcriptase control was produced by following the same protocol but replacing SuperScript II with dH_2O. For each sample, two PCR replicates were carried out for each of two independent cDNA syntheses. Fold changes for each gene in each sample were then calculated using the Pfaffl method (Pfaffl, 2001) and correlated against fold changes of the same genes in the same samples as measured by RNA-seq. For qPCR analysis of heterologous gene expression 1µg of total RNA was treated with RQ1 DNAse. cDNA was synthesised using AffinityScript Reverse Transcriptase (Agilent, www.agilent.com). Three PCR replicates were carried out for each sample. Expression was normalised to the transcript level of an endogenous *UBQ* fusion gene (rice – Os03g13170, *Setaria* – Si037467) and quantified using the delta delta Ct method (Livak and Schmittgen, 2001).

2.8 Protein manipulation

2.8.1 Protein extraction

Tissue was ground in liquid nitrogen before being added to an appropriate volume (3ml per g of tissue) of homogenisation buffer (18% sucrose, 10mM $MgCl_2$, 100mM Tris-HCl pH 8.0, 40mM β-mercaptoethanol). Samples were spun at 13,000rpm for 12 minutes before supernatants were transferred to new tubes and 0.6x volumes 2x stop dye (1% SDS, 0.1% bromophenol blue, 10mM EDTA, 20% ficoll) was added. Samples were boiled for 10 minutes and placed on ice prior to use, or stored at -80^0C.

2.8.2 SDS-polyacrylamide gel electrophoresis

12% acrylamide resolving gels (2M Tris-HCl pH 8.8, 10% SDS, 12% acrylamide, 10% APS, 0.001% temed) were first cast before overlaying with 15% acrylamide stacking gels (2M Tris-HCl pH 6.8, 10% SDS, 15% acrylamide, 10% APS, 0.002% temed). Samples were reboiled for 5 minutes if stored at -80^oC and electrophoresed for 45 minutes at 200v in protein running buffer (0.025M Tris base, 0.2M glycine, 0.001% SDS).

2.8.3 Coomassie staining

To assess sample protein integrity and relative abundance, gels were stained in Coomassie stain (0.5% Coomassie Blue R, 10% acetic acid, 50% methanol) for 5-10 minutes and destained for 4 hours – overnight using destain solution (10% acetic acid, 20% methanol). Destained gels were subsequently dried under vacuum and photographed using a Nikon Coolpix 990 camera (Nikon, www.nikon.co.uk).

2.8.4 Western blotting

Prior to blotting, gels were left to equilibrate in blotting buffer for 20 minutes (25mM Tris pH 8.3, 192mM glycine, 20% methanol). A standard western blotting apparatus was assembled and protein transferred in blotting buffer overnight at 30v, 40mA with an ice pack and constant stirring.

2.8.5 Western blot immunoreaction

Blots were first incubated in blocking buffer (5% non fat milk, 125mM NaCl, 10mM phosphate buffer ph 7.0) for 3 x 10 minutes while being shaken. Blots were then incubated in 10ml blocking buffer plus 10μl primary antibody serum (the antibodies used were raised in rabbits against maize malate dehydrogenase (MDH), wheat Rubisco large subunit, *Amaranthus viridis* phosphoenolyruvate carboxylase (PEPC) and maize nicotinamide adenine dinucleotide phosphate malic enzyme (NADP-ME)) and incubated for 12-24 hours at 4^{o}C. Blots were then again washed for 3 x 10 minutes in blocking

buffer before being incubated in 10ml blocking buffer with 5μl secondary antibody (goat anti-rabbit IgG, Sigma-Aldrich) for 1-3 hours at room temperature. If blots were to be developed using colorimetric detection they were washed in 3 changes of blocking buffer again before being placed in stain solution (500μg ml^{-1} 4-chloro-1-naphthol, 125mM NaCl, 10mM phosphate buffer pH 7.0, 0.00125% H_2O_2) until bands appeared. Where chemiluminescence was used for detection, blots were washed in blocking buffer containing 1mg ml^{-1} BSA instead of 5% milk. After dabbing the blots dry on paper, 400μl of chemiluminescent luminol reagent was mixed with 400μl chemiluminescent oxidising agent which was then poured over the blot. After 5 minutes of gentle agitation the blot was dried, wrapped in cling film and exposed to Kodak X-OMART AR film (Kodak, www.kodak.com) in a light-tight cassette. After 5-20 minutes exposure the film was developed using Kodak AR processing chemicals (Kodak).

2.9 Bundle sheath and mesophyll cell preparations

BS and M cells were separated from fully expanded 3rd leaves of *S. bicolor* inbred line BTx623 that were harvested during the morning. M cells were separated enzymatically from leaf tissue essentially as described previously (Sheen and Bogorad, 1985), but with vanadyl ribonucleoside complex omitted from the protoplast washing buffer. Bundle sheath strands were isolated mechanically using a household blender. Leaves were blended and filtered through 60μM mesh using buffers described previously by Westhoff et al

(1991). Cell preparations were checked microscopically for purity and immediately frozen in liquid nitrogen before storage at -80 °C.

2.10 Leaf sectioning

2.10.1 Hand sectioning

Thick fresh leaf sections (0.5-1mm in width) were cut by hand under a Leica Wild M8 stereo microscope (Leica, www.leica-microsystems.co.uk) using a standard lab scalpel. Sections were immediately mounted in water on slides and analysed using a Leica DMRB microscope with a QImaging MicroPublisher camera (QIamging, www.qimaging.com) and Image-Pro Insight software (MediaCybernetics, www.mediacy.com).

2.10.2 Microtome sectioning

2-3mm leaf sections were hand sectioned and fixed in either 3:1 ethanol:acetic acid for 30 minutes or FAA (4% formaldehyde, 5% acetic acid, 50% ethanol) overnight. Samples were then transferred to 70% ethanol for storage. Tissue was subsequently embedded in ParaPlast Plus using Tissue-Tek vacuum infiltrating (VIP) and embedding (TEC) equipment (Sakura, www.sakura.eu). Thin sections (8μm) were cut using a Leica RM2135 rotary

microscope. Sections were stained in Safranin/Fast Green as described previously (Langdale, 1994) and analysed as described in section 2.10.1.

2.11 Chlorophyll concentration measurement

For each leaf sample three 0.5cm^2 discs were removed using a leaf borer. Samples were ground in liquid nitrogen and resuspended in 1ml 80% acetone overnight at 4^{o}C in the dark. Cell debris was collected by brief centrifugation and the supernate absorbance measured at 645nm and 663nm. Chlorophyll concentration was calculated as follows: Chlorophyll concentration (µg/ml) = (OD_{645} x 20.2) + (OD_{663} x 8) (Arnon, 1949)

2.12 RNA-seq read mapping and modelling

2.12.1 Read mapping and transcript quantification

To validate the RNA-sequencing data returned by the Beijing Genomics Institue (BGI), and carry out additional bioinformatic analyses, Illumina paired-end read data was initially mapped to the maize genome version 4a.53 using the default options in the read mapping programme Stampy v1.0.0 (Lunter and Goodson, 2010). Using the Stampy SAM file outputs, transcript quantification was calculated using the default options in the Cufflinks programme v.0.9.0 (with the maize reference genome specified as an input for Cuffcompare component of the programme) (Trapnell et al., 2010).

Subsequent bioinformatics analyses, including the generation of candidate gene lists and the characterisation of leaf transcriptome signatures, were based on transcript quantification and differential expression estimates carried out by Dr Steve Kelly using the maize genome version 5b.60 and the programmes RSEM and DESeq (Anders and Huber, 2010; Li and Dewey, 2011) with the default parameter settings.

2.12.2 Statistical analysis of transcript abundance

Frequency distributions of transcripts were modelled using the R software package (www.r-project.org). Mixed distributions were modelled using the mixdist R package. Approximations for distribution means, standard deviations and component proportions were initially input manually and then refined by the mixdist package; other inputs were set as default. Modelling of the frequency of GC content at the third codon position was carried out by Dr Steve Kelly.

2.13 Phylogenetic analyses

2.13.1 Phylogenetic analyses of *GLK* genes

To identify *GLK* genes, BLASTP was used to search all of the annotated land plant proteomes on Phytozome v8.0 (http://www.phytozome.net) plus the

potato genome sequence (http://potatogenomics.plantbiology.msu.edu/), using the ZmGLK1 amino acid sequence as a query. Results for searches against each proteome were filtered manually to identify *GLK* genes (distinguished from other GARP family genes by an AREAEAA motif (consensus motif) at the C terminal of the DNA-binding domain). To ensure that all putative *GLK* genes were identified, the amino acid sequences encoded by 5 *GLK* genes representing a wide range of angiosperm lineages (*AtGLK1*, *GmGLKD*, *VvGLK*, *ZmGlk1*, *OsGLK2*) were aligned using MAFFT (Katoh et al., 2005). This alignment was converted to a hidden Markov model and used to search Phytozome v8.0 plant and algal proteomes with an iterative HMMer search algorithm described previously (Eddy, 1998; Kelly et al., 2011). No additional sequences were found. *Cleome GLK* sequences were provided by Dr Julian Hibberd (University of Cambridge).

Phylogenetic trees of the identified *GLK* genes were inferred using both Bayesian and maximum likelihood methods. Protein sequences were aligned using Merge-Align (Collingridge and Kelly, 2012). A 100 bootstrap maximum likelihood tree was inferred using RAxML (Stamatakis, 2006) employing the LG model of sequence evolution (Le and Gascuel, 2008) and CAT rate heterogeneity. A 50 % majority-rule consensus tree was calculated from the 100 bootstrap replicates using the python module DendroPy (Sukumaran and Holder, 2010). Bayesian phylogenetic trees were inferred using mrbayes v3.1.2 (Huelsenbeck and Ronquist, 2001) with gamma-distributed substitution rate variation approximated by four discrete categories and shape parameter estimated from the data. The ‘‘covarion’’ model (Galtier, 2001) was

implemented and four chains were employed, each with a temperature of 0.2. Tree inference was made from a random start tree and allowed to run for 2,500,000 generations. The time taken to reach stationary phase was approximately 700,000 generations and thus the final 1,800,000 trees sampled every 200 generations were used to infer posterior probabilities on topology. Tree output files were analysed using Dendroscope software (Huson et al., 2007).

2.13.2 Phylogenetic analyses of Kranz candidate genes

Amino acid sequences for individual genes were converted to hidden Markov models and used to search 22 Phytozome v8.0 plant and algal proteomes representative of different land plant lineages as described above. The amino acid sequences identified by the HMMer search were aligned using MAFFT. Phylogenetic trees were inferred using Bayesian methods as described above, except tree inference only ran for 1,000,000 generations and the final 825,000-900,000 trees, depending on the time taken to reach stationary phase, were used in the final analysis. For each candidate gene, phylogenetic trees were also constructed from sequences identified using only the Pfam defined domains (www.pfam.sanger.ac.uk) present in the candidate amino acid sequences as inputs for HMMer searches. Where Pfam domains have been annotated in figures, a Perl script developed by Dr Steve Kelly was used to highlight the ten most abundant Pfam v27.0 domains in the amino acid sequences used for phylogenetic tree construction.

2.13.3 Additional phylogenetic analyses

In addition to using alignment based tree inference programmes, DendroBLAST, an alignment-free method of phylogenetic tree reconstruction based on BLAST scores (Kelly and Maini, 2013), was used where specified in the text. BLASTP was used to search Phytozome v8.0 to identify closely related sequences to a protein of interest and approximate phylogenetic trees inferred using the default settings on DendroBLAST, which was run for 100 replicates.

2.14 *In silico* promoter sequence analysis

2kb of genomic DNA sequences 5' upstream of the translational start site of sorghum and maize *GLK* genes were identified on Phytozome. *GLK1* or *G2* sorghum and maize orthologues were used as a query sequence for *de novo* motif identification with MEME software (www.meme.ncbr.net). Equivalent 2kb regions of *Brachypodium* and rice DNA upstream of *GLK1* or *G2* were input as negative sequences as part of a discriminative motif analysis. Input options were selected as follows: zero or one motif per sequence, maximum 10 motifs, minimum motif width 6bp, maximum motif width 50bp. Identified motifs were searched against the JASPAR CORE Plants transcription factor binding site database (www.jaspar.genereg.net) using the defaults options in the MEME subsidiary programme TOMTOM. Equivalent 2kb regions upstream of *Setaria italica* and *Panicum virgatum GLK* genes were screened for the

presence of conserved domains by inputting individual sequences into MEME together with the appropriate maize and sorghum sequences. Individual grass promoter regions were also screened against the TRANSFAC database (www.gene-regulation.com/pub/databases.html) using the Match algorithm with the minimise false positive option selected to directly identify transcription factor binding sites.

Chapter 3: Deep-sequencing transcriptomic changes associated with Kranz development: Validation of an RNA-seq dataset

3.1 Introduction

The advent of next generation sequencing has allowed biologists to ask questions on a scale and depth not previously possible. Due to the development of high-throughput parallel sequencing technologies, genome and transcriptome sequences are available at a fraction of the time and cost compared to only a few years previously. Although there are differences in the underlying chemistry, next-generation sequencing platforms such as the 454 (www.454.com), Illumina (www.illumina.com) or SOLiD systems (www.lifetechnologies.com) all implement cyclic-array sequencing (Shendure and Ji, 2008). Dense aggregations of DNA are iteratively amplified in parallel with each round of amplification producing imaging data as to the nucleotide(s) that have been incorporated. The concomitant amplification and sequencing of DNA vastly reduces sequencing time compared to conventional Sanger sequencing methods, meaning genome sequences take weeks rather than years to complete. In addition to quickening genome sequencing, next-generation technology has allowed enhanced large-scale RNA quantification. The increased sensitivity, coverage and speed of such RNA-sequencing (RNA-seq) analyses compared to conventional techniques for measuring gene expression, such as microarrays or Serial Analysis of Gene Expression

(SAGE) methods, has revolutionised the capacity to quantify gene expression on a large-scale (Mardis, 2008; Shendure and Ji, 2008). Beyond genome and transcriptome sequencing, next-generation sequencing has also allowed genome-wide analyses of DNA-protein interactions and detection of DNA methylation sites (Lister et al., 2009).

In plants, early studies utilising RNA-seq demonstrated the power of the technology for revealing cell and tissue specific transcriptomic signatures, particularly in identifying lowly expressed transcripts and previously unannotated transcripts and splice variants (Cheung et al., 2006; Emrich et al., 2007; Weber et al., 2007; Lister et al., 2008). The potential of RNA-seq to help elucidate the transcriptomic networks that regulate development was clear. The transcriptomic networks that regulate Kranz development, a characteristic cellular arrangement crucial for C_4 photosynthesis in nearly all C_4 lineages, have previously been particularly difficult to resolve. This is despite the multiple independent origins of Kranz and the widespread international interest in introducing Kranz anatomy and C_4 photosynthesis into crop species (Hibberd et al., 2008; Sage et al., 2011a). Previous attempts to identify genetic regulators of Kranz development have largely depended on forward genetic screens and have identified few genes with roles in the patterning of Kranz anatomy (reviewed in Langdale, 2011).

The utilisation of RNA-seq provides an unparalleled opportunity for insight into the genes that regulate Kranz initiation. Crucial to any such RNA-seq investigation is the selection of appropriate biological samples. Maize

develops dimorphic leaves that exhibit Kranz and non-Kranz anatomy (Langdale et al., 1988b). Sequencing the mRNA from a range of developmental stages of both leaf types, pre and post anatomical patterning, should thus elucidate the transcriptomic changes that promote (or repress) Kranz development. This chapter describes the work done to validate RNA-seq data from such a set of samples, a necessary step before downstream transcriptome analyses or candidate gene filtration can be completed.

3.2 Results

3.2.1 The stages of leaf development sampled for RNA-seq show distinct developmental trajectories

To investigate the transcriptomic changes that occur during the patterning of Kranz anatomy in maize, RNA was extracted from leaves that do (foliar) and do not (husk) develop Kranz anatomy. Tissue harvesting, RNA extraction and tissue sectioning were carried out by Dr. Peng Wang. Figure 3.1A depicts where on the plant each sample was taken and the overall sample morphology. A total of 10 samples were isolated that cover a range of developmental stages, from plastochron (P)1 to 9 in foliar (F) (5 samples) and husk (H) (5 samples) leaves. A plastochron is the time interval between each newly initiated leaf primordia, with P1 being the most recently initiated primordium (Sharman, 1942; Esau, 1943). The anatomical characteristics of each sample can be seen in the transverse sections examined by light

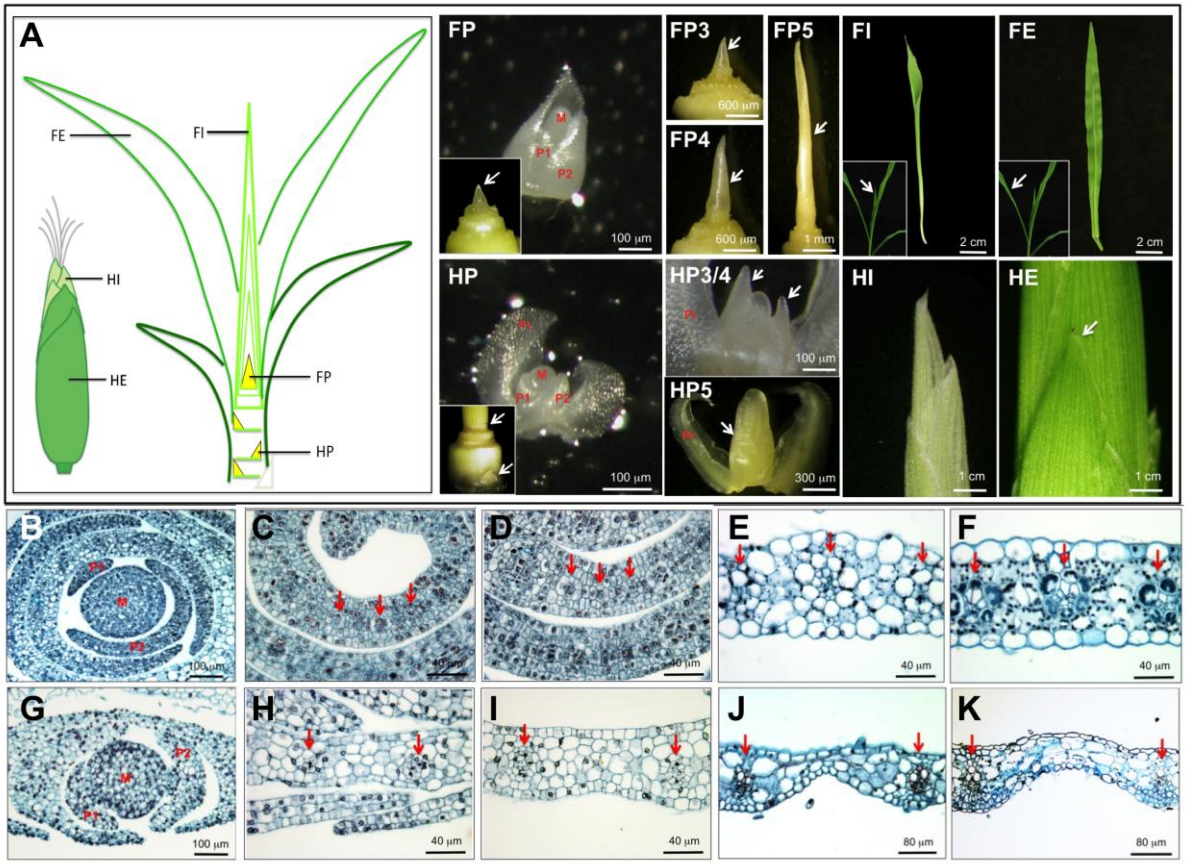


Figure 3.1. Biological samples used for transcriptome sequencing.

Figure legend is on opposite page.

Figure 3.1. Biological samples used for transcriptome sequencing. A) Schematic and images of the 10 tissues sampled. Foliar primordia (FP) and husk (HP) primordia samples included the vegetative apical and axillary inflorescence meristem (m) respectively, plus P1 and P2 leaf primordia. Arrows in the insets show the position of the samples prior to dissection from the plant shoot. The prophyll (Pr) was discarded from the HP sample prior to RNA extraction. FP3/4, FP5, HP3/4 and HP5 primordia (arrows) were isolated from the meristem and other leaf primordia prior to RNA extraction. Foliar immature (FI) and foliar expanded (FE) leaf blades were harvested above the ligule. Arrows in the insets show the position of the leaves on the shoot prior to harvesting. Husk leaf samples were the outermost leaf (HE) on the ear and third leaf in (HI) from the outside. **B-K)** Transverse sections showing leaf anatomy in FP (B), FP3/4 (C), FP5 (D), FI (E), FE (F), HP (G), HP3/4 (H), HP5 (I), HI (J) and HE (K) samples. Arrows point to developing or developed veins. Photos and sections courtesy of Dr. Peng Wang.

microscopy shown in Figure 3.1B-K. In the foliar primordia (FP) and husk primordia (HP) samples, the procambium associated with the midvein is barely visible (Figures 3.1B and G) in P1 and P2 of both leaf types. By P4, however, lateral veins have started to initiate and the distance between pairs of veins is already smaller in foliar, rather than husk, leaves (Figures 3.1C and H). At P5, foliar leaves have initiated intermediate veins and concentric rings of bundle sheath (BS) and mesophyll (M) cells are visible around each vein, with each vein separated by only two M cells from neighbouring veins (Figure 3.1D). However, at P5 in husk leaves, no intermediate veins have formed and cell division and expansion have increased the distance between lateral veins, with veins separated by approximately 10 M cells (Figure 3.1I). There are no further changes in vein spacing in foliar immature (FI) (Figure 3.1E) and foliar expanded (FE) (Figure 1F) samples, with the classic Kranz cellular arrangement clearly apparent, but there is additional cellular development. BS cells are further enlarged relative to M cells and chloroplast size increases from FI to FE as leaf tissue becomes increasingly photosynthetically active. Despite the active C_4 photosynthetic pathway that occurs locally around veins in light-exposed husk leaves (Langdale et al., 1988b; Pengelly et al., 2011), BS cells and chloroplasts remain smaller in the husk exposed (HE) sample than in FE (Figure 3.1E and K). The paucity of chloroplasts in M cells of husk tissue is consistent both with a lack of light exposure for HI (Figure 3.1J) and a lack of photosynthetic activity in between veins in husk leaves (Pengelly et al., 2011) for HI and HE. The developmental stage and anatomical characteristics of each sample are summarised in Table 3.1.

Sample name	Developmental stage	Kranz anatomy	Veins initiated	No. of M cells between veins	BS cell size	BS plastid size
FP	Foliar P1&2 plus SAM	No	Midvein	na	na	na
FP3/4	Foliar P3&4	Initiating	Midvein Laterals	4-6	na	na
FP5	Foliar P5	Elaborating	Midvein Laterals Intermediates	1-2	small	na
FI	Foliar P7	Anatomy fully differentiated	Midvein Laterals Intermediates	2	large	small
FE	Foliar P9	Fully differentiated	Midvein Laterals Intermediates	2	large	large
HP	Husk P1&2 plus AM	No	Midvein	na	na	na
HP3/4	Husk P3&4	No	Midvein Laterals	5	small	na
HP5	Husk P5	No	Midvein Laterals	10	small	na
HI	Third innermost leaf around developing ear	No	Midvein Laterals	~12	medium	small
HE	Outermost leaf around developing ear	No	Midvein Laterals	~15	medium	small

Table 3.1. Developmental stage and key anatomical traits of foliar and husk leaf samples.

3.2.2 Reads mapped using the SOAP2 programme fit a mixed distribution with high significance

RNA extraction from the 10 samples described above was sent to the Beijing Genomics Institute (BGI) for cDNA library construction and Illumina paired end sequencing. The BGI bioinformatics service also mapped and quantified the reads using the SOAP2 alignment package (Li et al., 2009). To validate the BGI quantification pipeline, and to carry out further statistical analyses, the frequency distribution of the BGI read data was modelled for the first six samples sequenced: FP, FI, FE, HP, HI and HE. Figure 3.2 shows histograms of log2 transformed global gene expression data for each of these six samples, with gene expression measured as reads per kilobase per million reads (RPKM) (this transformation normalises the read data for transcript length and library size) (Mortazavi et al., 2008). Although the reads appear to be similarly distributed in each of the samples, the presence of a large left-hand skew in each histogram suggests that the reads do not fit a single normal distribution (although see Soneson and Delorenzi (2013) for discussion on distribution types of RNA-seq) and that the histograms might represent two overlapping distributions.

To determine whether the read data could be more accurately fitted to two overlapping distributions rather than to a single distribution, the mixdist package in R was used (see Materials and Methods) to calculate the statistical likelihood of a mixed distribution. Figure 3.3A shows the probability density function of the RPKM data for FP (which is representative of the other

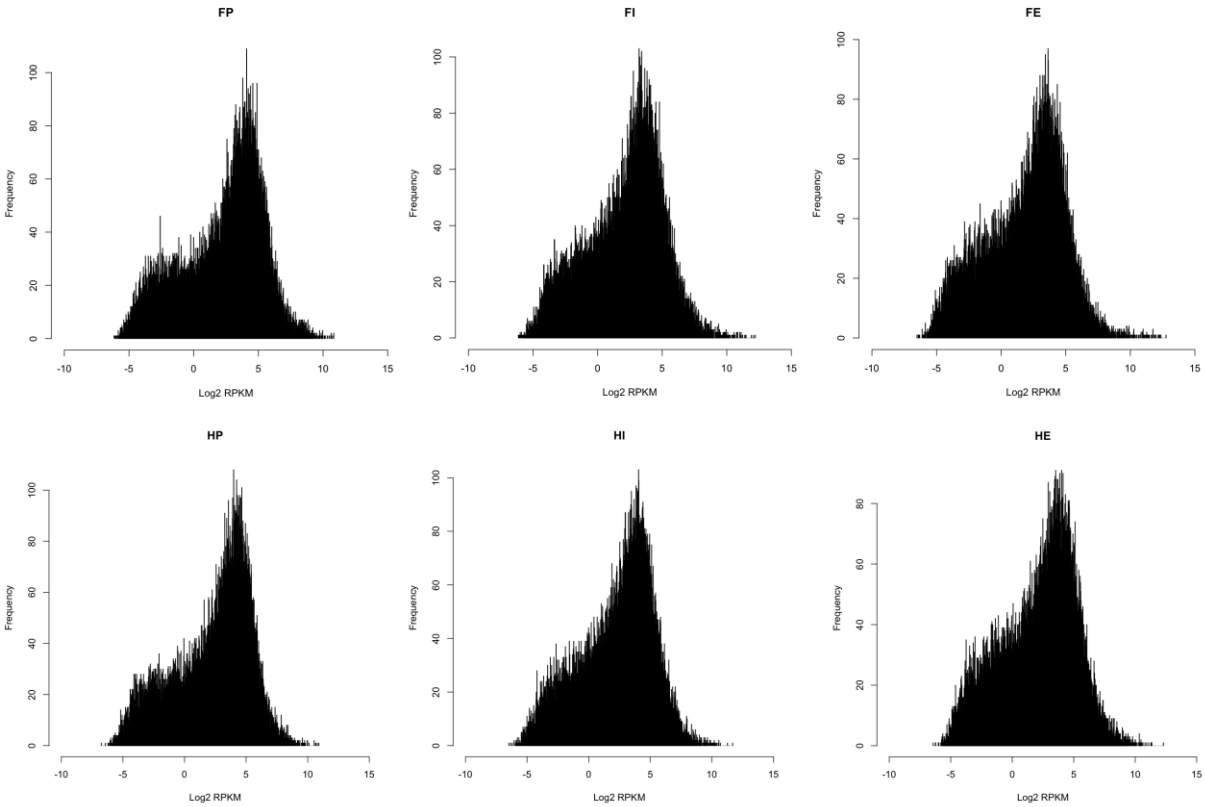


Figure 3.2. SOAP2 mapped reads fit a bimodal distribution. Histograms of read counts (normalised to RPKM values) calculated by the Beijing Genomics Institute using the SOAP2 alignment package for six samples: FP, FI, FE, HP, HI and HE. Data were split into 500 bins.

five samples), with the total distribution clearly made up of two component distributions. For all six samples, a chi-square test shows highly significant support for two distributions ($p < 0.000001$).

3.2.3 An alternative method of read mapping removes overlapping read distributions by adjusting for GC-skew

To determine what an appropriate cut off for 'noise' (defined as false values for gene expression as a consequence of experimental artefacts) might be at the lower end of the range of values for gene expression, it was important to identify why there were overlapping distributions in the read data. The lower distributions (which corresponded to between ~21-29% of the total read data) could be due to low levels of genomic DNA contamination, organelle DNA contamination or an amplification artefact during Illumina library preparation. If the lower distribution corresponded to reads that were a consequence of DNA contamination, regardless of origin, it would be more appropriate to set a noise threshold based solely on the properties of the upper distribution (assuming the upper distribution reflected the presence of reads derived only from RNA). If the bimodality of the read distribution was due to an amplification artefact during cDNA amplification prior to sequencing, it would be more appropriate to set two noise thresholds, one based on each distribution.

The summary statistics returned by BGI suggested some genomic contamination in the samples, as approximately 8% of reads mapped to

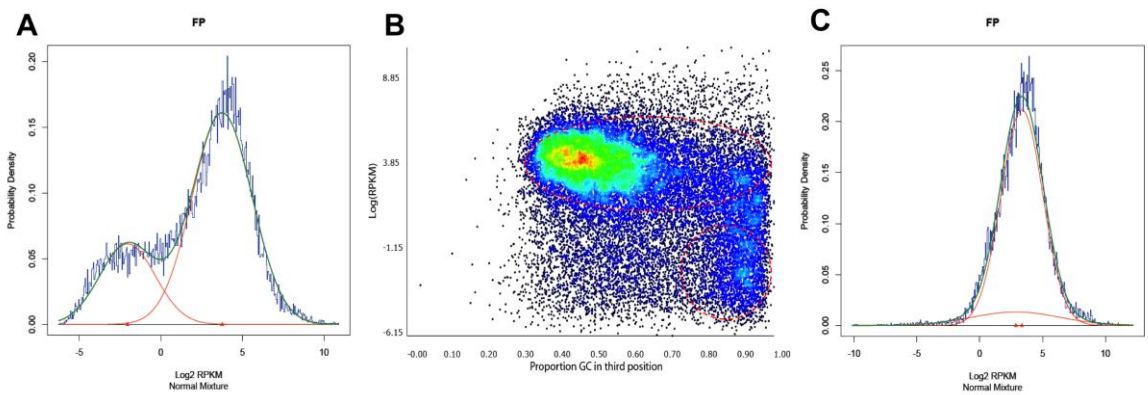


Figure 3.3. A GC-bias explains bimodal read distributions and can be corrected using Cufflinks. A) Probability density function of read data produced using SOAP2 for the representative FP sample; blue line - histogram of log2 transformed RPKM data (bins = 500); green line – fitted mixed distribution; red lines – fitted component normal distributions (arrow heads show component distribution means). **B)** Heatmap showing log2 transformed RPKM values and proportion of GC_3 for mapped genes; dotted red lines correspond to component distributions in (A) (adapted from work carried out by Dr. Steve Kelly). **C)** Probability density function of read data produced using Stampy and Cufflinks for the representative FP sample; blue line - histogram of log2 transformed RPKM data (bins = 500); green line – fitted mixed distribution; red lines – fitted component normal distributions (arrow heads show component distribution means).

intergenic regions of the maize genome. However, modelling the GC content at the 3rd codon position (GC_3) and RPKM of all the predicted coding regions in the genome suggested that inefficiencies during library amplification were a more likely cause of the overlapping distributions (Figure 3.3B; adapted from work carried out by Dr. Steve Kelly). A plot of log2 transformed RPKM values against transcript GC_3 content also revealed two overlapping distributions, highly similar in relative magnitude and proportion to those shown in Figure 3.3A. GC-rich genes were generally expressed at a much lower level than GC-poor genes. It has previously been shown that PCR amplification biases in the Illumina sequencing protocol can lead to under representation of GC_3 rich genes (Aird et al., 2011). Thus, the lower distribution shown in Figure 3.3A likely reflects inefficient amplification of GC-rich genes.

To confirm whether a GC-skew was the cause of the overlapping distributions of read data, an alternative method of read mapping and quantification was utilised that has been shown to correct for GC-bias during library amplification. Raw sequence data was mapped to the maize genome version 4.53a (the same version used for the SOAP2-based read quantification) using Stampy, and quantified using Cufflinks (see Materials an Methods for details). Cufflinks has previously been shown to correct for amplification biases and thus produces more accurate predictions of gene expression using RNA-seq data (Roberts et al., 2011). Using this combination of programmes led to an increased number of reads mapped to genes for each sample (Table 3.2) and a unimodal distribution of read data (Figure 3.3C). Due to the increased sensitivity and bias reduction of this mapping pipeline, expression values

	BGI mapping	Stampy/Cufflinks mapping
FP	19930323	23053007
FI	19015852	21963828
FE	21299638	24804159
HP	23095349	26527738
HI	19655694	22898853
HE	18412274	21799351

Table 3.2. Total number of reads mapped to selected samples using different mapping and quantification pipelines.

predicted by Stampy and Cufflinks based quantification were assumed to be more accurate than SOAP2 calculated values. The Stampy/Cufflinks data were therefore used to test the validity of the dataset through comparison with published datasets.

3.2.4 Sequence data correlate with existing gene expression datasets

To validate the Illumina read data, expression levels for a subset of well-characterised genes were compared with published data. All tested genes

and samples correlated as expected. For example, the meristematic homeobox gene *KNOTTED1* (GRMZM2G017087) was only found in the FP and HP samples (which contain meristematic tissue) (Figure 3.4A); the gamma tubulin gene *TUBG1* (GRMZM2G073888) gene was expressed in all tissues but most strongly in the primordia samples that are undergoing the highest degree of cytoskeletal rearrangement (Figure 3.4B); and genes encoding C_4 pathway enzymes *PEPC1* (GRMZM2G083841) (Figure 3.4C) and the *NADP-ME* isoform *ME3* (GRMZM2G085019) (Figure 3.4D) were upregulated in tissues with photosynthetic activity (FI, FE and HE).

To test for concordance between samples within this dataset and other previously published RNA-seq datasets, two studies that used comparable developmental stages were identified. The first was a transcriptome of four samples of the developmental gradient of a maize leaf (Li et al., 2010) and the second was a shoot apical meristem (SAM) transcriptome (Jia et al., 2009). Both of the published datasets were downloaded from the NCBI SRA archive (http://www.ncbi.nlm.nih.gov/sra). Figure 3.5 shows a heatmap of the Pearson product moment correlation coefficient (calculated by Dr. Steve Kelly) between the six samples used for initial validation in this study, and between the six samples and the four leaf gradient and SAM samples. Comparisons between the six samples from the current study show that the two primordia samples exhibit the strongest correlation. FI is most similar to FE and HI is most similar to HE. HE correlates more strongly with FI/FE than HI does. Comparison with the SAM and leaf gradient datasets also showed expected outcomes, with the SAM data having most similarity to FP. The leaf gradient

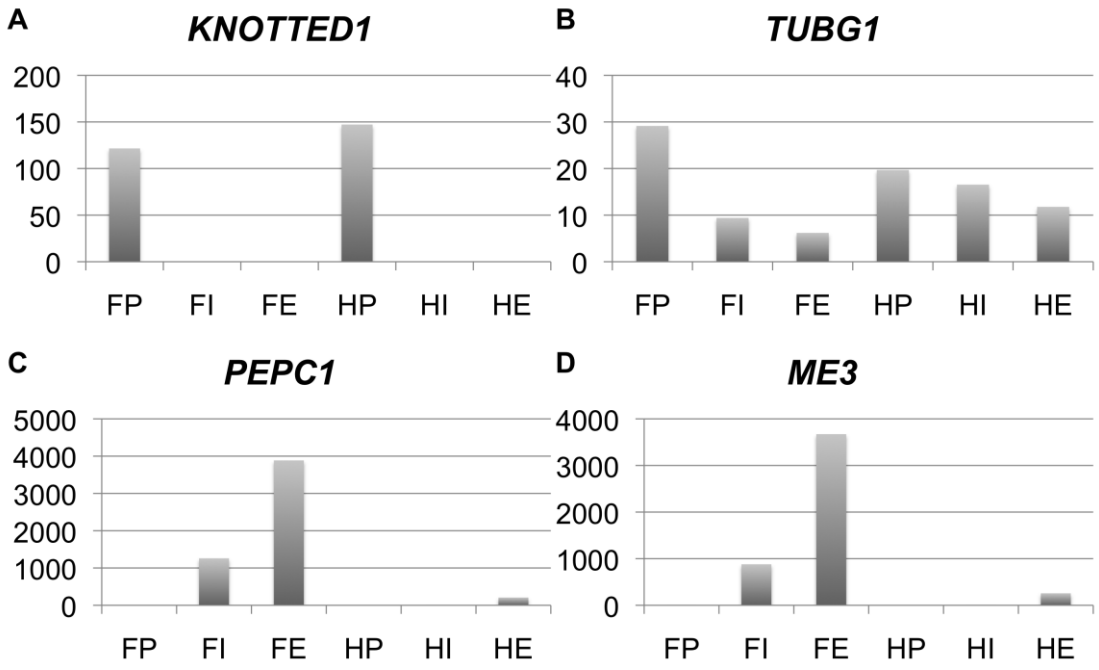


Figure 3.4. Expression of individual genes correlate as expected A) *KNOTTED1* (GRMZM2G17087), B) Gamma-tubulin chain *TUBG1* (GRMZM2G073888) **C)** *PEPC1* (GRMZM2G083841) **D)** *NADP-ME* isoform *ME3* (GRMZM2G085019). Expression values calculated using Stampy and Cufflinks, y-axis shows RPKM.

data, which reflects progressive increases in tissue differentiation and photosynthetic activity towards the leaf tip (Li et al., 2010), also correlates as predicted with the samples in this study: the base (least differentiated) sample correlates most strongly with the primordia samples, the 1cm sample (photosynthetic sink) has most similarity with FI and the 4cm and tip samples (photosynthetic source) have most similarity with FE. Taken together, the comparisons in Figure 3.5 show that global developmental trends are reflected in the sequencing data and that total sample expression profiles show expected relationships with published data taken from comparable developmental stages.

3.2.5 qPCR analysis supports the quantification estimates of the sequencing data

Although the correlation of sample expression profiles and individual gene expression values with existing datasets suggested that the RNA-seq data accurately reflected endogenous levels of gene expression, a comparison with estimates of gene expression measured by another method was necessary to conclusively validate the sequencing data. To this end, qPCR was carried out using RNA retained from the same FP, FI and HP extracts used for Illumina sequencing. Possible primer pairs were designed for 97 genes that had high expression levels in at least two of the FP, FI or HP samples, and 5 genes that showed consistent expression either across the samples in this study or in published datasets (see Materials and Methods for details). Primer pairs, read locations and read abundance (as predicted by

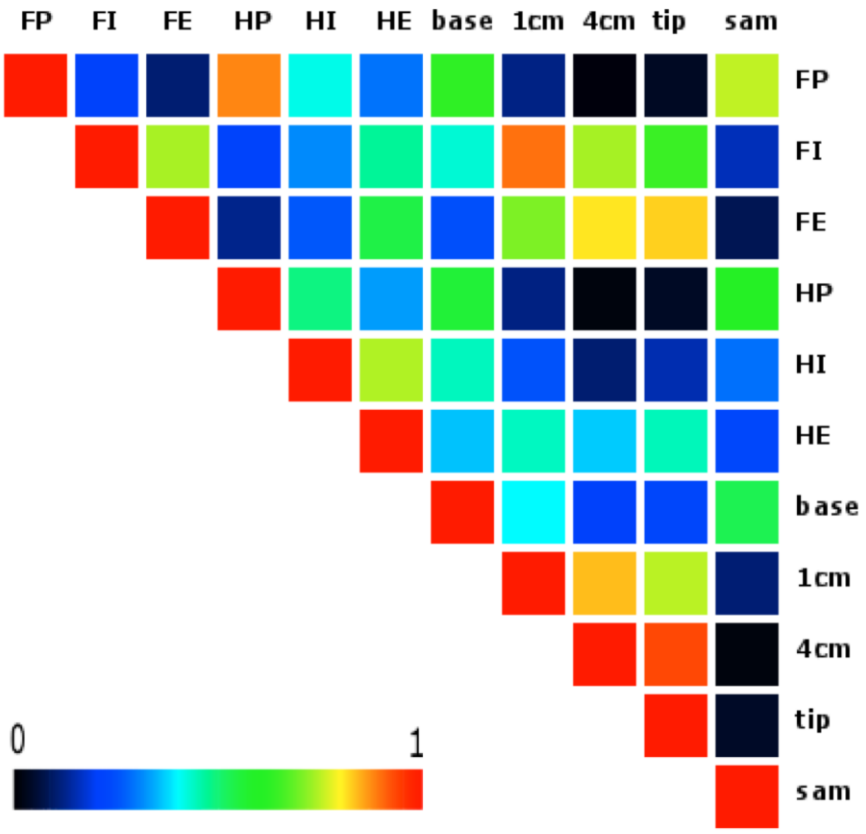


Figure 3.5. Sequencing data reflects developmental trends. Pairwise sample correlation coefficient for the global mRNA abundance for all pairwise sample comparisons. Scale bar indicates Pearson's correlation coefficient. FP, FI, FE, HP, HI and HE, this study; leaf gradient samples base, 1cm, 4cm, tip, see Li et al. (2010) ; sam, see Jia et al. (2009). Adapted from work carried out by Dr. Steve Kelly.

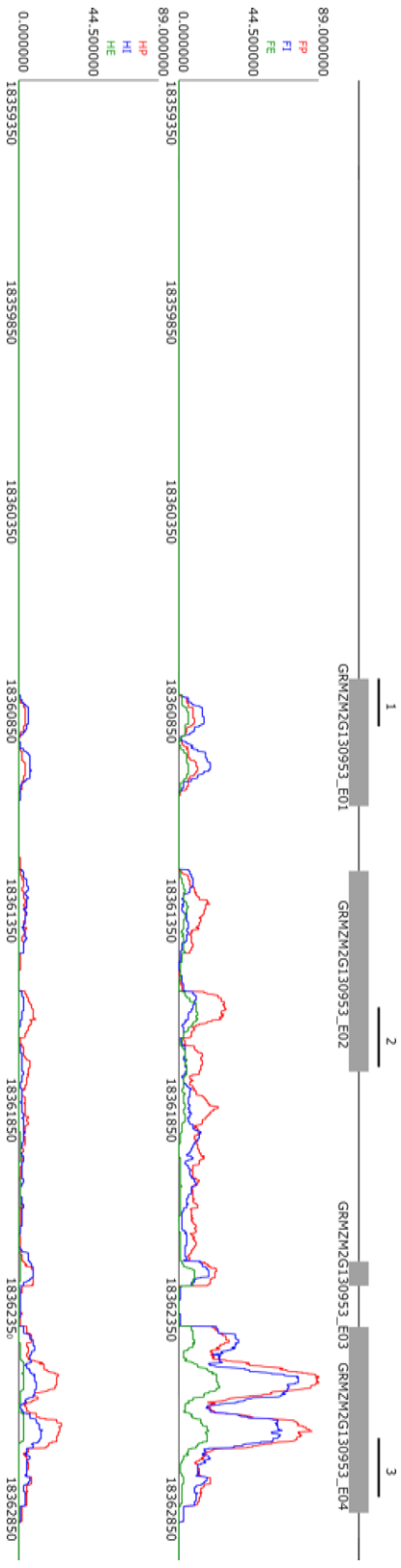


Figure 3.6. qPCR primer pairs were selected that overlapped with RNA-seq data. Figure legend is on opposite page.

Figure 3.6. qPCR primer pairs were selected that overlapped with RNA-seq data. Example of svg file used for primer pair selection (for gene GRMZM2G130953). Grey bars indicate exons, black lines above grey bars represent possible primer pairs. Size of peaks of coloured expression tracks represent number of reads mapped (values shown on axis along left) at that position along the chromosome (numbers beneath expression track show chromosome position in base pairs). Top expression track shows foliar samples and bottom track husk samples (red line – FP/HP, blue line – FI/HI, green line – FE/HE). Adapted from work carried out by Dr. Steve Kelly.

Figure 3.7. Example amplification plots and dissociation curves for FP cDNA. A) GRMZM2G011463, **B)** GRMZM2G138396. Amplification plots show measured fluorescence over increasing qPCR cycle number. Yellow lines represent manually set baseline fluorescence levels. Dissociation curves show the melting point for double stranded DNA amplicons. Grey and green lines are two technical replicates for RNA extraction 1, blue and red lines are two technical replicates for RNA extraction 2. Baseline fluorescence levels were set manually to peak background fluorescence after amplification cycle 5. (A) shows high similarity between technical replicates and RNA extractions (amplification plot) and no primer dimers (dissociation curve). In (B) the amplification curves show slightly reduced RNA quantity in RNA extraction 2 (increased number of cycles taken for blue line to reach log-linear phase) and faulty amplification due to primer dimer formation (melting point for the red line in the dissociation curve plot shows primer dimer melting at a lower temperature than amplicon melting (blue, green and grey lines). In this instance the data represented by the red line were not used for Ct calculation.

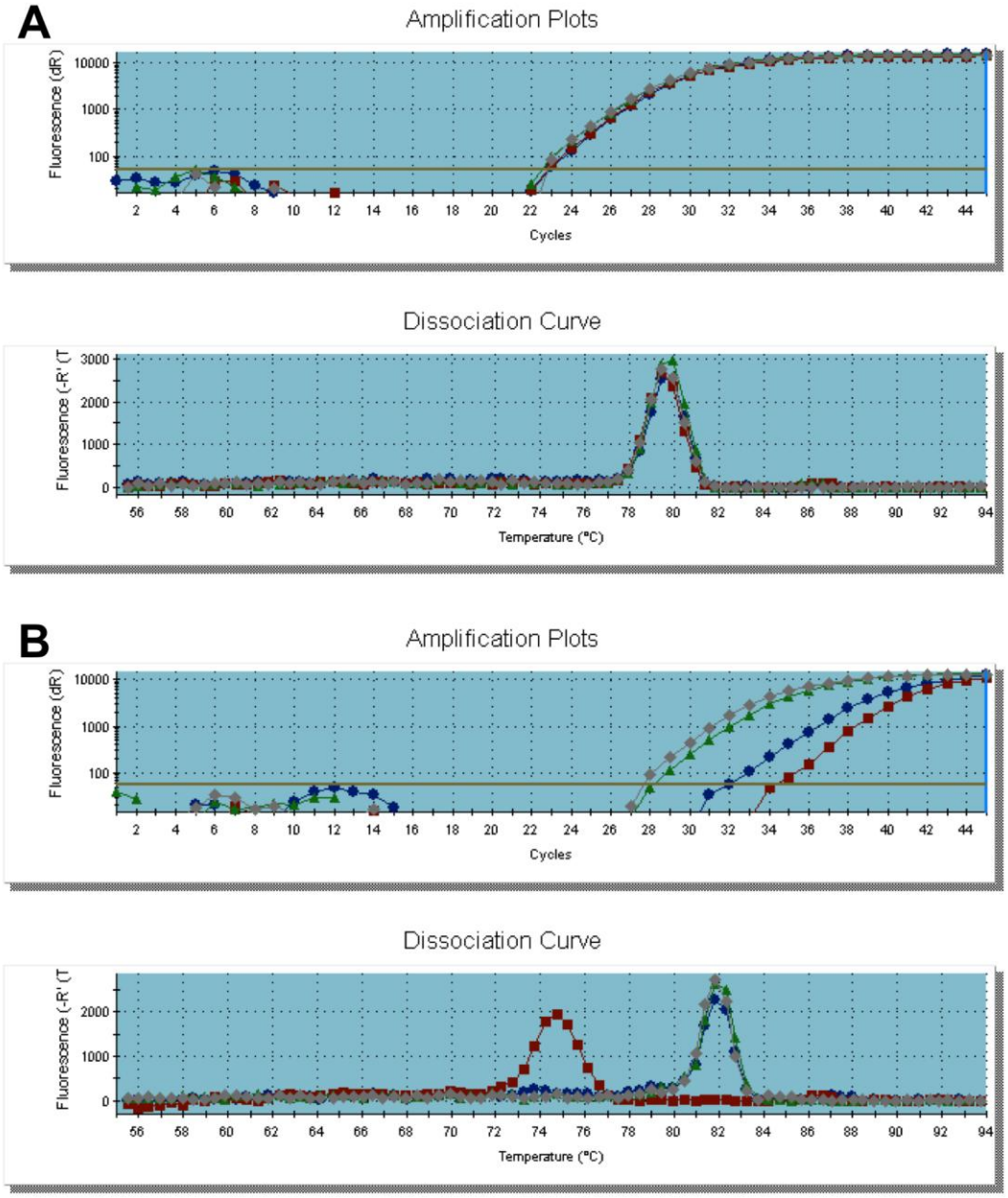


Figure 3.7. **Example amplification plots and dissociation curves for FP cDNA.** Figure legend is on opposite page.

RNA-seq) were mapped to gene models annotated in maize genome version 4.53a (Figure 3.6; adapted from work carried out by Dr. Steve Kelly). Individual primer pairs were selected manually from candidate pairs by identifying pairs that overlapped with regions of the transcripts that were identified as part of true mRNA by RNA-seq. This ensured primer pairs were selected that would amplify regions of genes also found to be expressed in the RNA-seq data, and would give more precise comparisons with the RNA-seq data (through comparison of exon specific read values, rather than comparing PCR data with RNA-seq data averaged across transcripts).

To enable quantification of changes in gene expression between samples by the Pfaffl method (Pfaffl, 2001) (an accurate method of estimating changes in gene expression by qPCR that requires knowledge of primer amplification efficiencies), standard curves were calculated for each primer pair using 5 serial 2-fold dilutions of genomic DNA. From the standard curve analyses, of the 102 primer pairs, 20 were found to have amplification efficiencies (90-110%) and amplification correlation coefficients (R-sq values > 98.5) required for optimal performance and minimisation of amplification artefacts. These 20 were used to calculate qPCR cycle threshold (Ct) values using cDNA produced from the FP, FI and HP samples. As part of the Ct value calculation process, baseline levels of fluorescence and primer dimers were checked manually (see Figure 3.7 for a detailed example, raw qPCR data shown in Appendix 3.1). Data from amplification reactions affected by primer dimer formation or reaction failure were discounted from further analyses. Amplification of 20 genes across the 3 named samples thus produced 25

values for ratios of gene expression between samples. Pairwise comparison of the same 25 ratios as predicted by the Stampy/Cufflinks RNA-seq data produced a highly significant correlation ($p < 0.000001$) (Figure 3.8). Using the SOAP2 mapped RNA-seq data, the correlation was very marginally weaker, although still highly significant (Stampy/Cufflinks correlation, $R^2 = 0.84$, $p = 1.28 \times 10^{-12}$; SOAP2 correlation, $R^2 = 0.83$, $p = 1.78 \times 10^{-12}$). The strength of the correlation between independently derived expression values validates the use of these RNA-seq data for biological interpretation. The qPCR results also reinforce the validity of using Stampy and Cufflinks in read quantification rather than SOAP2, although this analysis suggests that there is very little difference between the two when correlated to qPCR based expression values for a subset of genes.

3.3 Discussion

3.3.1 Choice of read quantification method affects estimates of transcript abundance

As deep sequencing of RNA becomes an increasingly affordable and widely used method of measuring gene expression, the associated bioinformatics pipelines necessary for mapping and quantifying sequencing data are becoming increasingly refined (Garber et al., 2011). When the Illumina sequencing for the current study was carried out, next-generation sequencing had only relatively recently emerged as a tool for sequencing genomes or

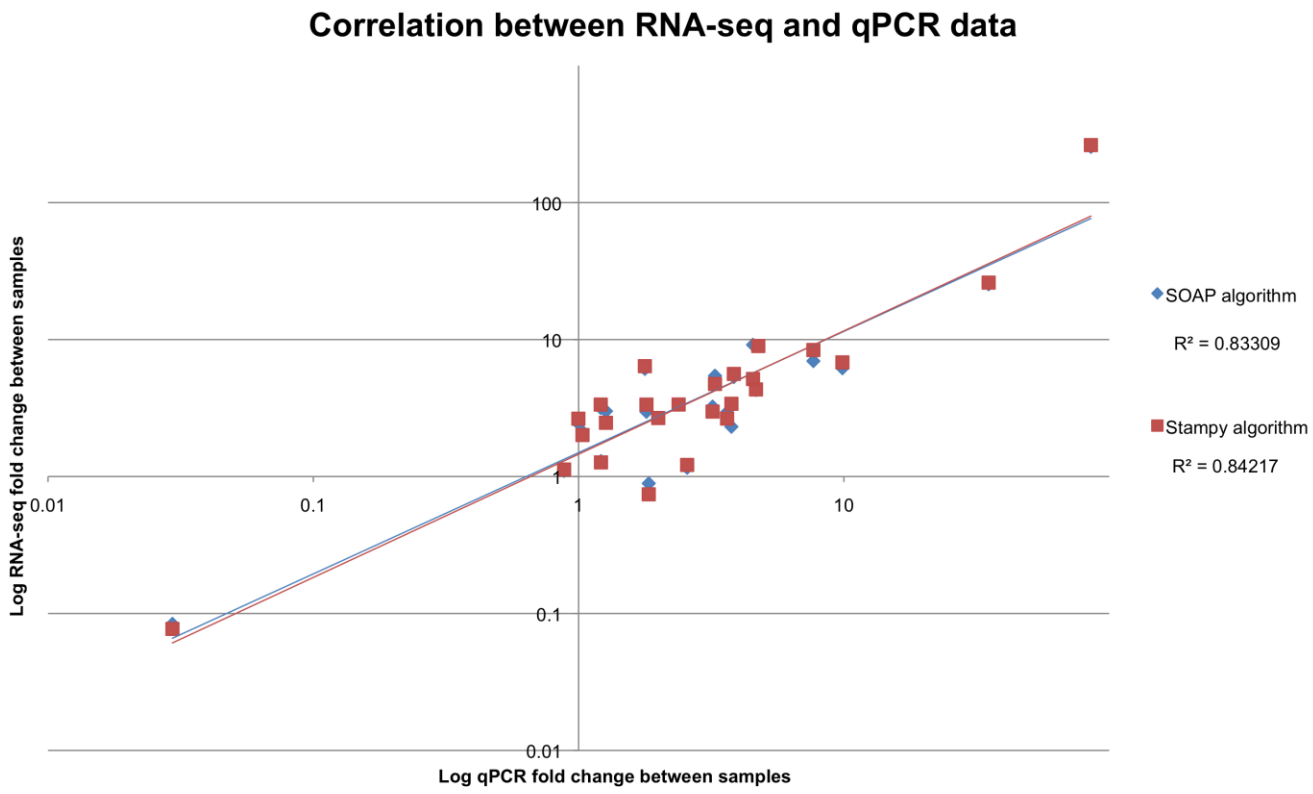


Figure 3.8. qPCR and RNA-seq data show a highly significant correlation. Blue data points show RNA-seq expression values calculated using SOAP2, red data points show RNA-seq expression values calculated using Stampy and Cufflinks.

transcriptomes (Shendure and Ji, 2008). Indeed, there had been only a handful of studies using next-generation sequencing carried out in plants, and only a single mRNA-seq study published in maize (using 454 rather than Illumina sequencing) (Emrich et al., 2007; Lister et al., 2009). As such, it is perhaps unsurprising that initial attempts to quantify the Illumina read data using early versions of different bioinformatic pipelines produced conflicting results (Figures 3.2 and 3.3). Despite these differences, the correlation between expression values predicted by both bioinformatic pipelines and qPCR data suggests that there is still much overlap between the different mapping methods (Figure 3.8). That said, the subset of genes chosen for qPCR were selected due to their relatively high levels of expression in the samples tested, whereas lowly expressed genes appeared to have the most distorted expression estimates in the SOAP2 pipeline (Figure 3.3A). Thus, qPCR amplification of genes with very low levels of expression would be likely to produce a weaker correlation with SOAP2 expression estimates than either Stampy/Cufflinks expression estimates for the same genes, or the SOAP2 correlation shown in Figure 3.8.

Using Stampy and Cufflinks to quantify reads led to a unimodal read distribution (Figure 3.3C), correcting a GC-skew that led to a bimodal read distribution using SOAP2 (Figure 3.3B). It has previously been shown that maize coding sequences have bimodal GC_3 distributions, although the functional significance of this distribution is not clear (Alexandrov et al., 2009; Tatarinova et al., 2010). Due to increased thermostability, GC-rich sequences can be problematic to amplify during PCR, leading to under representation in

PCR amplified Illumina sequence reads (Aird et al., 2011). In addition to correcting for amplification biases (as a consequence of both sequence specific and sequence position effects (Roberts et al., 2011)), Stampy and Cufflinks based read quantification was also more sensitive than SOAP2, mapping an increased number of reads to the maize genome (Table 3.2). Studies into gene expression or genome sequences that utilise next-generation sequencing are rapidly increasing in number and are often outsourced to companies or research facilities independent of the lab where the samples for sequencing were obtained. The results presented here highlight the importance of carrying out downstream bioinformatic validation and analyses 'in-house', and of testing the effects of different bioinformatic pipelines.

As the current study progressed the computational methods available for transcriptome quantification became more advanced. As such, sequencing reads were remapped and quantified using RSEM, and differential expression calculated using DESeq (Anders and Huber, 2010; Li and Dewey, 2011) (these analyses were carried out by Dr Steve Kelly). All the bioinformatic analyses presented in subsequent chapters in this thesis used the expression estimates produced by these programmes. To validate these expression estimates, the same analyses were carried out as described above. Individual gene expression (Figure 3.9) and total sample expression (Figure 3.10) correlated as expected. The correlation between the qPCR data and RSEM expression estimates was also much stronger (R^2 = 0.994, p = 5.55 x 10^{-17}) than the correlations between the qPCR data and either SOAP2 or

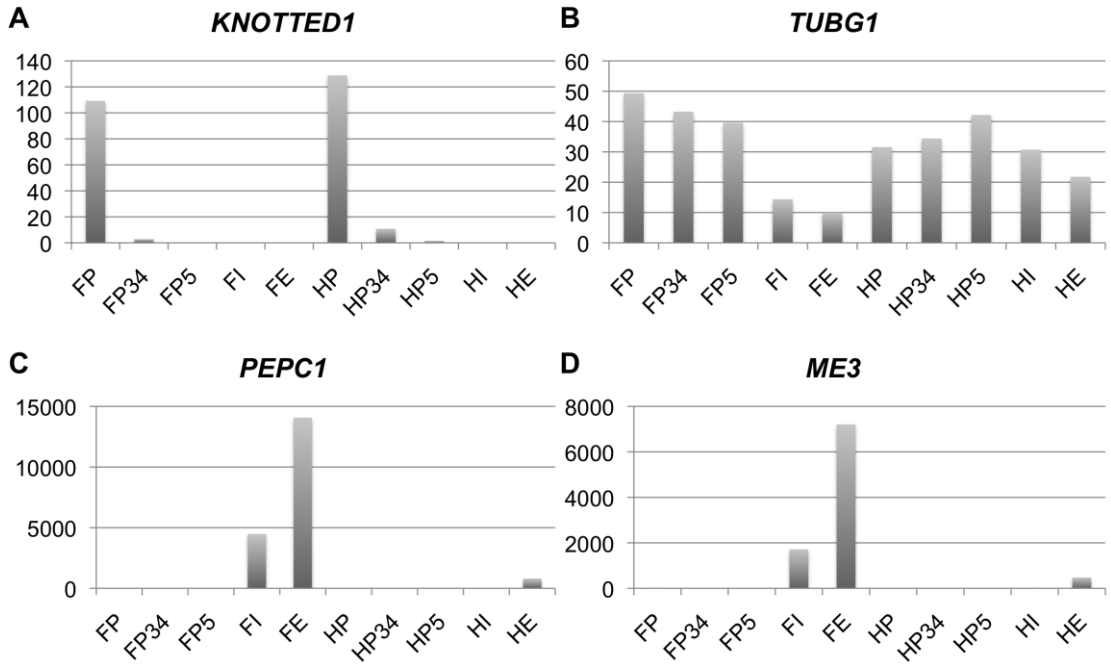
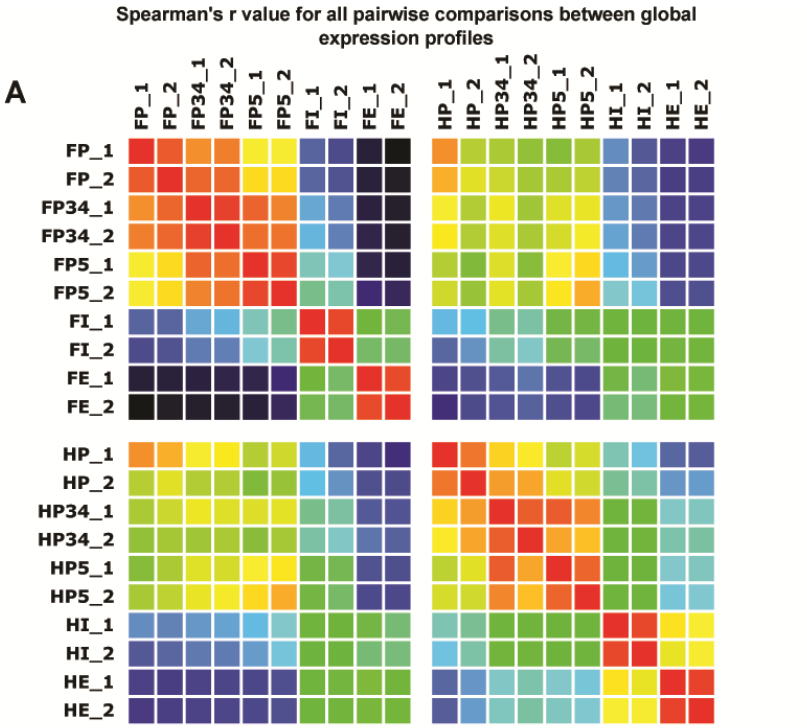


Figure 3.9. Expression of individual genes correlates as expected. A) *KNOTTED1* (GRMZM2G17087), **B)** Gamma-tubulin chain *TUBG1* (GRMZM2G073888) **C)** *PEPC1* (GRMZM2G083841) **D)** *NADP-ME* isoform *ME3* (GRMZM2G085019) Expression values calculated using RSEM, y-axis shows RPKM.



B

Raw values for heatmap shown in A

	FP		FP34		FP5		FI		FE		HP		HP34		HP5		HI		HE	
	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
FP_1	1.00	0.97	0.95	0.95	0.91	0.91	0.68	0.65	0.58	0.55	0.95	0.88	0.88	0.87	0.86	0.88	0.70	0.67	0.63	0.62
FP_2	0.97	1.00	0.96	0.96	0.92	0.92	0.68	0.66	0.58	0.56	0.94	0.90	0.89	0.88	0.88	0.89	0.69	0.68	0.63	0.63
FP34_1	0.95	0.96	1.00	0.98	0.96	0.95	0.70	0.69	0.58	0.57	0.91	0.88	0.91	0.87	0.90	0.91	0.70	0.69	0.63	0.63
FP34_2	0.95	0.96	0.98	1.00	0.96	0.96	0.71	0.69	0.59	0.57	0.92	0.88	0.90	0.88	0.90	0.91	0.70	0.69	0.64	0.64
FP5_1	0.91	0.92	0.96	0.96	1.00	0.98	0.74	0.73	0.59	0.58	0.88	0.85	0.90	0.86	0.91	0.92	0.71	0.70	0.64	0.64
FP5_2	0.91	0.92	0.95	0.96	0.98	1.00	0.76	0.75	0.61	0.60	0.90	0.87	0.90	0.88	0.92	0.94	0.73	0.72	0.66	0.66
FI_1	0.68	0.68	0.70	0.71	0.74	0.76	1.00	0.98	0.79	0.78	0.71	0.71	0.76	0.75	0.78	0.79	0.83	0.82	0.81	0.80
FI_2	0.65	0.66	0.69	0.69	0.73	0.75	0.98	1.00	0.77	0.77	0.68	0.70	0.75	0.73	0.78	0.77	0.81	0.82	0.79	0.79
FE_1	0.58	0.58	0.58	0.59	0.59	0.61	0.79	0.77	1.00	0.98	0.64	0.66	0.67	0.69	0.66	0.65	0.79	0.77	0.85	0.84
FE_2	0.55	0.56	0.57	0.57	0.58	0.60	0.78	0.77	0.98	1.00	0.61	0.65	0.67	0.68	0.66	0.64	0.77	0.77	0.84	0.84
HP_1	0.95	0.94	0.91	0.92	0.88	0.90	0.71	0.68	0.64	0.61	1.00	0.96	0.92	0.92	0.89	0.90	0.74	0.72	0.68	0.68
HP_2	0.88	0.90	0.88	0.88	0.85	0.87	0.71	0.70	0.66	0.65	0.96	1.00	0.94	0.94	0.90	0.90	0.76	0.75	0.70	0.70
HP34_1	0.88	0.89	0.91	0.90	0.90	0.90	0.76	0.75	0.67	0.67	0.92	0.94	1.00	0.97	0.97	0.95	0.80	0.79	0.73	0.74
HP34_2	0.87	0.88	0.87	0.88	0.86	0.88	0.75	0.73	0.69	0.68	0.92	0.94	0.97	1.00	0.94	0.93	0.80	0.80	0.75	0.75
HP5_1	0.86	0.88	0.90	0.90	0.91	0.92	0.78	0.78	0.66	0.66	0.89	0.90	0.97	0.94	1.00	0.96	0.81	0.81	0.73	0.74
HP5_2	0.88	0.89	0.91	0.91	0.92	0.94	0.79	0.77	0.65	0.64	0.90	0.90	0.95	0.93	0.96	1.00	0.80	0.80	0.72	0.73
HI_1	0.70	0.69	0.70	0.70	0.71	0.73	0.83	0.81	0.79	0.77	0.74	0.76	0.80	0.80	0.81	0.80	1.00	0.98	0.92	0.91
HI_2	0.67	0.68	0.69	0.69	0.70	0.72	0.82	0.82	0.77	0.77	0.72	0.75	0.79	0.80	0.81	0.80	0.98	1.00	0.91	0.91
HE_1	0.63	0.63	0.63	0.64	0.64	0.66	0.81	0.79	0.85	0.84	0.68	0.70	0.73	0.75	0.73	0.72	0.92	0.91	1.00	0.98
HE_2	0.62	0.63	0.63	0.64	0.64	0.66	0.80	0.79	0.84	0.84	0.68	0.70	0.74	0.75	0.74	0.73	0.91	0.91	0.98	1.00

Figure 3.10. Pairwise sample correlations using RSEM calculated expression values. A, B) Spearman ranked correlation coefficient for the global mRNA abundance for all pairwise sample comparisons. (A) Correlation shown as a heatmap, scale bar indicates Spearman correlation coefficient. (B) Numerical values used in (A). Both sequencing replicates for each of the 10 samples are shown. Adapted from work carried out by Dr. Steve Kelly, reproduced from Wang et al (2013a).

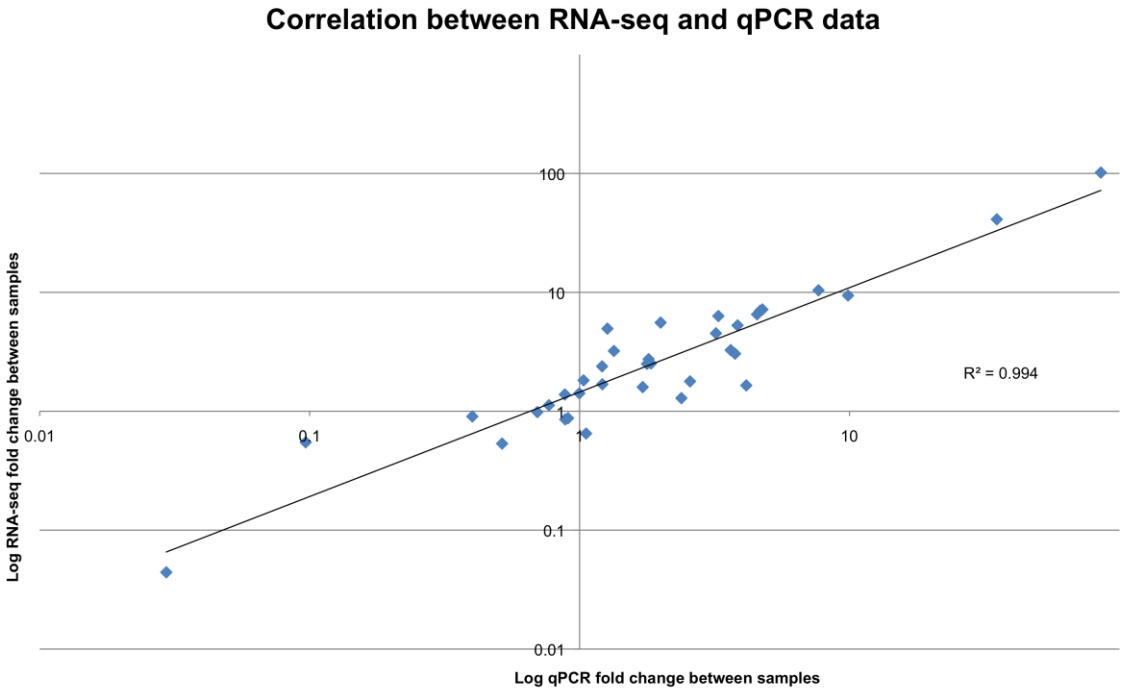


Figure 3.11. **qPCR and RNA-seq data show a highly significant correlation.** Blue data points show RNA-seq expression values calculated using RSEM.

Stampy/Cufflinks based estimates of gene expression (compare Figure 3.8 and Figure 3.11).

3.3.2 Different bioinformatic pipelines affect comparisons of gene expression between samples

The samples shown in Figure 3.1 were selected for sequencing with a view to using gene expression profiles to identify candidate Kranz regulators. When assessing appropriate expression profiles it is important to first define a threshold, or cut off, level of gene expression to be considered in downstream analyses. It is then necessary to define the comparison method to be used between samples. For example comparisons can be based on ratios, absolute expression values or significance testing. The bimodal read distribution originally identified in this study complicated expression analysis because threshold values could be assigned to both distributions and the under-representation of GC-rich genes led to distorted absolute expression values for a subset of genes. In addition, expression ratios could not be calculated for all genes in all samples as all genes were not detected in all samples (i.e. ratios could not be calculated using a zero value). Attempts were made to solve these problems statistically. For example, distribution overlap and means were calculated and genes were assigned to individual distributions with different degrees of statistical confidence. Ultimately these problems were solved computationally through the use of different read mapping and quantification programmes. The use of either Stampy and Cufflinks or RSEM removed difficulties associated with bimodal read distributions. The use of DESeq, which utilises significance testing to calculate differential gene

expression between samples, meant that ratio or threshold calculations were no longer necessary.

3.3.3 Multiple analyses support the validity of the RNA-seq data

Functional characterisation of genes in grass species is a lengthy and time-consuming process. Therefore, it is of great importance to ensure the data used for candidate gene selection are robust. Importantly, technical sequencing replicates produced very similar estimates of gene expression (Figures 3.7, 3.10). In addition, data from individual genes, sample expression profiles and qPCR all supported the validity of this dataset (Figures 3.4, 3.5, 3.8, 3.9, 3.10, 3.11). Based on these validity analyses, this dataset can be used to make reasoned predictions about the genes that regulate Kranz development. This dataset should also provide additional data for gene model refinement and isoform prediction within the maize genetics community (Figure 3.6). With the quality of the data assured, in the next chapter I will discuss how it was used to define transcriptomic profiles of different stages of development and to identify putative Kranz candidate regulators.

Chapter 4: Comparative classification of foliar and husk leaf transcriptomes reveals candidate Kranz regulators

4.1 Introduction

Kranz anatomy is essential for C_4 photosynthesis in almost all known C_4 species (Sage et al., 2011a). However, while much is known about the C_4 pathway and its regulation, the genetic factors that control Kranz development are still poorly understood (Langdale, 2011). Without further knowledge of how Kranz tissue is both initiated and elaborated, recent large-scale international efforts to introduce the C_4 pathway into C_3 crops species stand little chance of success (Hibberd et al., 2008; von Caemmerer et al., 2012).

Improvements in sequencing technologies have enabled advances in the understanding of C_4 leaf development recently. Transcriptome analyses of a maize leaf developmental gradient (Li et al., 2010; Pick et al., 2011), maize BS and M cells (Li et al., 2010; Chang et al., 2012), maize meristematic growth (Takacs et al., 2012), maize germination (Liu et al., 2013) and comparisons of C_3 vs C_4 leaf growth in eudicot genera (Brautigam et al., 2011; Gowik et al., 2011) have all provided insights into the genetic regulatory networks that define C_4 leaves. However, to date there have been no studies of gene expression in the developmental stages most relevant to Kranz

development i.e. P3-5. The work in this chapter describes the transcriptomic signatures of these and other stages of development in foliar and husk leaves. Supervised classification systems of gene expression across a broad developmental context are then used to identify putative regulators of Kranz development.

4.2 Results

4.2.1 Expression of distinct sets of signature genes define stages of foliar and husk leaf development

Five morphologically distinct developmental stages were isolated from both foliar and husk leaves to represent a gradient of maturation across leaf type (Figure 3.1). To determine whether individual stages are associated with unique transcriptomic signatures, sets of signature genes were defined for each stage. Signature genes were identified as those whose relative mRNA abundance was significantly greater ($p < 0.05$) in one sample than in all other samples. Importantly, datasets were normalised to the total number of sequence reads in each sample, thus signatures represent proportional differences in transcript abundance as opposed to differences in overall levels of transcripts accumulated in each tissue. This statistical analysis was carried out by Dr. Steve Kelly and the data are summarised in Table 4.1. The signature genes for each sample were annotated with gene ontology (GO) terms (http://www.geneontology.org), MaizeCyc pathway terms

Sample	No. of signature genes in dataset	Metabolic pathways significantly enriched in signature gene sets
FP	20	
FP3/4	10	
FP5	42	Ribonucleotide synthesis DNA synthesis
FI	1341	Protein synthesis Lipid metabolism Tetrapyrrole biosynthesis
FE	1479	Light harvesting reactions C_4 carbon assimilation Gluconeogenesis Photorespiration
HP	24	
HP3/4	10	
HP5	26	Lipid biosynthesis
HI	344	Cell wall biosynthesis
HE	713	2^o metabolite biosynthesis Amino acid degradation Lipid degradation

Table 4.1. Different metabolic pathways are enriched in the signature gene sets of different samples. Signature genes are those whose transcripts were detected at significantly higher levels in the named sample compared to the other nine samples ($p < 0.05$). Statistical analysis carried out by Dr. Steve Kelly.

(http://www.maizecyc.maizegdb.org) Mapman terms (http://www. http://mapman.gabipd.org) and Pfam domains (http://www. http://pfam.sanger.ac.uk) and are shown in Appendices 4.1-4.10 (annotation carried out by Dr. Steve Kelly).

Table 4.1 shows a general increase in the number of signature genes identified as tissues mature and become more metabolically complex, with the highest number of signature genes found in the most photosynthetically active sample FE (1479). The one exception is seen between FP/HP and FP3/4/HP3/4 stages where the number decreases from ~20 to 10. This decrease probably reflects the inclusion of meristematic tissue in the FP and HP samples such that suites of meristem-specific genes are represented in the FP and HP, but not in the FP3/4 and HP3/4, transcriptomes.

Due to the paucity of signature genes in the earliest developmental stages, no metabolic pathways were identified as significantly enriched in the FP/HP and FP3/4/HP3/4 samples (Table 4.1, Appendices 4.1, 4.2, 4.6, 4.7). Analyses of GO term enrichment indicate transcriptional activity in both FP and HP samples, and also in the FP5 sample (Appendix 4.3). The shared terms between FP and HP, and the lack of corresponding terms in FP3/4 and HP3/4 samples (Appendix 4.11), suggest that the transcriptional activity represented in FP/HP tissues is associated with meristem function, as opposed to leaf formation. The GO term enrichment for transcriptional activity is not seen in HP5, suggesting that by P5 foliar and husk leaf differentiation is already distinct. Further evidence that distinct developmental trajectories are already

apparent by P5 comes from analysis of transcription factor (TF) families that are expressed in each leaf type (Figure 4.1). 22 genes from 10 TF families are expressed in the signature gene set of FP5 whereas only 13 TFs are enriched in HP5.

Both shared and organ specific functionality is manifest in the FI/HI and FE/HE samples. GO term enrichments for these four tissues are shown in Figure 4.2. In combination with pathway enrichment data (Appendices 4.4, 4.5, 4.9, 4.10, 4.12), analyses of GO term enrichments show that cell wall biosynthesis occurs in both the FI and HI samples, however, enriched protein synthesis, chlorophyll biosynthesis, electron transport activity and lipid metabolism are specific to foliar leaves. The FE and HE samples are both enriched for components of C_4 photosynthesis, but FE is specifically enriched for post-translational modification and the HE sample is enriched for protein and lipid mobilisation. In summary, the profiles of the 10 samples are consistent with a foliar leaf trajectory that transitions from proliferating, non-photosynthetic primordia (P1-5), through the induction of photosynthetic competence and chloroplast biogenesis (FI) to a fully photosynthetically active leaf (FE). A similar trajectory is also apparent in early husk leaves (P1-4), however, the process of tissue differentiation is delayed compared to the foliar leaf (HP5 vs FP5). The HI profile is consistent with a non-photosynthetic organ that is expanding to surround the ear while the HE profile is that of a weakly photosynthetic tissue that has started to senesce.

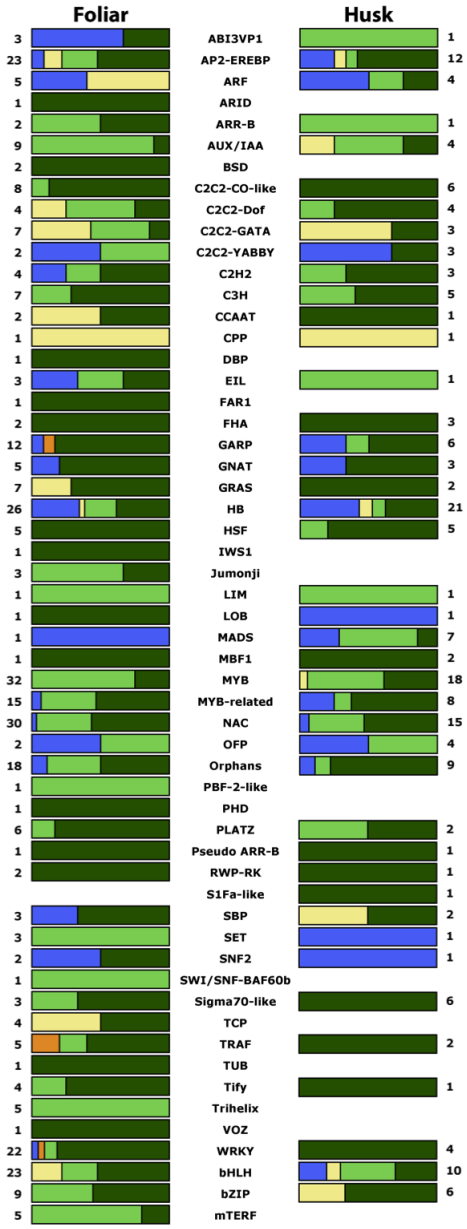


Figure 4.1. Transcription factor complements in the signature genes of foliar and husk leaf samples. Numbers represent the total number, and bar length denotes the relative proportion, of gene family members represented in each foliar or husk sample. FP/HP samples are depicted as purple bars; FP3/4 & HP/3/4 – orange bars, FP5/HP5 – yellow bars; FI/HI – light green bars; FE/HE – dark green bars. Produced in collaboration with Dr. Steve Kelly, reproduced from Wang et al (2013a).

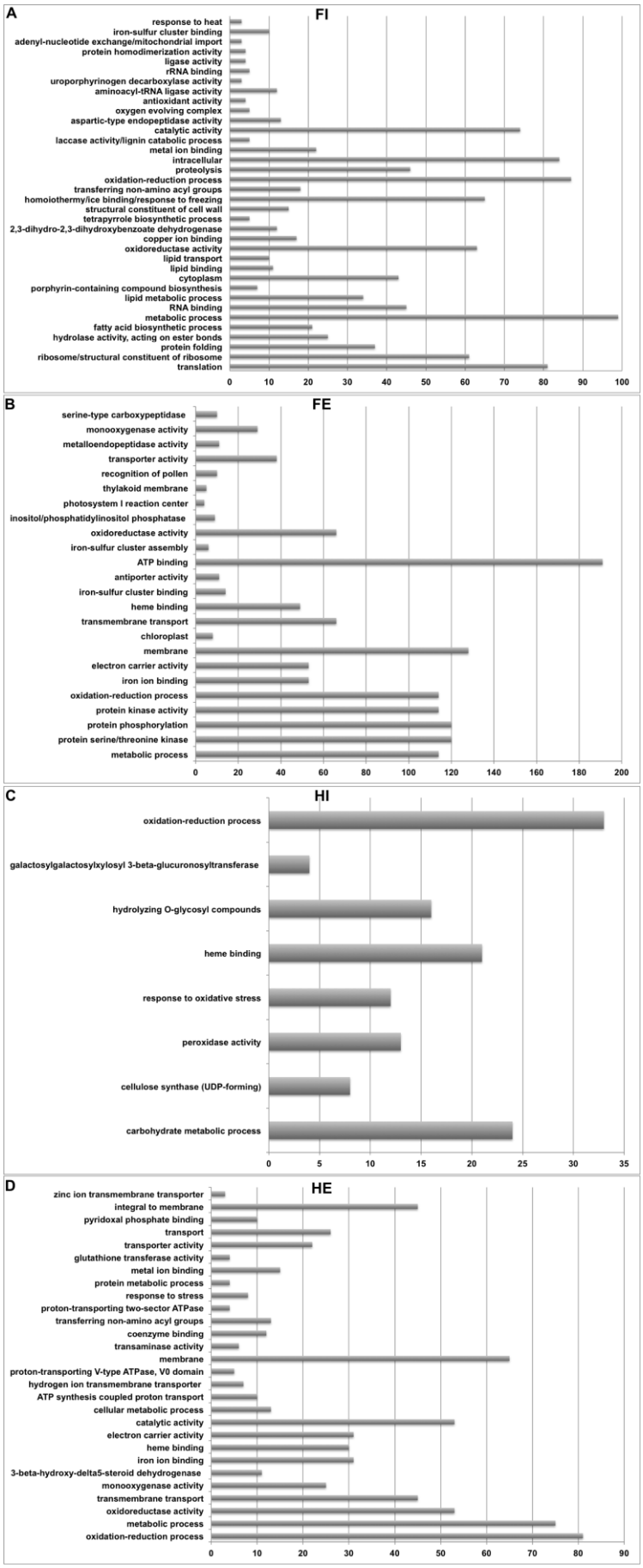


Figure 4.2. Enriched GO term annotations in a subset of signature gene sets. A) FI, **B)** FE, **C)** HI, **D)** HE. GO term enrichment statistically significant ($p < 0.05$). X-axis represents number of genes annotated with that GO term.

4.2.2 Classification of primordia gene expression into profiles identifies distinct developmental trajectories

To identify genes that regulate the earliest patterning processes of leaf development, genes expressed in the primordia samples were assigned to one of seven predetermined expression profiles (Figure 4.3). Profiles were designed to filter genes that either decrease in expression from P1-P5 in a single leaf type (descending profiles D1, D2, D3), or increase from P1-P5 in a single leaf type (ascending profiles A1, A2, A3). Significant ($p < 0.05$) changes in gene expression between primordia samples were necessary for inclusion of a gene in a specific profile. Genes that showed no significant change were assigned to a distinct neutral profile (N). The computational assignation of genes into profiles was carried out by Dr. Steve Kelly. Annotated gene lists for the seven profiles for both foliar and husk leaves are shown in Appendices 4.13-4.26.

As expression in the descending profiles decreases from the FP/HP to FP5/HP5 samples, these profiles were designed to identify genes that regulate processes such as meristem maintenance, leaf initiation and early axial patterning. Conversely, as expression in the ascending profiles increases from the FP/HP to FP5/HP5 samples, these profiles were designed to identify genes that regulate processes that occur slightly later than in the D profiles, such as lateral and intermediate vein formation, plastid biogenesis and cell type specification. Importantly, manual inspection of these profiles revealed that genes with known roles in leaf patterning were assigned to

A

P P3/4 P5	No. of genes/ TFs in F cluster	TF groupings	No. of genes/ TFs in H cluster	TF groupings
D1	56/12	2 FD1 only 6 FD1/HD1 4 FD1/HD2	13/7	6 HD1/FD1 1 HD1/FD3
D2	178/20	12 FD2 only 8 FD2/HD2	175/28	12 HD2 only 4 HD2/FD1 8 HD2/FD2 4 HD2/FD3
D3	413/50	42 FD3 only 1 FD3/HD1 4 FD3/HD2 3 FD3/HD3	157/22	18 HD3 only 3 HD3/FD3 1 HD3/FA3
A1	171/25	13 FA1 only 1 FA1/HA1 8 FA1/HA2 3 FA1/HA3	60/4	1 HA1/FA1 3 HA1/FA3
A2	463/59	37 FA2 only 21 FA2/HA2 1 FA2/HA3	573/64	31 HA2 only 8 HA2/FA1 21 HA2/FA2 4 HA2/FA3
A3	580/59	41 FA3 only 3 FA3/HA1 4 FA3/HA2 10 FA3/HA3 1 FA3/HD3	469/33	19 HA3 only 3 HA3/FA1 1 HA3/FA2 10 HA3/FA3
N	26160/ 1263	Not determined	28467/ 1457	Not determined

B

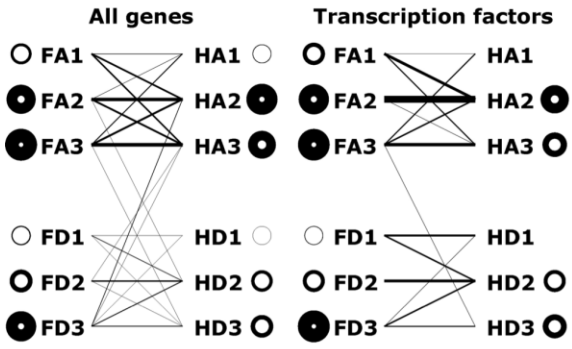


Figure 4.3. **Comparative profile analysis of transcriptomes in developing leaf primordia.** Figure legend is on opposite page.

Figure 4.3. Comparative profile analysis of transcriptomes in developing leaf primordia. A) Table showing schematics of descending (D1-D3), ascending (A1-A3) and neutral (N) transcriptome profiles. The total number of genes in each foliar (F) and husk (H) profile is shown along with the number of genes encoding transcription factors (TFs). The specificity of TFs with respect to each profile is shown, along with reference to the number of TFs that are shared with other profiles. **B)** Schematics showing the extent of overlap between all ascending and descending profiles with respect to all genes (left) and genes encoding transcription factors (right). The circles adjacent to profile names indicate the number of genes specific to that profile, and the lines between profiles indicate the number of genes shared between profiles. In each case, the thickness of the line/circle perimeter denotes the number of shared/unique genes (thicknesses are normalised to the maximum number off genes/TFs). Figure produced in collaboration with Dr. Steve Kelly, reproduced from Wang et al (2013a).

profiles as expected (Table 4.2). For example, *KNOTTED1-LIKE HOMEOBOX* (*KNOX*) genes, that are known to regulate both shoot and axillary meristem maintenance (Jackson et al., 1994), are present in the descending profiles. The same is true for the majority of genes that regulate abaxial/adaxial patterning in the leaf (Husbands et al., 2009). An exception is *KAN1*, which is present in both the FA2 and HA2 profiles, however, this gene has been shown to fall within a grass *KAN* clade that is expressed in the later stages of abaxial development (Candela et al., 2008; Zhang et al., 2009). Genes implicated in mediolateral and proximodistal patterning are also expressed as expected (i.e. mediolateral specification occurs prior to proximodistal specification). An observed difference in transcript accumulation of the *LIGULELESS2* gene (*LG2*), which regulates ligule formation in the developing leaf (Walsh et al., 1998), between foliar and husk leaves is consistent with the lack of ligule formation in husk leaves. The accumulation patterns of chloroplast regulatory genes *GOLDEN2* (G*2*) and *G2-LIKE1* (*GLK1*) are also consistent with the loss of *g2* function phenotype and the proposed role for *GLK1* in specifically regulating C_4 chloroplast development (Langdale and Kidner, 1994; Rossini et al., 2001). Taken together, these results validate the identification of candidate leaf developmental regulators based on profile-analysis.

Analysis of the GO terms enriched in each profile reveals that, with one exception, all the enriched terms in both the foliar and husk descending profiles relate to the regulation of transcription (compiled GO term enrichments for each profile are shown in Appendix 4.27). The exception is

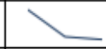
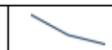
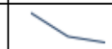
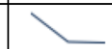
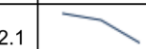
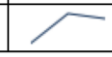
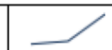
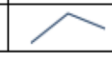
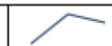
		F Cluster	H Cluster	FP	FP3/4	FP5	F Trendline	HP	HP3/4	HP5	H Trendline
Meristem maintenance											
GRMZM2G017087	*KN1*	FD1	HD1	109.3	2.9	0.0		128.8	10.8	1.5	
GRMZM2G028041	*RS1*	FD1	HD1	37.7	2.1	0.0		66.6	9.3	1.6	
GRMZM2G452178	*GN1*	FD2	HD2	31.4	0.9	0.0		28.6	2.0	0.4	
GRMZM2G087741	*LG3*	FD1	HD1	52.7	2.7	0.0		67.9	8.2	1.7	
GRMZM2G094241	*LG4A*	FD1	HD1	27.7	2.1	0.1		58.4	7.3	1.4	
GRMZM5G832409	*LG4B*	FD2	HD2	5.3	0.3	0.0		14.7	2.8	2.1	
Lateral organ initiation and adaxial/abaxial patterning											
GRMZM2G088309	*DL1*	FD1	HD1	85.0	42.0	22.0		104.2	12.5	1.1	
GRMZM2G403620	*RS2*	FN	HN	113.8	76.0	64.5		95.2	74.6	75.3	
GRMZM2G441325	*ARF3*	FD2	HD2	67.4	28.8	26.6		45.5	15.7	22.1	
GRMZM2G074543	*ZYB9*	FD3	HD3	69.7	37.8	0.8		88.2	74.3	22.1	
GRMZM2G056400	*KAN1*	FA2	HA2	2.4	7.9	7.1		0.0	3.1	3.9	
Mediolateral patterning											
GRMZM5G892991	*RGD2*	FD1	HN	64.4	29.0	11.7		90.3	45.5	23.4	
GRMZM2G069028	*NS1*	FN	HN	7.4	5.1	1.7		7.9	3.6	1.3	
Proximodistal patterning											
GRMZM2G036297	*LG1*	FA3	HA1	0.0	0.1	1.9		0.2	6.4	24.2	
GRMZM2G060216	*LG2*	FA2	HN	3.8	15.3	9.6		40.2	20.9	7.6	
Chloroplast biogenesis											
GRMZM2G026833	*GLK1*	FA2	HN	0.2	4.2	3.1		0.3	0.2	0.0	
GRMZM2G087804	*G2*	FN	HN	25.9	42.9	62.8		44.4	71.7	82.3	

Table 4.2.. **mRNA abundance dynamics for regulatory genes known to be associated with meristem maintenance, axis patterning in the leaf and chloroplast biogenesis.** Columns show gene ID, foliar and husk profile ID, average transcript reads per kilobase per million (RPKM) in FP, FP3/4, FP5, HP, HP3/4 and HP5 samples, and non-normalised trendlines of transcript abundance. Discrepancies between profile ID and trendline trajectories (e.g. *RS*2 is in FN but read data shows descending trajectory) are due to the fact that profiles are defined by significant differences between samples as opposed to general trends, reproduced from Wang et al (2013a).

the enrichment of the GO term for ‘trehalose biosynthetic process’ in the HD2 profile. This profile contains transcripts that are detected at significantly higher levels in meristem-containing HP samples than in either HP3/4 or HP5 samples. As such, the observation is consistent with the role of the *RAMOSA3* gene and the trehalose metabolic pathway in branching of the maize inflorescence meristem (Satoh-Nagasawa et al., 2006). Enriched GO terms in ascending profiles are more varied than those in descending profiles (Appendix 4.27). In the FA2 and HA2 profiles, where transcript levels are equivalent in P3/4 and P5 leaf samples but are higher in those samples than in meristem-containing FP/HP samples, enriched GO terms are related to metabolism (e.g. steroid biosynthesis, cell wall). In contrast, photosynthesis-related terms are enriched in FA1, FA3 and HA3 profiles, where transcript levels are significantly higher in P5 than in P3/4 primordia. This pattern clearly reflects a developmental trajectory towards photosynthetic competence. The FA3 and HA3 profiles, however, are quite distinct. Whereas the HA3 profile contains over 30 enriched GO terms, reflecting a broad range of biological processes, the FA3 profile contains a majority of terms related to transcription factor activity. This difference presumably reflects a more complex differentiation process between P3/4 and P5 in foliar as opposed to husk leaf primordia. In addition to the distinct patterns of GO term enrichment that foliar and husk leaves exhibit, the unique developmental trajectories that the two leaf types follow from P3/4 onwards is reflected in the increasingly unique sets of genes identified in each leaf type during leaf maturation (Figure 4.3). Noteworthy TFs identified in each profile are outlined in the sections below. Manual annotations for all the TFs identified in the ascending and descending

profiles are shown in Appendix 4.29 and the information is summarised in Table 4.3.

4.2.3 Expression profile D1: meristematic function

The D1 profiles (Appendices 4.13, 4.20) contain transcripts that decrease from FP/HP (meristem, P1-P2) to FP3/4/HP3/4, and from FP3/4/HP3/4 to FP5/HP5 (i.e. decrease from samples that contain meristems and leaf tissue to samples that contain only leaf tissue), and therefore it was expected that D1 profiles would identify regulators of meristematic growth. Four class I *KNOX* genes, which regulate meristematic growth (Jackson et al., 1994; Hay and Tsiantis, 2010), are present in both the FD1 and HD1 profiles. Homologues of the rice *DROOPING LEAF* (*DL*) gene, which acts to regulate midvein formation (Yamaguchi et al., 2004), and the *DL* repressor *OsMADS32* (Sang et al., 2012), were also identified in both D1 profiles. Four TFs were identified in both the FD1 and HD2 profiles, suggesting that conserved developmental mechanisms operate over a different timeframe in the two different leaf contexts. Of these genes, two are homologues of *Arabidopsis* genes that regulate photomorphogenesis (*ATHB4* (Sorin et al., 2009); *HLS1/COP3* (Li et al., 2004)) and one is a homologue of an *Arabidopsis* gene that functions in RNA directed DNA methylation (*DRD1-like* (Kanno et al., 2004)). The fourth gene plays a redundant role in maize spikelet meristem formation (*SISTER OF INDETERMINATE SPIKELET1* (*SID1*) (Chuck et al., 2008)), however, it has not previously been implicated in vegetative meristem

Cluster	Developmental processes associated with profile	TFs in F cluster	TFs in H cluster
D1	Meristem function Axial patterning in leaf	KN1, RS1, LG3, LG4, SID1, DL, MADS32-like	KN1, RS1, LG3, LG4, DL, MADS32-like, KAN5
D2	Meristem function Axial patterning in leaf Auxin signalling Epidermal patterning Ethylene signalling Floral transition Chromatin remodelling	WOX9B, NAM2, BEL-like, OVATE-like, WOX3A, RBR-like, KAN-like, LG2-like, ARF4, ARF11, ARF29, ERF4-like, EIN3-like	WOX9B, NAM2, BEL-like, OVATE-like, ABPH1, SID1, KAN-like, DL-like, ARF4, ARF29, IAA23, SPH-like, LHY, CO-like5, MADS1, DDM1-like, SUVH-like
D3	Dividing primordia Axial patterning in leaf Auxin signalling Epidermal patterning Phase change Stress response Floral transition Axillary Branching Integrating signals	ANT-like, PAN-like, ULT-like, AS2, ZFL1, ZMM16, ZYB9, KAN-like, DL-like, IAA2, IAA7, IAA26, SPH-like, SCRM-like, MIXTA-like, ZFL1, SBP-like	ANT-like, CUC-like, ZYB9, IAA6, SNAC-like, DREB-like, MYB, NAP-like, WRKY, CO-like6, U2AF34, TB1-like, GT1, NAC2-like, LAF1-like
A1	Cell division/expansion Axial patterning in leaf Epidermal patterning Organ outgrowth Photomorphogenesis Plastid biogenesis Integrating signals	PLT-like, TSO-like, SPCH1, SPCH2, CRF-like, ATHB16-like, DIV-like, TCP-like, CO-like, OBP3-like, GATA2-like, GNC/CGA1-like, BME3-like	LG1, GATA2-like, BME3-like
A2	Cell division/expansion Axial patterning in leaf Auxin signalling Epidermal patterning Organ outgrowth Photomorphogenesis Plastid biogenesis Vascular differentiation Stress response Floral transition Senescence	RAV-like, MYB-like, SCZ-like, GRF-like, KAN1, LG2, LG2-like, HDZIV3, HDZIV13, AS2-like, TCP-like, YAB-like, DIV-like, CO-like, OBP3-like, GLK1, GNC/CGA1-like, SIG70-like, XND1-like, DOT5-like, APL-like, SHR	RAV-like, MYB-like, SCZ-like, NAC, KAN1, ATHB4-like, LG2-like, IAA24, HDZIV3, ATHB16-like, AS2-like, YAB-like, TCP-like, DIV-like, NAC, CO-like, OBP3-like, CCA1-like, ERF-like, DREB-like, STZ-like, MYB, NAC, NF-YA3, NF-YA1, VRS1-like, ZMM31, MADS3, ZAP1, WRKY
A3	Axial patterning in leaf Auxin signalling Epidermal patterning Organ outgrowth Photomorphogenesis Plastid biogenesis Vascular differentiation Mucilage/Lignin Integrating signals	LG1, ARF9, ARF16, ARF18, ARF22, IAA6, IAA13, IAA14, IAA24, IAA30, IAA31-like, SPCH2-like, SCRM1-like, SCRM2-like, MIXTA-like, CRF-like, EIL1-like, ATHB16-like, ATHB13-like, GATA2-like, GATA12-like, PDE191, CIA2-like, APL-like, SHR-like, HAT14-like	LG1-like, ZmARF20, IAA14, SCRM1-like, MIXTA-like, SHN1-like, ATHB13-like, ATHB1-like, GATA2-like, MYB, HAT14-like, PRE1-like

Table 4.3. Overview of transcription factors identified in each primordia expression profile. Overview is not exhaustive, in that not all transcription factors identified in profiles are shown. No indication of the number of appearances of a particular transcription family in a profile is implied, table produced in collaboration with Prof Jane Langdale.

formation. Interestingly, *SID1* was shown to be specifically expressed in the BS of expanded maize leaves (Chang et al., 2012). Two further genes that are specific to the FD1 profile may play novel roles in meristem maintenance: one (GRMZM2G151223) has been annotated as maize *HISTIDINE KINASE 1* (*ZmHK1*) (Yonekura-Sakakibara et al., 2004) and is part of a gene family known to be involved in cytokinin signal transduction (Muller and Sheen, 2007); the other (GRMZM2G065496) is part of the Reproductive Meristem (REM) subfamily of B3 domain transcription factors, named after the *Brassica oleracea* gene *REM1*, which is specifically expressed in reproductive meristems (Wang et al., 2012a).

4.2.4 Profile D2: meristem function and early leaf development

The D2 profiles also select for genes that are expressed most strongly in the meristematic FP/HP samples, however, in these profiles there is no significant decrease in expression between P3-4 and P5. Hence the D2 profiles are expected to identify genes that play roles in both the meristem and early leaf development (Appendices 4.14, 4.21). As such the D2 profiles are not enriched in *KNOX* genes, however, genes known to be expressed in the meristem such as *NAM2* (Zimmermann and Werr, 2005) and *WOX9B* (Nardmann et al., 2007) are present in both FD2 and HD2. Additional genes found in both the FD2 and HD2 profiles include homologues of *Arabidopsis* genes that regulate meristem specification (*BEL-like* (Rutjens et al., 2009)), axial patterning (*KAN-like* (Kerstetter et al., 2001)) and auxin signaling (*ZmARF4, ZmARF29* (Xing et al., 2011)). This overlap demonstrates

conserved mechanisms that regulate meristematic growth and early primordia development in both foliar and husk leaves. However, developmental specificity exists between the two leaf types and is manifest in the 40-60% of TFs that are specific to either FD2 (60%) or HD2 (40%) (Figure 4.4A, 4.4C). This specificity is apparent both in the unique members of TF families that are shared among FD2 and HD2 (i.e. different members of the *OVATE* family, which repress *KNOX* genes in *Arabidopsis* (Wang et al., 2011), are present in FD2 and HD2) and the unique TF families that are represented in the foliar and husk D2 profiles. The presence of different TF families implies a shift in the biological processes that are taking place in the two leaf types. Examples of this include the presence of ethylene response homologues only in the FD2 profile (*AtEIN3-like* and *OsERF4* (Guo and Ecker, 2004)). Likewise, the HD2 profile features homologues of *Arabidopsis* genes that function in chromatin remodelling (*DDM1-like* (Verbsky and Richards, 2001); *SUVH-like* (Baumbusch et al., 2001)) and the floral transition (*LHY*, *CO* (Putterill et al., 2004)) as well as maize genes known to operate in reproductive phase change (*ZmMADS1* (Heuer et al., 2001)). The specificity of these developmental programmes thus reflects distinct developmental trajectories: active vegetative growth in the foliar meristem/primordia and floral induction in the husk meristem/primordia.

4.2.5 Profile D3: early leaf development

The D3 profiles were designed to filter genes that are active in developmental processes in leaf primordia prior to P4, as mRNA accumulation is equivalent in P and P3/4 samples, but significantly decreases in the P5 samples (Appendices 4.15, 4.22). Only three TFs are shared between the FD3 and HD3 profiles (although there is overlap between the D3 profiles with other profile types (Figure 4.3)): two genes closely related to *AINTEGUMENTA* (*ANT*), which has been shown to regulate the rate of cell division in leaf primordia (Mizukami and Fischer, 2000) and the axial patterning gene *ZYB9* (Juarez et al., 2004). TFs specific to FD3 include further axial patterning genes (*ZmKAN5*, (Zhang et al., 2009); *ZmDL2* (Yamaguchi et al., 2004; Ishikawa et al., 2009), homologues of rice and *Arabidopsis* genes expressed in dividing primordia (*ULT-like* (Carles et al., 2005); *PAN-like* (Chuang et al., 1999); *IG1* (Evans, 2007); *ZMM16* (Munster et al., 2001); *ZFL1* (Bomblies et al., 2003)), homologues of epidermal patterning genes (*SPCH-like*, *SCRM-like* (Pires and Dolan, 2010); *MIXTA-like* (Dubos et al., 2010)) and auxin responsive genes (*ZmIAA4*, *ZmIAA16*, *ZmIAA26* (Wang et al., 2010)). In contrast, HD3 features homologues of genes that function in drought/ABA/stress responses (*SNAC-like*, *NAP-like* (Zhu et al., 2012); *MYB* (Dubos et al., 2010); *WRKY* (Wei et al., 2012); *TINY-like* (Sun et al., 2008)), light signalling (*LAF1-like* (Ballesteros et al., 2001)), flowering time (*U2A35* (Wang and Brendel, 2006); *CO-like* (Putterill et al., 2004)) and axillary bud outgrowth (*GT1* (Whipple et al., 2011); *TB1-like* (Martin-Trillo and Cubas, 2010); *OsNAC2* (Mao et al., 2007)). The HD3 profile is thus consistent with that of an axillary meristem that has transitioned into reproductive development (see extended inflorescence meristem in Figure 3.1A). Over

80% of the TFs identified in the D3 profiles are specific to each profile (Figure 4.3). This presumably reflects the distinct developmental mechanisms that function in early development in foliar and husk leaves. Of these TFs it was not possible to assign putative functions for many through homology informed analyses (Appendix 4.29). As such, these may represent novel regulators of pre-P4 leaf development that distinguish foliar and husk type differentiation. Putative novel regulators include ABI3VP1, MYB, bHLH, SBP, SRS and Zinc Finger proteins (specific to FD3) and PLATZ proteins that are present in both FD3 and HD3 profiles, but with distinct family members in each.

4.2.6 Profile A1: organ identity, outgrowth and metabolic function

Gene expression in the A1 profiles significantly increases in consecutive samples from P to P3/4 to P5 (Appendices 4.16, 4.23). These profiles were thus designed to select genes that regulate the establishment of organ identity and expansion and the onset of derived metabolic activity. Only four TFs were identified in HA1. Of these one is shared with FA1, a BME1-like GATA protein that may play a role in the integration of environmental and developmental signals (Reyes et al., 2004; Liu et al., 2005; Manfield et al., 2007). The other three HA2 TFs were also identified in the FA3 profile. Similarly, overlap is also found between the FA1 and HA2/HA3 profiles. This pattern is indicative of shared developmental mechanisms that exhibit variant timing in different leaf contexts. Of the genes that are found in FA1 and HA2/HA3, most are homologues of TFs involved in organ differentiation/outgrowth (*AmDIV-like*

(Galego and Almeida, 2002); *ATHB16-like* (Harris et al., 2011); *TCP-like* (Martin-Trillo and Cubas, 2010); *TSO-like* (Hauser et al., 1998)) and photomorphogenesis (e.g. *OBP3-like* (Ward et al., 2005); *GATA2-like* (Reyes et al., 2004)). Of the five genes in the FA1 profile that encode TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR (TCP) transcription factors, three are in the *PROLIFERATING CELL FACTOR* (*PCF*)*-like* clade (which is associated with regulating cell proliferation) and two are in the *CINCINNATA* (*CIN*)*-like* clade (which is associated with leaf margin specification) (Martin-Trillo and Cubas, 2010). Both *CIN-like* genes and one of the *PCF-like* genes are common to FA and HA profiles whereas two *PCF-like* genes are foliar specific. Similarly, the *OBP3-like* and *GATA–like* genes, which are implicated in photomorphognenesis (Reyes et al., 2004; Ward et al., 2005), are also enriched in foliar ascending profiles relative to husk ascending profiles. These observations suggest greater complexity in the differentiation and photomorphogenetic programmes that operate in foliar, as opposed to husk, primordia. The increased complexity of early foliar leaf development is also reflected in the range of foliar specific genes identified in the FA1 cluster. These genes include homologues of TFs that are required for cell division/expansion (e.g. *PLT-like* (Dhonukshe et al., 2012); *CRF-like* (Rashotte et al., 2006)), epidermal patterning (*ZmSPCH1, ZmSPCH2* (Liu et al., 2009)) and chloroplast biogenesis (*GNC/CGA1-like* (Reyes et al., 2004)) plus three novel genes. The novel genes include a forkhead domain family member (GRMZM2G177895) that GO term analysis suggests may be involved in chloroplast biogenesis, and a bHLH TF (GRMZM2G164341) that

is in the same clade of bHLH genes as *INDEHISCENT,* which functions in margin specification (Liljegren et al., 2004; Pires and Dolan, 2010).

4.2.7 Profile A2/A3: leaf venation patterning and cell type specification

A2 and A3 profiles were designed to capture genetic regulators of vascular initiation and of the elaboration and specification of cell types. The profiles comprise transcripts that accumulate to a higher level in P3/4 and P5 than P samples (A2, Appendices 4.17, 4.24) or to a similar level in the P and P3/4 samples but significantly increase in P5 (A3, Appendices 4.18, 4.25). As lateral veins continue to extend and intermediate veins emerge during P4 and P5 (Figure 3.1C,D,H,I), the presence of homologues of regulators of xylem (e.g. *XND1* (Zhao et al., 2008)) and phloem (e.g. *APL1* (Bonke et al., 2003)) differentiation in the A2 and A3 profiles was expected (Table 4.3). Similarly, the enrichment of *AUXIN RESPONSE FACTOR* (*ARF*) and *AUX*/*IAA* genes in FA3 is expected given the known role of auxin in vascular differentiation ((Mockaitis and Estelle, 2008)). However, the majority of genes that regulate vascular patterning and BS and M cell differentiation, key events in the onset of Kranz development that coincide with the FA2 profile, remain unknown. As such, it was expected that analysis of the TFs present in the FA2 profile would identify previously uncharacterised or unannotated genes.

Inspection of the FA2 profile reveals sixteen genes of note (specific to FA2) that may play key roles in vascular patterning and cellular differentiation

(Figure 4.4A, Appendix 4.17, 4.29). First, there are three bHLH genes that, given the known roles of bHLH genes in a wide range of cellular patterning processes (Pires and Dolan, 2010), may regulate foliar specific cellular differentiation. One of these genes is found in the same clade of bHLH genes as a target of the auxin regulated TF *MONOPTEROS*, which is known to regulate vascular differentiation in *Arabidopsis* (Schlereth et al., 2010). Second, there is one homologue of an *Arabidopsis* bZIP gene and three homologues of *Arabidopsis DoF* genes (two share the same clade as *Dof AFFECTING GERMINATION* (*DAG*) and one shares a clade with *HIGH CAMBIAL ACTIVITY 2* (*HCA2*) (Wang et al., 2013a)) that are all known to be expressed in leaf vascular tissue (Gualberti et al., 2002; Gibalova et al., 2009; Guo et al., 2009; Gardiner et al., 2010). Notably, the two *DAG-like* genes were found to be expressed specifically in the BS cells of maize foliar leaves (Li et al., 2010). Third, two co-orthologues of the *Arabidopsis SHORTROOT* (*SHR*) gene that is known to function in radial patterning of the root, hypocotyl and stem vasculature (Fukaki et al., 1998; Helariutta et al., 2000) are present in FA2. Fourth, seven C2H2 zinc finger (ZnF) proteins are of particular note, Two are homologues of the *Arabidopsis* gene *DEFECTIVELY ORGANISED TRIBUTARIES 5* (*DOT5*), which has been shown to regulate leaf vascular patterning (Petricka et al., 2008), and three are closely related to the *INDETERMINATE1* (*ID1*)-like (Colasanti et al., 1998) genes *SHOOTGRAVITROPISM 5* (*SGR5*) (Morita et al., 2006) and *JACKDAW* (*JKD*) (Welch et al., 2007) which function in the *SHR* pathway in *Arabidopsis*. The final two C2H2 ZnF proteins are closely related to maize MYB-related protein 1-interacting (MRPI) proteins (Royo et al., 2009). Intriguingly, the

A

FA2			FA3		
GRMZM2G027068	bHLH family	AtTMO5 clade	GRMZM2G073750	ARF family	ZmARF9
GRMZM2G045883	bHLH family	Subfamily Ia (SPCH-like)	GRMZM2G028980	ARF family	ZmARF16
GRMZM2G081816	bHLH family	Subfamily VIIIb	GRMZM2G035405	ARF family	ZmARF18
GRMZM2G052102	bZIP family	bZIP family	GRMZM2G089640	ARF family	ZmARF22
GRMZM2G114998	C2C2-Dof family	AtDAG1 clade	GRMZM2G077356	AUX/IAA family	ZmIAA14
GRMZM2G171852	C2C2-Dof family	AtDAG1 clade	GRMZM2G138268	AUX/IAA family	ZmIAA7
GRMZM2G140694	C2C2-Dof family	AtHCA2 clade	GRMZM2G079957	AUX/IAA family	ZmIAA1
GRMZM2G028046	C2H2 family	ZmMRPI clade			
GRMZM2G136494	C2H2 family	ZmMRPI clade	**HA3**		
GRMZM2G074032	C2H2 family	AtSGR5 clade	GRMZM2G085678	AP2-EREBP family	SHN1-like Subfamily Va
GRMZM2G129261	C2H2 family	AtJKD clade	GRMZM2G106591	AP2-EREBP family	SHN1-like Subfamily Va
GRMZM2G143723	C2H2 family	AtJKD clade	GRMZM2G139284	MYB family	R2R3-type Subfamily 1
GRMZM2G134998	C2H2 family	AtDOT5 clade	GRMZM2G064744	MYB family	R2R3-type Subfamily 13
GRMZM2G150011	C2H2 family	AtDOT5 clade	GRMZM2G127490	MYB family	R2R3-type Subfamily 13
GRMZM2G132794	GRAS family	AtSHR clade	GRMZM2G147698	MYB family	R2R3-type Subfamily 13
GRMZM2G172657	GRAS family	AtSHR clade	GRMZM2G171781	MYB family	R2R3-type Subfamily 13

B

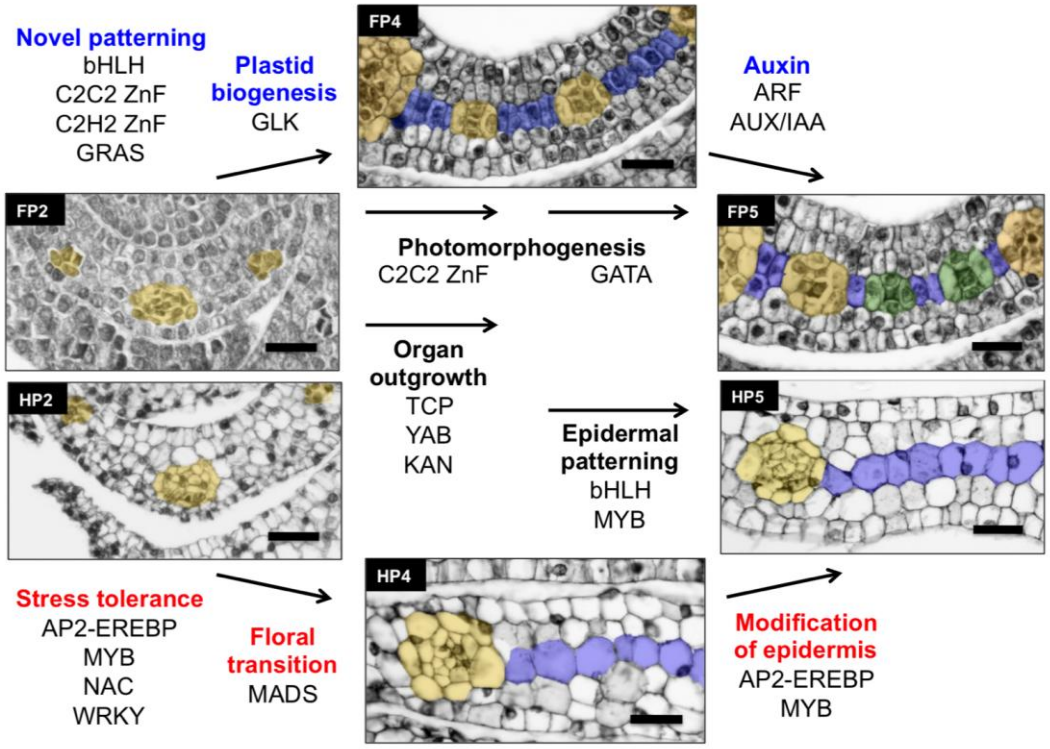


Figure 4.4. Distinct transcription factor cohorts associated with early differentiation events in foliar and husk leaf primordia. Figure legend is on opposite page.

Figure 4.4. Distinct transcription factor cohorts associated with early differentiation events in foliar and husk leaf primordia. A) Transcription factors identified in FA2, FA3 and HA3 profiles that are specific to foliar or husk leaves, and have no previously identified role in early maize leaf development. **B)** Schematic overview of novel insight into early differentiation events. Transverse sections of foliar and husk leaf primordia are colour coded to indicate the midvein and lateral veins (yellow), intermediate veins (green) and interveinal mesophyll cells (blue). Images courtesy of Dr. Peng Wang. Transcription factors listed under black headings are common to both foliar and husk leaves, under red headings are husk-specific and under blue headings are foliar-specific. Scale bar = 20□m, reproduced from Wang et al (2013a).

C-terminal domain by which MRPI proteins interact with MYB-related protein 1 is also found in proteins identified as SHR targets (Levesque et al., 2006; Royo et al., 2009; Cui et al., 2011). Further support for the activity of this cohort of sixteen genes in vascular patterning comes from a transcriptomic analysis of *Arabidopsis* provascular cells that shows upregulation of homologues of four of this cohort (bZIP, two *DAG*-like genes and a *SGR5*-like gene) during vascular specification (Appendix 4.30) (Gandotra et al., 2013), suggesting conserved roles in vascular development. In combination, expression profiles and phylogenetic relationships of these sixteen genes imply that they are likely to function in cellular patterning in maize foliar leaves (Figure 4.4B).

The identification of a cohort of genes in FA2 as candidate regulators of vascular and cellular patterning is evidence for the distinct developmental trajectories that foliar and husk leaves follow even from very early development. However transcripts shared between the FA2/HA2 and FA3 /HA3 clusters shows conservation of core developmental programmes between the two leaf types. For example regulators of organ outgrowth (e.g. *YAB* (Juarez et al., 2004); *AS2-like* (Evans, 2007); *TCP-like* (Martin-Trillo and Cubas, 2010)), photomorphogenesis (*OBP3-like* (Ward et al., 2005); *CO-like* (Putterill et al., 2004)), abaxial-adaxial patterning (*KAN1* (Candela et al., 2008)) and epidermal patterning (*SCRM-like* (Pires and Dolan, 2010); *MIXTA-like* (Dubos et al., 2010)) are all shared. HA2 is distinguished by a unique enrichment of ten genes implicated in floral transition (Appendix 4.24) whereas HA3 features fewer genes with roles in reproductive development

(Appendix 4.25). This observation is consistent with the fact that the inflorescence meristem is clearly differentiated at the stage at which HP5 primordia are harvested, and thus expression should be elevated by HP3/4 (as in HA2). HA3 is defined by a set of eight R2R3 MYBs, of which five are husk specific (Figure 4.4). Four of these HA3 specific genes share a clade with an *Arabidopsis* gene that regulates phenylpropanoid metabolism and mucilage deposition (Penfield et al., 2001; Newman et al., 2004) while the other forms a clade with genes involved in defense responses (Dubos et al., 2010). These observations, in addition to the presence in HA3 of homologues of *SHINE* (*SHN*)*1*, which regulates epidermal wax deposition in *Arabidopsis* and rice (Aharoni et al., 2004; Wang et al., 2012b), suggest that the husk leaf epidermis is undergoing considerable modification between P4 and P5, potentially to provide protection to the developing inflorescence.

4.2.8 Supervised classification of gene expression patterns across the dataset identifies candidate Kranz regulators

To identify further candidate regulators of Kranz development, a supervised classification system complementary to the primordia profile analyses described above was constructed. This utilised knowledge of the anatomical and developmental characteristics of all ten samples from this dataset (Figure 3.1) and an additional study on the developmental gradient across a single maize leaf (Li et al., 2010). Two broad mechanisms for the regulation of Kranz development can be envisaged. It is possible that Kranz anatomy is induced in foliar leaves by positive regulators of Kranz initiation, which would suggest

that non-Kranz tissue is the default developmental state in a leaf. Conversely, it is possible that Kranz tissue is the default state, and that a lack of Kranz initiation in husk leaves is a consequence of the activity of negative Kranz regulators. To ensure that candidate genes generated by both hypotheses were captured, supervised classification systems were designed to identify both putative positive and negative regulators.

The filtration steps used to identify candidate Kranz regulators are summarised in Figure 4.5. For positive regulators it was assumed that gene expression would be higher in foliar compared to husk leaves, and that expression would peak in the foliar primordia samples, concomitant with or just prior to the induction of Kranz development (Figure 3.1). It is possible that some signal for a positive regulator of Kranz development may exist in the husk primordia samples, as vascular differentiation also occurs in these samples, however, it is expected that if this is the case it would be significantly lower than in the foliar primordia samples. Without precise *a priori* knowledge of how Kranz regulators are expressed, a conservative approach to gene filtration was employed where possible. Hence some flexibility in relative expression levels among foliar primordia samples and among husk primordia samples was allowed. Due to heterogeneity in tissue maturation along the proximodistal axis of the expanding FI leaf, the extent of active Kranz elaboration was unknown and therefore this sample was only peripherally informative in the positive regulator filtration steps. However, it is well understood that development across an expanding leaf such as the FI sample occurs basipetally (youngest tissue at the base).

A

Positive Regulators	Negative Regulators
All genes **N = 35,770**	All genes **N = 35,770**
Step 1: FP or FP3/4 or FP5 significantly > FE + HP + HI +HE Genes must pass test in 2 out of 3 FP samples. **N = 918**	**Step 1:** HP or HP3/4 or HP5 significantly > FE + HI + HE Genes must pass test in 2 out of 3 HP samples. **N = 2,567**
Step 2: FP3/4 or FP5 significantly > or not different to FP FP3/4 not significantly < FP HP3/4 not significantly > FP3/4 HP5 not significantly > FP5 HP not significantly > FP HP5 significantly < FP3/4 HP significantly < FP3/4 FP5 or FP3/4 > 0 HI not significantly > FP, FP3/4, FP5, FI or FE HE not significantly > FP, FP3/4, FP5, FI or FE FE not significantly > FP, FP3/4, FP5 or FI **N = 334**	**Step 2:** FP not significantly > HP FP3/4 not significantly > HP3/4 FP5 not significantly > HP5 FE or HI or HE not significantly > FP or FP3/4 or FP5 HP3/4 significantly > FP HP5 significantly > FP **N = 160**
Step 3: If in leaf gradient dataset: Base not significantly < 1cm, 4cm or tip 1cm not significantly < 4cm, tip 4cm not significantly < tip **N = 283**	**Step 3:** If in leaf gradient dataset: Tip not significantly > 4cm, 1cm or base 4cm not significantly > 1cm or base 1cm not significantly > base **N = 142**

B

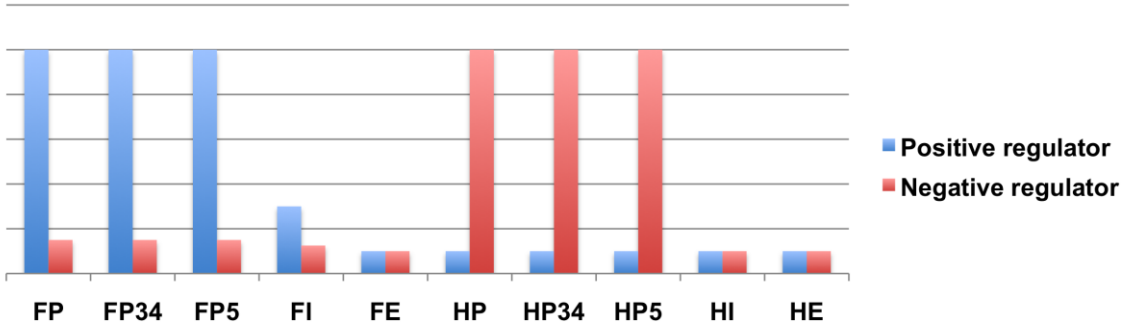


Figure 4.5. Filtration steps to identify putative regulators of Kranz anatomy. A) Stepwise filtration process, all tests were carried out at a *p*-value cut-off of 0.05 (by Dr. Steve Kelly). **B)** Schematic representation of expected expression patterns for positive (blue bars) and negative (red bars) regulators.

As such a gene was discarded if it was found to be expressed to a significantly higher level in older compared to younger (distal compared to proximal) tissue in a study along a proximodistal gradient of a similar leaf stage to FI (Li et al., 2010). The only rudimentary differentiation of Kranz tissue that is apparent at the base of foliar leaves supports the use of this filter criterion (Majeran et al., 2010).

For negative regulators it was assumed that expression would be higher in husk leaves than in foliar leaves and that expression would be highest in the husk primordia samples. As a putative negative regulator would act to inhibit the initiation of procambium, it is possible that there may be some signal in the foliar primordia samples (i.e. to allow for M cell formation between veins). The FI sample was excluded from the filtration criteria for negative regulator selection due to the reasons described above. Where genes were present in the maize leaf developmental gradient data (Li et al., 2010), expression was expected not to increase distally. No (or significantly less) expression of either negative or positive regulators is expected in the fully differentiated HI, HE and FE samples. The filters designed to identify positive and negative putative Kranz regulators both assume that Kranz development is regulated at the transcriptional level.

Based on these assumptions, systematic gene filtration identified 283 candidate positive regulators and 142 candidate negative regulators of Kranz development in maize (Figure 4.5A). Homology and protein domain based manual annotation of the candidate lists identified TFs as the biggest single

functional grouping for both the positive and negative candidates (Figure 4.6), although genes with a wide range of molecular roles were identified in both lists. Further automated GO term, MaizeCyc, MapMan and Pfam domain annotations are supplied with the candidate gene lists in Appendices 4.31 (positive regulators) and 4.32 (negative regulators) (automated annotations carried out by Dr. Steve Kelly). Analysis of the GO term annotations for both candidate gene lists confirms the enrichment of transcriptional regulators among the putative candidates. Figure 4.7 shows the enriched GO terms identified in the list of positive candidates. The enrichment of microtubule associated and ribosomal genes is consistent with the design of the filters to select genes expressed in rapidly dividing primordia samples, however, these are unlikely to function in Kranz development. The enrichment of genes annotated with the term 'ice binding' is also unlikely to reflect an association with the regulation of Kranz development. The remaining enriched GO terms in the positive list, and all the GO terms enriched in the negative candidate list (data not shown) are associated with transcriptional regulation and are therefore of more interest.

The integration of manual and automated annotation methods revealed 71 TF encoding genes within the positive and negative candidate lists (48 positive, 23 negative (Table 4.4)). Support for these genes as candidates comes from a number of independent sources. First, orthologues of negative regulators of Kranz development must be expressed in developing rice leaves. This was found for the majority of the putative negative regulators (Wang et al., 2013a). Second, both *SHR1* and the *DOT5* orthologue identified in the FA2 profile are

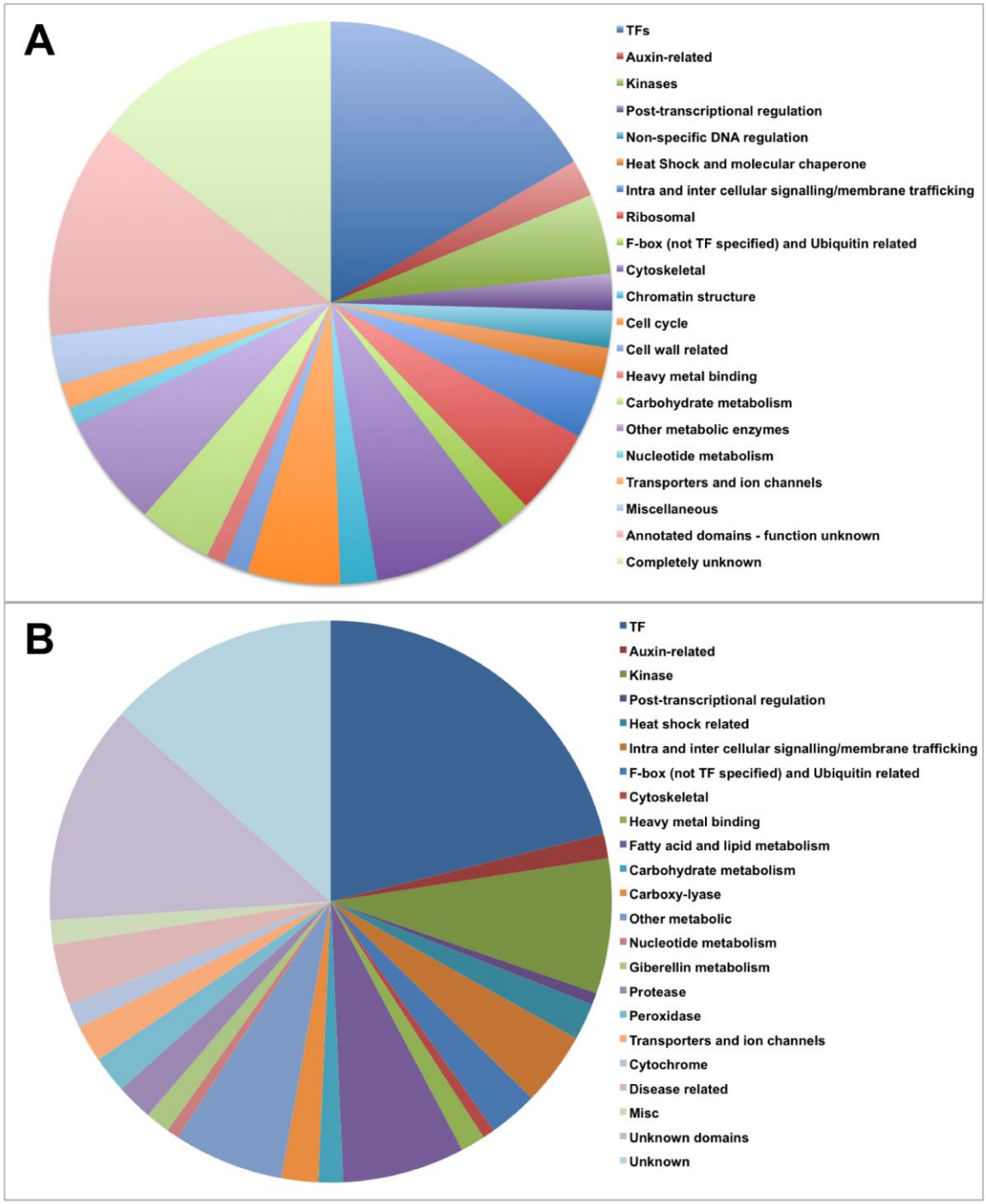


Figure 4.6. **Annotation of putative Kranz candidate regulators. A)** Putative positive regulators, **B)** putative negative candidates. Functions annotated based on protein domain and similarity to known proteins.

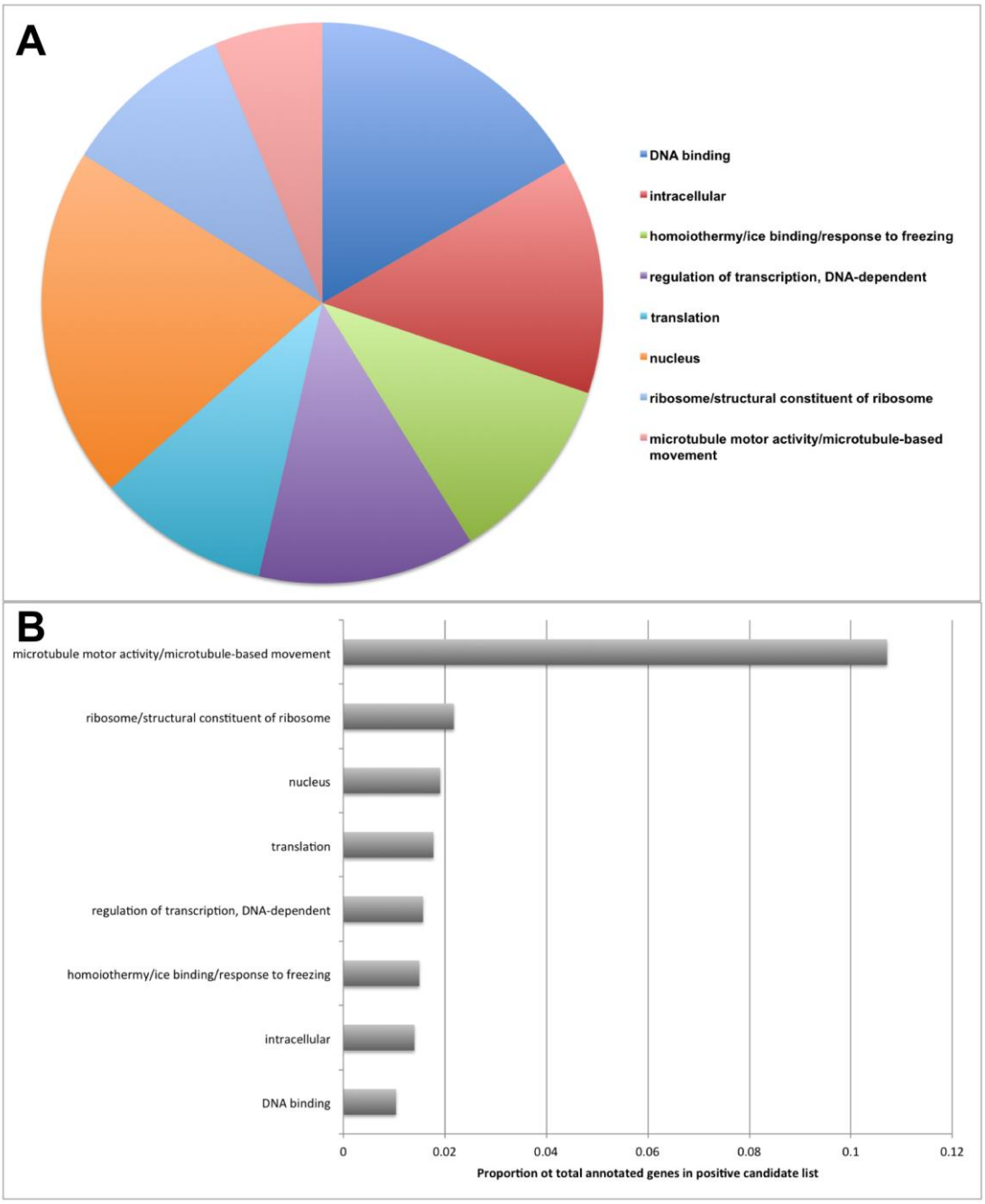


Figure 4.7. **Significantly overrepresented GO terms in putative positive Kranz regulator list. A)** Segment size represents number of genes annotated with associated GO term **B)** Proportion of total number of genes annotated with particular GO term in maize genome in putative positive candidate list.

In the positive candidate gene list. *SHR1* expression has previously been shown to predict sights of leaf vasculature formation (Gardiner et al., 201) and mutant alleles of key partner gene *scr1* show some disruption to Kranz patterning (Slewinski et al., 2012), while *DOT5* regulates vascular patterning in *Arabidopsis* (Petricka et al., 2008). Third, orthologues of a further five transcription factors (plus thirteen non-TF positive candidates) were found to be enriched in *Arabidopsis* provascular cells (bHLH, *GATA* ZnF, *SCR*, 2 *SBP-like*) (Gandotra et al., 2013). Finally, approximately 81% of the candidate positive regulators annotated as TFs were identified in coexpression modules associated with increased Kranz development in a pooled embryonic leaf dataset, compared with approximately 27% of TFs in the total embryonic leaf dataset (Liu et al., 2013). These observations support the filtration criteria used to select candidate genes and suggest that key regulators of Kranz development are likely to be found within the lists shown in Table 4.4.

4.3 Discussion

4.3.1 Signature gene sets and primordia expression profiles reveal unique and shared components of foliar and husk leaf development

The gene expression data described here provide insights into the early patterning of maize leaves that has not previously been possible. In particular, the classification of gene expression by signature gene set (Table 4.1, Appendices 4.1-4.12) and primordia profile (Figure 4.3, 4.4, Table 4.3,

Putative Positive Regulators			
GRMZM2G146688	AP2-EREBP family	GRMZM2G119359	GRF
GRMZM2G151542	AP2-EREBP family	GRMZM5G850129	GRF
GRMZM2G021573	AP2-EREBP family	GRMZM5G893117	GRF
GRMZM2G399072	AP2-EREBP family	GRMZM2G178102	HD-ZIP III family
GRMZM2G121309	Aux/IAA family	GRMZM2G098813	LFY family (ZFL1)
GRMZM2G163975	bHLH family	GRMZM2G471089	MADS family
GRMZM2G098988	bHLH family	GRMZM2G171365	MADS family (ZmMADS1)
GRMZM2G045883	bHLH family	GRMZM2G469304	Putative ternary complex factor MIP1
GRMZM2G015666	bHLH family	GRMZM2G039074	Myb family KAN-like
GRMZM2G082586	bHLH family	GRMZM2G374986	Myb family
AC215201.3_FG008	bHLH family	GRMZM5G887276	Myb family
GRMZM2G095899	bHLH family	GRMZM2G111045	Myb family MIXTA-like
GRMZM2G178182	bHLH family	GRMZM2G040924	Myb family MIXTA-like
GRMZM2G123900	C2C2-Dof OBP3-like	GRMZM2G312419	Myb family LOF-like
GRMZM2G318592	C2H2 family	GRMZM2G131577	Basic TF (NAC domain)
GRMZM2G028046	C2H2 family MRPI-like	GRMZM2G126018	SBP family
GRMZM2G136494	C2H2 family MRPI-like	GRMZM2G061734	SBP family
GRMZM2G150011	C2H2 family DOT5-like	GRMZM2G148467	SBP family
GRMZM2G002280	C3HC4 RING ZnF	GRMZM2G097275	SBP family
GRMZM2G462623	DP-1 family	GRMZM2G472945	TLP-family
GRMZM2G140669	GATA ZnF family	GRMZM2G377217	WRKY family
GRMZM2G132794	GRAS family (SHR)	GRMZM2G425236	ZnF-HD family
GRMZM2G172657	GRAS family (SHR)	GRMZM2G417229	ZnF-HD family
GRMZM2G131516	GRAS family (SCR1)	GRMZM2G069365	ZnF-HD family
Putative Negative Regulators			
GRMZM2G086573	AP2-EREBP family	GRMZM2G140694	C2C2-Dof family
GRMZM2G156006	AP2-EREBP family	GRMZM2G171600	CAMTA family
GRMZM2G028980	ARF family	GRMZM2G132367	HB family
GRMZM2G085751	bHLH family	GRMZM2G062244	HB family
GRMZM2G064638	bHLH family	GRMZM2G060544	LOB family
GRMZM2G045109	bHLH family	GRMZM2G005155	MADS family
GRMZM2G180406	bHLH family	GRMZM2G137510	MADS family
GRMZM2G137541	bHLH family	GRMZM2G181030	MYB-related family
GRMZM2G077124	bZIP family	GRMZM2G003715	NAC family
GRMZM2G052102	bZIP family	GRMZM2G065451	SBP family
GRMZM2G000842	bZIP family	GRMZM2G170034	LIM domain ZnF
GRMZM2G176063	C2C2-Dof family		

Table 4.4. Putative positive and negative Kranz candidate transcription factors.

Appendices 4.13-4.26) reveals novel cohorts of genes that are both shared by, and distinguish, foliar and husk leaf development. The fact that much of what is already known about the genetic regulation of early leaf development, particularly in the grasses, is recapitulated in this study (Table 4.2, Appendix 4.29) is a strong validation of the expression data and the analytical approaches used.

The novel insights that these expression data provide into early maize leaf development can be classified into four groups as follows. First, analysis of the primordia profiles provides further information about homologues of genes already known to function in leaf patterning. For example there are 13 *BEL-like* (*BELL*) genes in *Arabidopsis* (Mukherjee et al., 2009), of which some are necessary for meristem specification (Rutjens et al., 2009) while others function in lateral organ growth (Kumar et al., 2007). There are 17 *BELL* genes in maize, none of which have been functionally characterised to date (Mukherjee et al., 2009). Six maize *BELL* genes were identified in the primordia profile analyses, four in the D profiles (suggesting a role in meristem specification) and two in the A profiles (suggesting a role in leaf growth). Thus analysis of the primordia profiles of TF families known to be involved in leaf development will provide information both about which genes within the family are being expressed, and what function they may be carrying out (for multi-functional families). This strategy for inferring gene function will be particularly useful for TF families with very closely related genes in which it is difficult to resolve phylogenetic relationships.

Second, the results presented here implicate novel TF families in the regulation of leaf development. For example *PLATZ* genes are known to act as transcriptional repressors (Nagano et al., 2001), but as yet no members of this TF family have been functionally characterised. Identification of *PLATZ* family members in the primordia profiles suggests they may function in leaf development. Third, the clustering of genes into different expression profiles provides further information on possible co-expression modules of maize gene expression and putative genetic regulatory networks. This is demonstrated by the appearance of homologues of the rice genes *DROOPING LEAF* (*DL*) (Yamaguchi et al., 2004) and *OsMADS32/CHIMERIC FLORAL ORGANS* (Sang et al., 2012) in the same profiles (both are present in FD1 and HD1). In rice *OsMADS32* is known to repress *DL* (Sang et al., 2012). The appearance of other cohorts of genes in the same, or developmentally consecutive, profiles may imply shared regulatory networks. Data such as these will help to enhance recent genome wide models of transcriptional networks in maize (e.g. Ficklin and Feltus, 2011; Downs et al., 2013).

Fourth, this study provides unique insights into the distinct trajectories that foliar and husk leaves follow during early development. Previous comparative work on foliar and husk leaves has focused on photosynthetic function in differentiated leaves (Langdale et al., 1988b; Pengelly et al., 2011) or on gene expression in SAM versus axillary meristems (without any leaf primordia samples post P1) (Takacs et al., 2012). The patterns of primordia gene expression identified in this work (Appendices 4.13-4.26) imply that conserved

biological processes exist between the two leaf types at the same stage of development. Of particular note, *KAN*, *YAB* and *ARF* (D2, D3), *ANT, TCP* and *SCHIZORIZA* (*SCZ*) (D3, A2) and bHLH and *HDZIV* (A2, A3) family members are shared between equivalent foliar and husk profiles (Table 4.2, 4.3; Figure 4.4B). This suggests conserved mechanisms of organ initiation/axial patterning, outgrowth/expansion and epidermal patterning respectively. Conversely, sets of TFs unique to foliar and husk profiles (in addition to differentially enriched GO terms and MaizeCyc pathways) suggest distinct developmental trajectories, consistent with the biological functions of each leaf type. For example, genes that function in chloroplast biogenesis and vascular patterning are more common to foliar profiles while genes that regulate floral transition and epidermal wax deposition are upregulated in husk profiles. These differences were expected, but analysis of genes with unknown functions that exhibit similar expression patterns may reveal novel foliar and husk specifying factors. The fact that more than 80% of TFs are unique in the D3 profiles, in which expression is highest in the P and P3/4 samples, suggests foliar and husk leaves follow different trajectories from the moment of leaf initiation. Furthermore, observation that a minority of genes are expressed in non-equivalent profiles between foliar and husk leaves (e.g. FA1 and HA2) (Figure 4.3) implies that core developmental mechanisms operate at different rates in the two leaf types. Changes in the temporal expression of independently regulated processes have contributed to developmental novelty during evolution (Gould, 1977). Such shifts may explain the ontogeny of the husk leaf, which is effectively an expansion of leaf sheath tissue (which itself differentiates after blade tissue in foliar leaves). The

altered expression dynamics of the ligule specifying gene *LG2* (Walsh et al., 1998) between foliar and husk leaf primordia is thus intriguing (Table 4.2). In summary, the primordia profile analyses presented here provide a unique overview of early leaf development in foliar and husk leaves.

4.3.2 The combination of classification systems identifies candidate regulators of early Kranz patterning

The primordia expression profiles discussed above were designed to identify genes that regulate different aspects of early leaf development in both foliar and husk leaves, and to capture the biological processes that distinguish the two leaf types. As part of this analysis, sixteen genes were identified as potential regulators of Kranz type development in foliar leaves (Figure 4.4A). To complement this approach, a set of filters were constructed that utilise expression data from all ten samples within this dataset and an additional study on the developmental gradient within a single maize leaf (Li et al., 2010). The integration of data from a broader developmental framework allowed the identification of 283 putative positive regulators and 142 putative negative regulators of Kranz anatomy.

Three pieces of information suggest that it is more likely that Kranz develops through positive regulation. First, Kranz anatomy is known to have evolved independently on over 60 occasions in a wide variety of angiosperms (Sage et al., 2011a). Despite the remarkable extent of convergent Kranz evolution, it is more parsimonious for Kranz to have evolved ~60 times from a C_3 state than

for Kranz to be the default, and negative regulation to have evolved in the ~97% of angiosperms that are not C_4. Although it is possible that Kranz is the default angiosperm state and that negative regulation evolved at the base of the angiosperms, only to be lost on over 60 occasions, the additional evolutionary step this hypothesis requires also makes it a less parsimonious solution. Second, given that C_4 plants develop leaf-like organs without Kranz, it is most likely that Kranz evolved by superimposing a positive regulatory process on a default non-Kranz background. Third, components of the C_4 pathway are known to be derived and to require a variety of signals for appropriate induction (Sheen and Bogorad, 1985; Langdale et al., 1988a; Langdale et al., 1988b). Taken together, these observations support a positive regulatory mechanism of Kranz induction. 48 of the putative positive candidates identified are TFs, of which 6 were also identified in the FA2 candidate list. In combination, this suggests an upper limit of 58 TFs that could act as positive regulators of Kranz development in maize.

Candidate Kranz regulators were selected based on a combination of primordia profiles and a wider range of developmental stages to ensure that all stages of Kranz differentiation were identified. In this regard, Kranz development can be thought of in terms of three distinct (but overlapping) stages. First, procambium initiates at regular intervals along the mediolateral axis of leaf primordia. Second, concentric rings of BS and M cells differentiate around vascular strands. Third, components of the C_4 pathway accumulate in a cell specific manner. As C_4 pathway components are not enriched in the primordia samples (Appendices 4.1-4.3, 4.6-4.8) despite obvious anatomical

differentiation (Figure 3.1), the data presented here suggest that vascular differentiation and BS and M cell specification occurs prior to C_4 pathway initiation and not concomitantly as previously suggested (Langdale et al., 1987).

The reported expression classification systems were designed so that regulators of vein spacing should appear in both the FA2 profile (Appendix 4.17) and the list of putative Kranz regulators (Appendix 4.31). The six TFs found in both lists (Figure 4.4A, Table 4.4) are therefore strong candidates for vein patterning regulators. Of these, three are homologues of *Arabidopsis* genes with known roles in vascular development (*SHR* (Levesque et al., 2006; Gardiner et al., 2011) (two maize homologues) and *DOT5* (Petricka et al., 2008)), two encode MRPI proteins that share a C-terminal domain with SHR target proteins (Levesque et al., 2006; Royo et al., 2009; Cui et al., 2011) and one encodes a bHLH protein that is closely related to the epidermal cell patterning protein SPEECHLESS (SPCH) (Lampard et al., 2008). It is feasible that the maize *SPCH* homologue has been co-opted into regulating procambial, rather than epidermal, cell patterning. Alternatively, it may be involved in the regulation of epidermal patterning associated with the underlying Kranz tissue. Expression data for orthologues of additional genes implicated in vascular development in *Arabidopsis* are compiled in Appendix 4.33.

In addition to the identification of vein patterning candidates, the interrogation of published transcriptome data from isolated maize BS and M cells (Li et al.,

2010; Chang et al., 2012) revealed a number of genes in the FA2 profile and putative positive candidate list that are upregulated in a cell specific manner (Appendix 4.34). Indeed, of all the profiles, FA2 contained the highest number of BS or M specific genes, suggesting that BS and M cell differentiation occurs in quick succession to vascular patterning. Notably, two proteins in the FA2 profile are closely related to JKD, which is known to act in the *SHR* pathway in cell layer specification in the root (Welch et al., 2007; Ogasawara et al., 2011). Both proteins are upregulated in either BS or M cells (one in BS and one in the M), supporting the role of a *SHR* regulated pathway in patterning Kranz anatomy. The identification of a large cohort of genes that are expressed in a BS or M cell-specific manner within the expression classification systems suggests that further BS and M cell-specification factors are present in the candidate gene lists, which may or may not act within *SHR* regulated pathways.

The emphasis placed on members of the *SHR* pathway in potentially regulating Kranz patterning is a consequence of the observed expression patterns and mutant phenotypes of *SHR* and its targets. In *Arabidopsis*, *SHR* is expressed in the vasculature of the root and shoot (Helariutta et al., 2000; Gardiner et al., 2011) and loss of *shr* function leads to defects in endodermal patterning around root, hypocotyl and stem vasculature (Fukaki et al., 1998; Helariutta et al., 2000). In addition to the appearance of orthologues of *SHR* and *SCR* in the candidate gene lists presented here (Table 4.4, Figure 4.4A), support for the role of a *SHR*/*SCR* regulatory module in patterning Kranz development comes from the observation that *SCR1* is expressed in the

vasculature in maize leaves (Lim et al., 2005) and that mutant *scr1* alleles perturb BS development in maize (Slewinski et al., 2012). In addition to the SHR targets already mentioned, a further four targets of SHR were identified in the putative positive candidate list (two F-box proteins, a receptor like kinase and a truncated C2H2 ZnF protein (Appendices 4.31, 4.35)) (Levesque et al., 2006; Cui et al., 2011). In summary, components of the *SHR* regulatory network, which is known to regulate vascular patterning in a number of different tissue contexts and is active in maize vasculature, are overrepresented in gene lists produced by filters designed to identify candidate Kranz regulators.

4.3.3 Many candidate Kranz regulators are not transcription factors

With regard to the identification of candidate Kranz regulators, an emphasis has been placed on the TFs that have been revealed by the classification systems described above. However, TFs comprise only a minority of these candidate gene lists. Of the 463 genes in the FA2 profile only 59 are TFs (Figure 4.3A), and only 48 of the 283 putative positive regulators are annotated as TFs (Appendix 4.31). Considering the fundamental roles that TF play in the regulation of development (Alberts et al., 2008), such an emphasis is justifiable. However, it should not preclude consideration of the potential roles of other classes of proteins. For example, within the list of 283 positive candidates 13 kinases are present, including 10 leucine rich repeat receptor like kinases (LRR-RLK). LRR-RLKs are known to regulate key aspects of plant development, including patterning of shoot, root and cambial meristems

(reviewed by Miyashima et al., 2013). The maize orthologue of the LRR-RLK *TDIF RECEPTOR*/*PHLOEM INTERCALATED WITH XYLEM* (*TDR*/*PXY*), which regulates cambial stem cell divisions in *Arabidopsis* (Fisher and Turner, 2007; Hirakawa et al., 2008), is present in both the FA2 and putative positive candidate list. While the role of *TDR*/*PXY* in cambial stem cell maintenance is not directly related to Kranz patterning, it highlights the fact that genes identified as vascular regulators are present in the candidate gene lists that are not TFs. This is an important consideration when determining which candidates to functionally characterise. The validation of the candidate gene lists through comparison with transcriptome data from other studies of relevant developmental stages (Li et al., 2010; Chang et al., 2012; Gandotra et al., 2013; Liu et al., 2013) suggests that TF or otherwise, the candidates outlined here are likely to include key regulators of Kranz patterning in maize. In the next chapter I will discuss attempts to functionally characterise three putative positive regulators of Kranz development through overexpression in rice and RNA interference in *Setaria viridis*.

Chapter 5: Functional characterisation of Kranz candidates in rice and *Setaria viridis*

5.1 Introduction

The bioinformatic analyses described in the previous chapter have provided support for the role of a cohort of candidate genes in the regulation of Kranz anatomy. Despite the strong support that expression analyses and phylogenetic insights provide, however, without definitive functional characterisation candidate genes remain simply candidates. Two strategies exist for testing the developmental potential of genes: they can either be ectopically expressed (spatially, temporally or both) or endogenous expression can be reduced. To date, although the effects of ectopic expression of a number of C_4 pathway genes have been determined in rice and other C_3 species (reviewed in Hibberd and Covshoff, 2010), no ectopic expression analyses of candidate C_4 anatomical regulators have been published. The finding that a series of rice deletion mutants tolerate plasticity of vascular patterning (Smillie et al., 2012) suggests that overexpression of Kranz candidates may yield results, and that grass species are able to tolerate perturbed vascular patterning. The work in this chapter aimed to test whether ectopic expression of three Kranz candidate genes is able to induce Kranz like structures in rice, and whether disruption of candidate expression affects C_4 development in *Setaria viridis.*

5.2 Results

5.2.1 Three candidate genes were selected for functional characterisation

To test candidate Kranz regulator gene function, three genes were selected for characterisation. Two of these genes, a bHLH-like transcription factor (TF) (AC215201.3_FG008) and an R2R3-type MYB TF (GRMZM2G040924) were identified as part of the 283 positive candidates described in Chapter 4. The third, a leucine rich repeat receptor-like kinase (LRR-RLK) (GRMZM2G040263), was selected based on preliminary candidate filtration analyses prior to the completion of the candidate filtration process described in Chapter 4. Due to a delay in the sequencing of the P3/4 and P5 samples, these preliminary analyses were based on expression data from the six P, I and E samples. The expression patterns for these three genes across all 10 developmental stages used in this study are shown in Figure 5.1. Although the LRR-RLK was not identified as one of the 283 putative positive regulators (0.5 RPKM in the FE sample and 0 RPKM in FP5 and FI led to it failing Step 2 of Figure 4.5A), its expression pattern (high expression in FP and FP34 relative to the other 8 samples) is still consistent with that of a putative positive regulator of Kranz anatomy (Figure 4.5B).

In addition to exhibiting expression patterns indicative of potential roles in Kranz patterning, these three genes were selected as priorities for functional testing on the basis of maize expressed sequence tag (EST), rice microarray

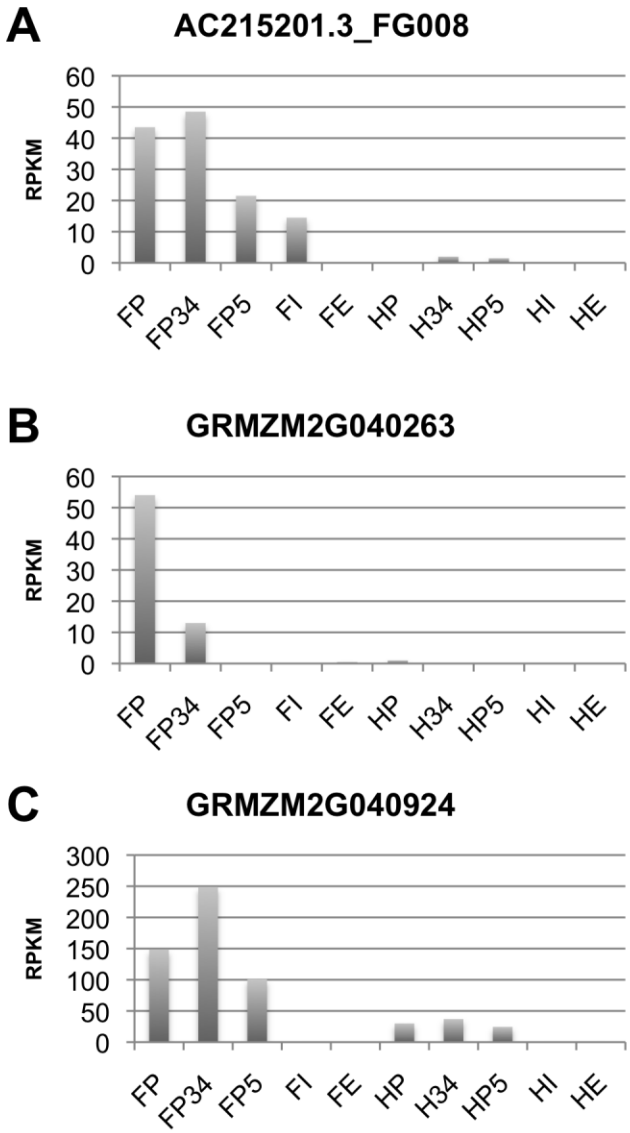


Figure 5.1. Expression patterns of candidate Kranz regulators selected for functional characterisation. Expression values represent average over two technical replicates determined by RNA-seq data.

and phylogenetic data. EST libraries from a wide range of maize developmental stages, including male and female reproductive tissues, were downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov). They were screened on the basis that a Kranz regulator is unlikely to appear in multiple tissue types that don't exhibit Kranz anatomy. None of the three selected genes were found to be expressed in any tissue type aside from shoot apices and young leaves (data not shown). Based on their expression patterns, it is possible that the candidate genes identified by the filtration process describe in Chapter 4 are blade, rather than Kranz, specifying factors. Therefore data from a microarray study on rice leaf blades and sheaths (Sato et al., 2011) was also used to discriminate between candidates. Orthologues of the three genes are not upregulated in rice leaf blades relative to rice leaf sheaths (data not shown). Finally, phylogenetic trees were constructed for each gene to determine whether any closely related genes had been previously characterised. An iterative HMMer search (Kelly et al., 2011) for each gene against a subset of 22 phylogenetically representative plant genomes on the Phytozome resource (v8.0, www.phytozome.net) was used to identify the most closely related genes from across the Viridiplantae. Phylogenetic trees were built using Bayesian methods. Figures 5.2 (bHLH-like), 5.3 (LRR-RLK) and 5.4 (MYB) show that none of the three genes are present in clades with roles that would contradict a potential role in Kranz development. For example, none of the genes identified as closely related to the bHLH-like candidate have been characterised (Figure 5.2). The LRR-RLK is situated in a clade with the *Arabidopsis* gene *RECEPTOR-LIKE KINASE 902* (*RLK902*) (Figure 5.3), which is known to regulate root development in an

auxin responsive manner (ten Hove et al., 2011a; ten Hove et al., 2011b), and is in a sister clade to *GSO1* and *GSO2*, which regulate embryonic epidermal development (Tsuwamoto et al., 2008; Xing et al., 2013). *RLK902*, its sister gene *RECEPTOR KINASE-LIKE1* (*RKL1*) and *LEUCINE RICH REPEAT1* (*LRR1*) (which also sits in *RLK902*/*RKL1* clade) have also been implicated in pathogen response, suggesting a mixed set of roles for this family of LRR-RLK genes. Figure 5.4 shows that the most closely related genes to the MYB candidate also have mixed developmental roles, including regulating floral transition (*LATE MERISTEM IDENTITY2* (*LMI2*) (Pastore et al., 2011)) and epidermal patterning (*MYB16* and *NOECK* (*NOK*) (Folkers et al., 1997; Jakoby et al., 2008; Gilding and Marks, 2010; Oshima et al., 2013)). Taken together, phylogenetic inference lends some support for the functional characterisation of the three selected candidates.

5.2.2 Rice transformation led to strong transgene expression for two Kranz candidates

5.2.2.1 Construct design

To determine whether the three candidates selected for characterisation function in Kranz patterning, a combined strategy of overexpression in C_3 (rice) and knockdown in C_4 (*Setaria viridis*) plants was employed. For overexpression in rice, full length genomic sequences for each of the three candidates were cloned into the Gateway adapted (Invitrogen,

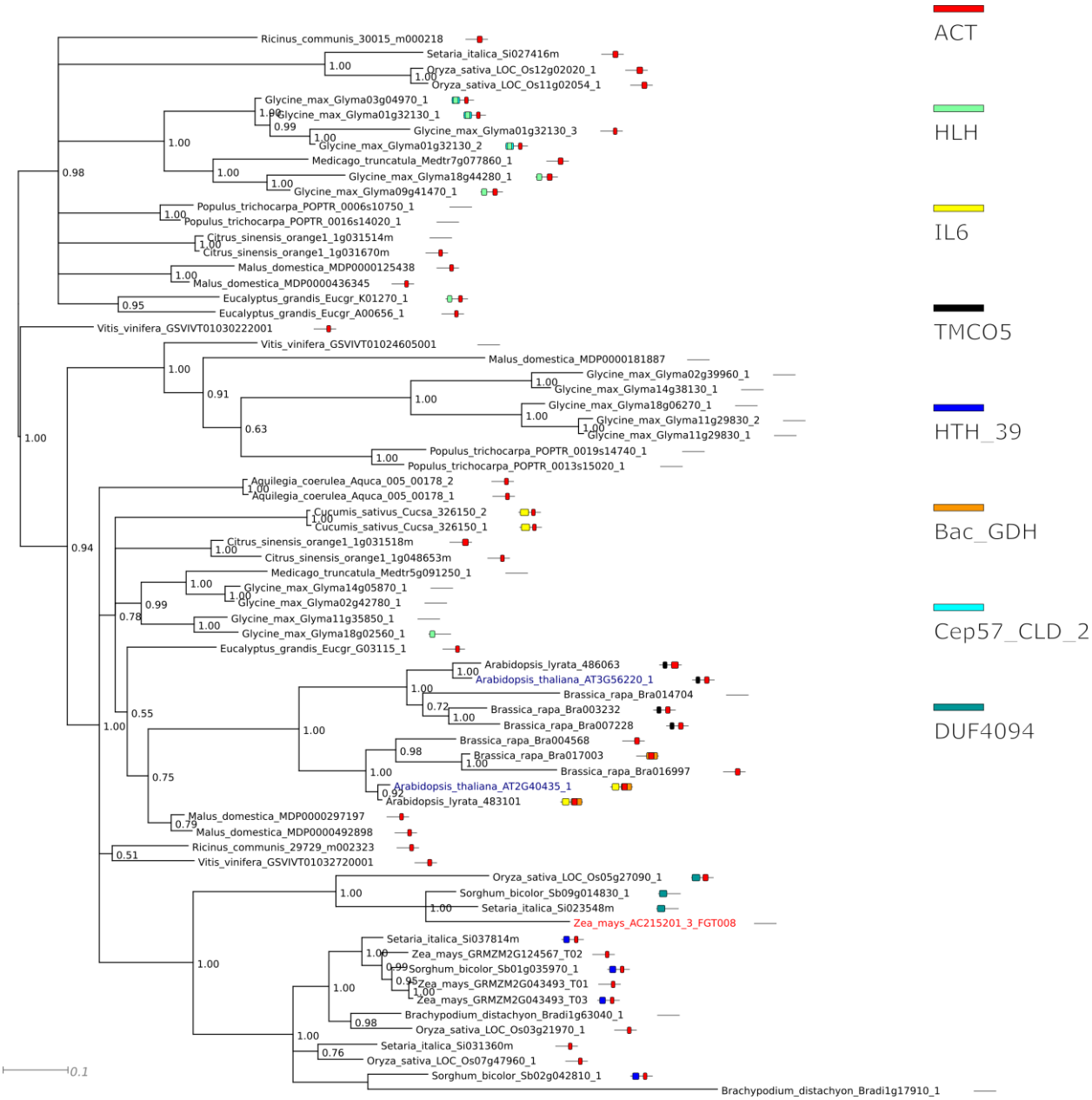


Figure 5.2. **Bayesian phylogenetic tree of bHLH-like candidate gene AC215201_3.FG008.** Posterior probabilities are shown at branch nodes. Maize candidate gene is highlighted in red, *Arabidopsis* genes are highlighted in blue. *Arabidopsis* gene names are provided based on TAIR annotations. Cartoons next to sequences depict the domain organisation of the predicted protein sequence. The sequence is indicated by the black line and domains are coloured according to Pfam domains shown in legend in the top right corner.

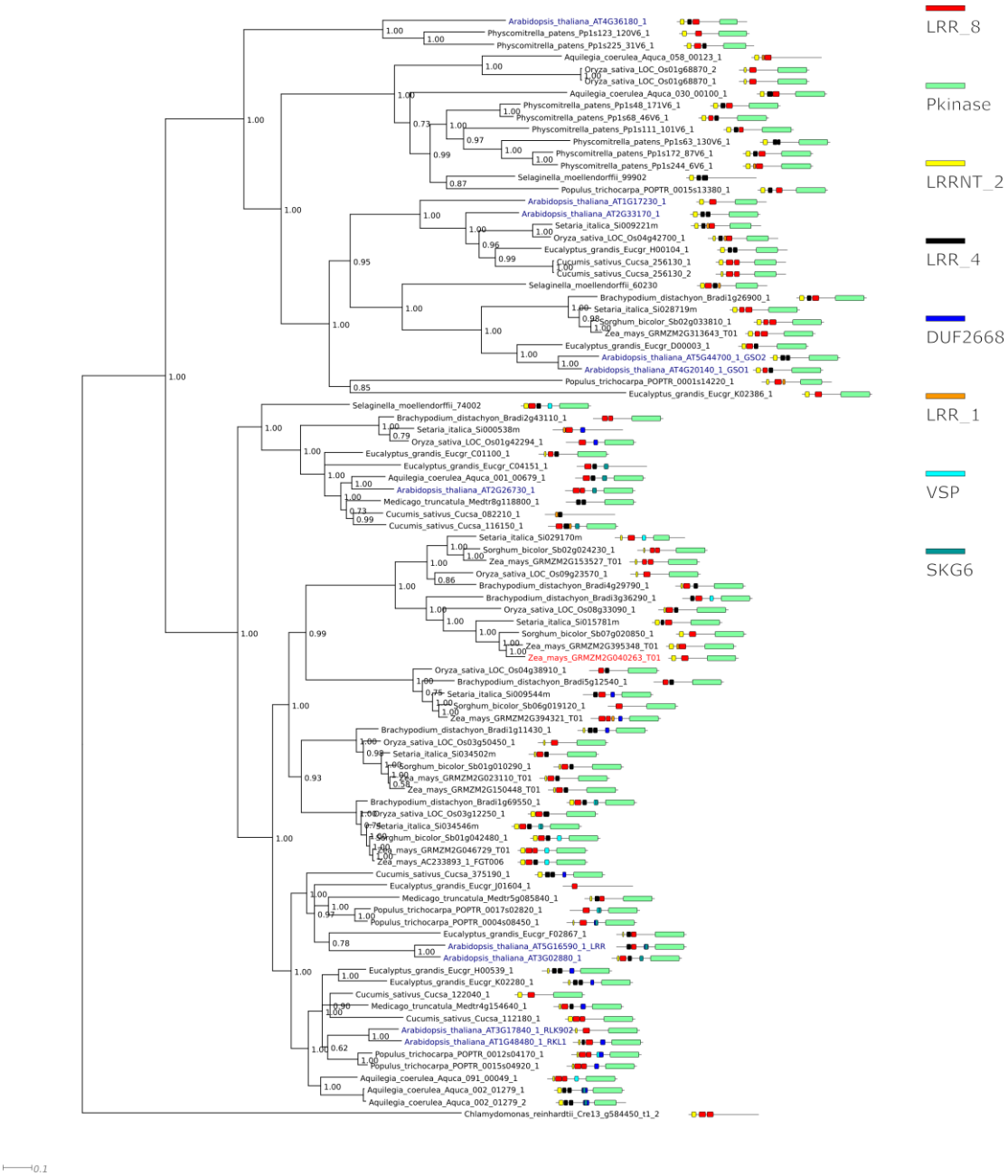


Figure 5.3. **Bayesian phylogenetic tree of LRR-RLK candidate gene GRMZM2G040263.** Posterior probabilities are shown at branch nodes. Maize candidate gene is highlighted in red, *Arabidopsis* genes are highlighted in blue. *Arabidopsis* gene names are provided based on TAIR annotations. Cartoons next to sequences depict the domain organisation of the predicted protein sequence. The sequence is indicated by the black line and domains are coloured according to Pfam domains shown in legend in the top right corner.

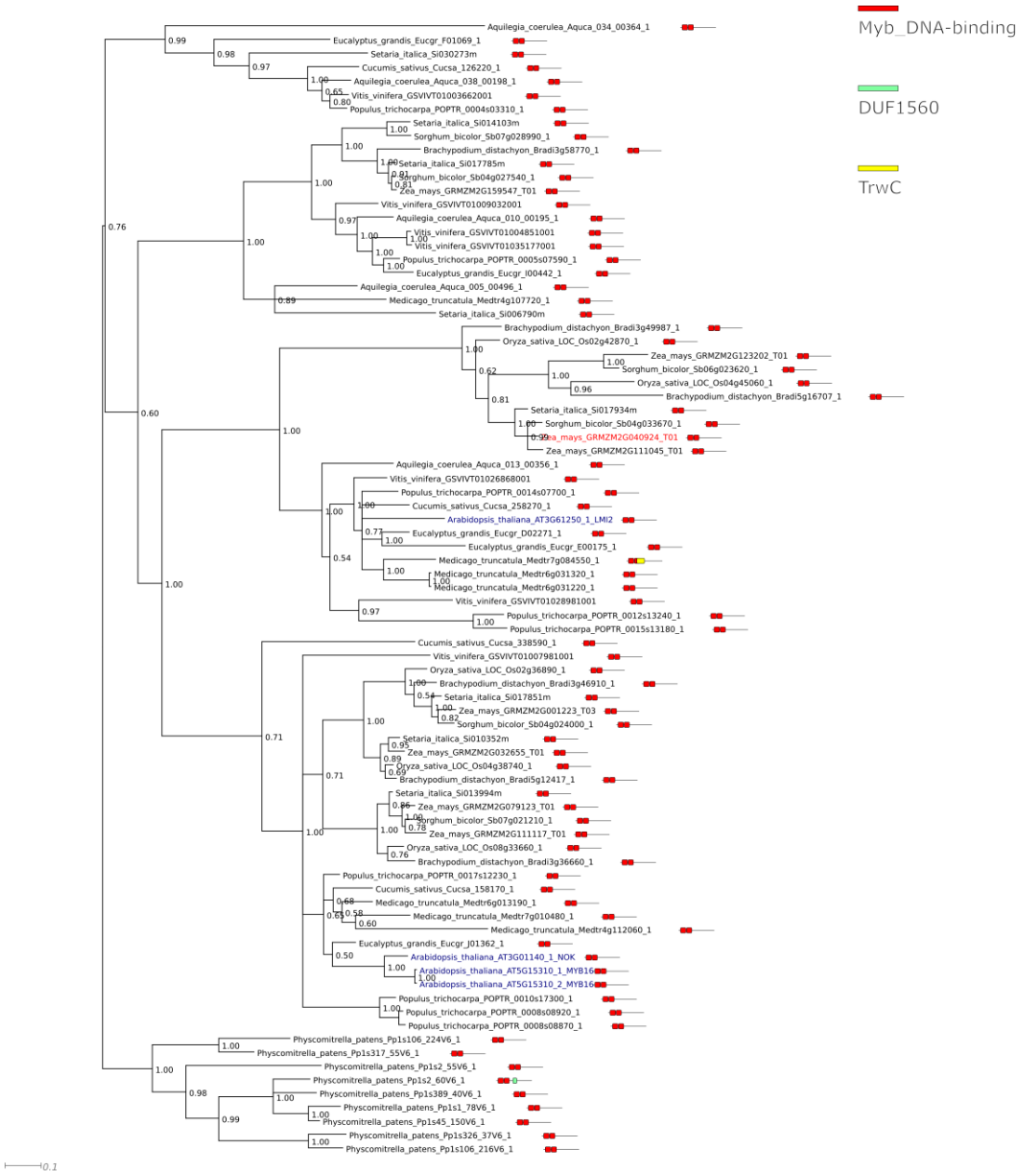


Figure 5.4. **Bayesian phylogenetic tree of MYB candidate gene GRMZM2G040924.** Posterior probabilities are shown at branch nodes. Maize candidate gene is highlighted in red, *Arabidopsis* genes are highlighted in blue. *Arabidopsis* gene names are provided based on TAIR annotations. Cartoons next to sequences depict the domain organisation of the predicted protein sequence. The sequence is indicated by the black line and domains are coloured according to Pfam domains shown in legend in the top right corner.

www.invitrogen.com) pVec8 vector (Wang et al., 1998). pVec8 has previously been shown to induce strong heterologous gene expression in grasses through use of the maize *UBIQUITIN1* promoter plus intron1 sequence (Murray et al., 2004; Alves et al., 2009).

5.2.2.2 Rice transformation

Agrobacterium mediated inoculation of rice (Kitaake variety) embryogenic callus induced from germinating seed was used as a means for plant transformation. This is a well-developed and relatively optimised process (Hiei et al., 1994; Nishimura et al., 2006; Hiei and Komari, 2008) as reflected by the high transformation efficiencies obtained (Table 5.1). As the different rice lines were generated contemporaneously from the same batches of callus, the lack of MYB-OE (R3 in Table 5.1) transformants suggests that overexpression of the MYB candidate in rice is lethal. Positive transformants were identified by genomic PCR using primers designed to amplify the pVec8 *HYGROMYCIN PHOSPHOTRANSFERASE* (*HPT*) resistance gene and a portion of the maize candidate sequence (data not shown).

Preliminary phenotypic analyses were carried out on three PCR-positive T0 lines for each construct, however, to mitigate for the potential effects of callus regeneration the majority of phenotypic characterisation was carried out in the T1 generation. Self-fertilised seed from three individual lines per construct were germinated on filter paper and transferred to soil. The rates of transgene

Construct	Regenerant plants	Transformants	Transformed (%)	Construct	Regenerant plants	Transformants	Transformed (%)
R1	55	23	41.8	**S1**	146	5	3.4
R2	63	60	95.2	**S2**	139	3	2.1
R3	32	0	0	**S3**	126	7	5.6
EV	15	9	60.0	**EV**	14	5	35.7

Table 5.1. T0 transformation rates. Rice lines are shown in blue, *Setaria* lines are shown in green. R1 = bHLH-OE, R2 = LRR-RLK-OE, R3 = MYB-OE, S1 = bHLH-RNAi, S2 = LRR-RLK-RNAi, S3 = MYB-RNAi, EV = Empty Vector.

Line	Germinated seed	Transformants	Transformed (%)	Line	Germinated seed	Transformants	Transformed (%)
R1 B6 8	7	1	14.2	S1 B2 1	15	3	20.0
R1 B6 18	10	9	90.0	S1 B2 6	15	1	6.7
R1 B6 23	10	7	70.0	S1 B2 7	7	3	42.9
R2 B6 2	10	5	50.0	S2 B1 4	15	4	26.7
R2 B6 20	10	5	50.0	S2 B1 7	15	2	13.3
R2 B6 34	10	7	70.0	S2 B3 15	15	3	20.0
				S3 B1 9	15	5	33.3
				S3 B1 11	15	2	13.3
				S3 B2 10	13	2	15.4
				S3 B3 2	15	2	13.3
EV B6 1	10	3	30.0	EV B3 1	9	2	22.2
EV B6 2	9	5	55.6	EV B3 4	15	2	13.3
EV B6 7	5	4	80.0	EV B3 6	13	2	15.4

Table 5.2. T1 transformation rates. Rice lines are shown in blue, *Setaria* lines are shown in green. R1 = bHLH-OE, R2 = LRR-RLK-OE, R3 = MYB-OE, S1 = bHLH-RNAi, S2 = LRR-RLK-RNAi, S3 = MYB-RNAi, EV = Empty Vector. B denotes transformation batch, final number denotes line identifier.

transmission to the T1 generation are shown in Table 5.2. In Table 5.2 and subsequent text the B after the construct name refers to the transformation batch, and the final number the individual line identifier from that batch (the additional inclusion of '1.' identifies this plant at part of the T1 generation). The percentage of transformed offspring from each transformed T0 plant implies that individual T0 inflorescences consist of chimeric tissue. Assuming a single transgene insertion event that would lead to a hemizygous progenitor plant, it would be expected that ~75% of T1 plants carry the transgene. In most of the rice lines, the percentage of transformed offspring is considerably less than this. This could be explained by weak penetrance of reproductive or embryonic developmental defects caused by the heterologous constructs, however, transformed T1 plants did not appear stunted compared to non-transgenic siblings and seed set did not appear to be affected in the T1 generation. If T0 plants are chimeric in form, this strongly supports delaying phenotypic screening of transformed plants until the T1 generation.

5.2.2.3 Expression analyses of T1 rice lines

Transformation of T1 lines was also confirmed by genomic PCR using primers designed to amplify the *HPT* gene and a region of the maize candidate sequence (Figure 5.5). Despite previous successes with Southern blotting, a series of technical difficulties meant that Southern blot analyses failed to validate the results of the PCR screens, or to provide information about transgene copy number in all cases. However, the results did suggest the presence of two insertion events in the R2 B6 34 line (data not shown).

Following confirmation of transformation, effects on gene expression were quantified in T1 plants by qRT-PCR. A single fully expanded leaf was sampled for each plant. Figure 5.6 shows that there is strong transgene expression in all bHLH-OE and LRR-RLK-OE lines, higher in all cases than an endogenous *UBQ* control. No transgene expression was detected in wild type rice. A semi-quantitative PCR analysis also produced similar results (Appendix 5.1). In summary, pVec8 led to strong transgene expression in all transformed rice lines.

5.2.3 Overexpression of Kranz candidate bHLH-like and LRR-RLK genes do not affect leaf development in rice

To determine the effects of Kranz candidate overexpression, a number of different developmental traits were measured in rice. In terms of gross plant morphology, overexpression of the bHLH-like or LRR-RLK candidate genes did not lead to noticeable changes in growth compared to WT and EV lines (Figure 5.7). As the aim of this study was to investigate the regulation of Kranz development, the phenotypic characterisation of transgenic lines was focused on vascular patterning. This was carried out through analysis of hand cut fully expanded leaf sections using UV light microscopy (Figure 5.8). Leaf sections from neither bHLH-OE nor LRR-RLK-OE lines appeared noticeably different from WT or EV plants in terms of vascular number, vascular spacing, BS size, BS chloroplast content or M cell number. To determine whether there

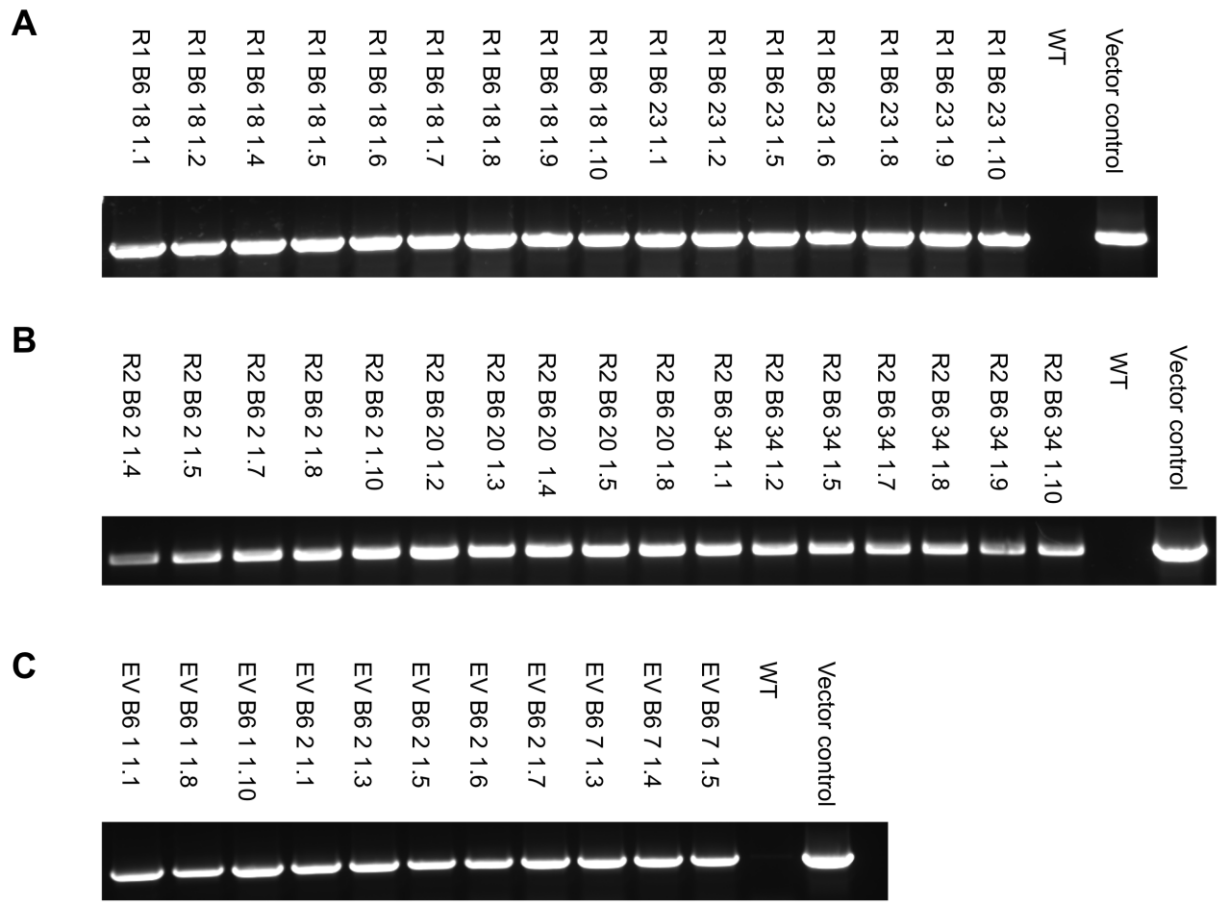


Figure 5.5. **Genomic PCR of T1 rice lines.** PCR from genomic DNA using primers designed to amplify the pVec8 *HPT* gene. Forward primer sequence 5'- CTTCTACACAGCCATCGGTC-3', reverse primer sequence 5'-CCGATGGTTTCTACAAAGATCG-3'. The same results were seen when using primers designed to identify a portion of the maize candidate gene sequence. In each case vector control represents PCR from original binary vector, WT from untransformed DNA. **A)** bHLH-OE lines B) LRR-RLK-OE lines **C)** Empty Vector lines.

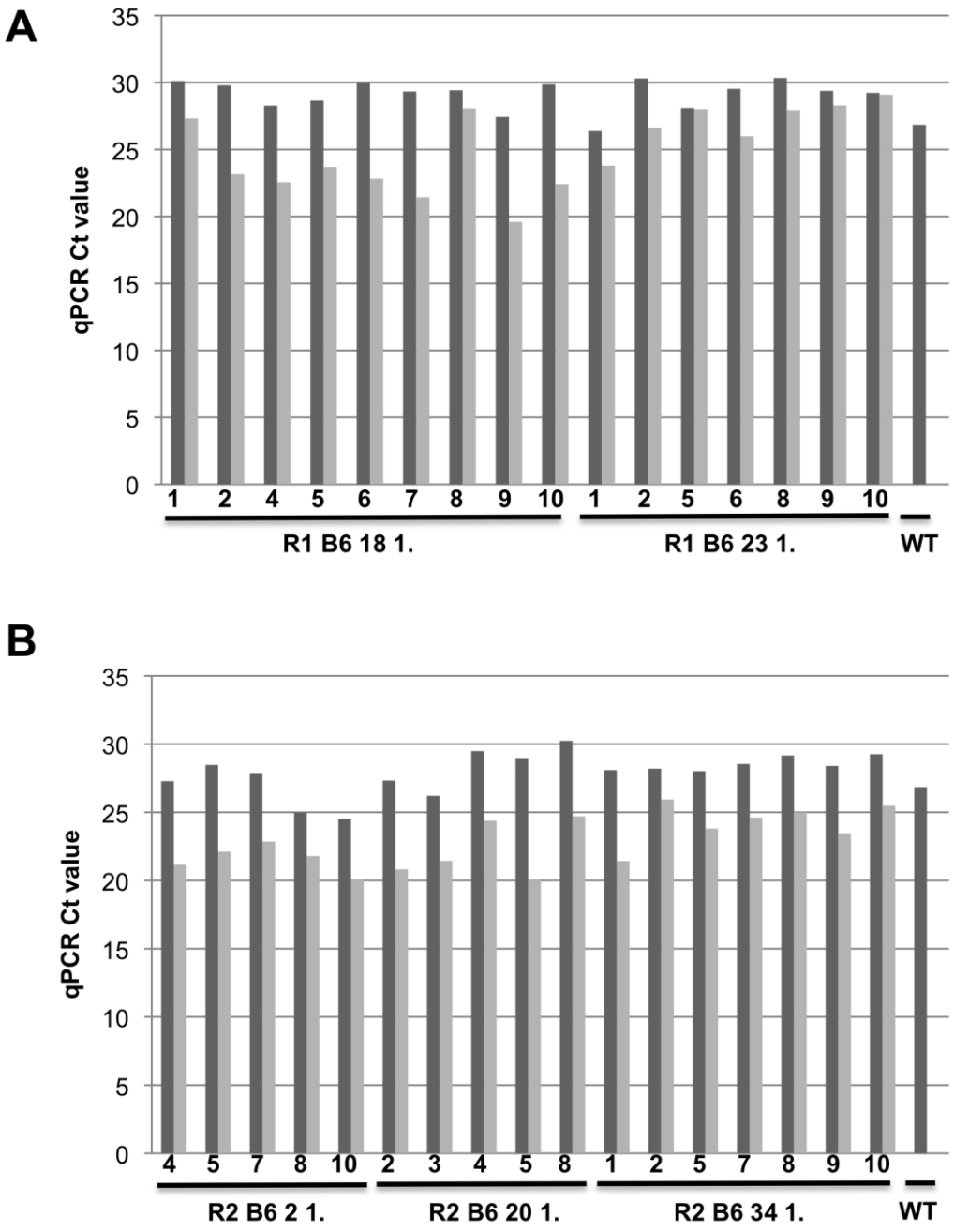


Figure 5.6. **qRT-PCR analysis of transgene expression in rice.** Average Ct (threshold fluorescence) value over three technical replicates is shown. Lower values represent increased expression. RNA was extracted from a single expanded leaf, 1µg of RNA was used for cDNA production. Dark grey bars show rice *UBQ* gene (Os03g13170) expression, light grey bars show transgene expression. **A)** bHLH-OE lines **B)** LRR-RLK-OE lines

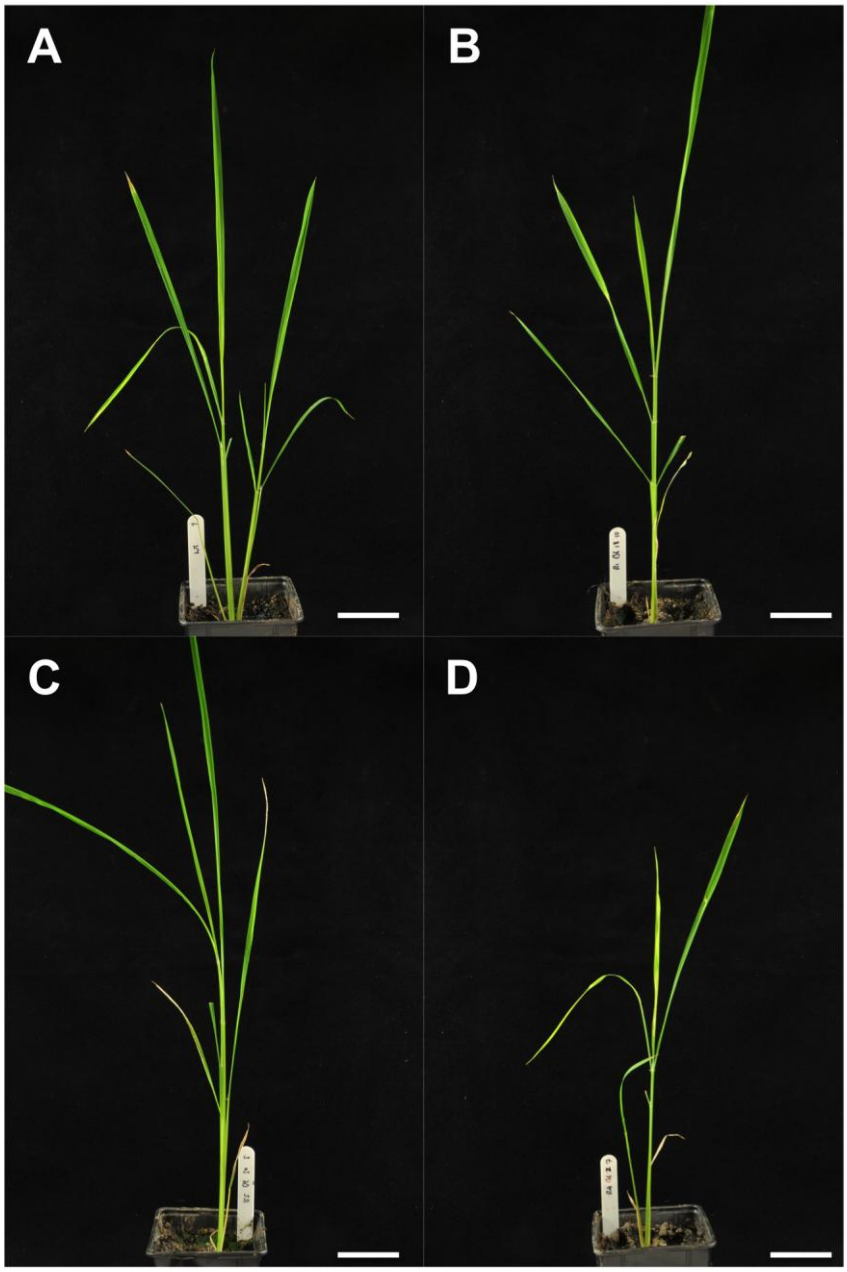


Figure 5.7. **Gross morphology of transformed T1 rice lines and control plants.** Photos taken 47 days after planting. **A)** WT **B)** R1 B6 18 1.10 **C)** R2 B6 20 1.5 **D)** EV B6 2 1.7 White scale bar = 5cm

were any quantitative differences in Kranz anatomy between the different lines, measurements of vein number, M cell number, inter-bundle sheath distance and inter-vein distance (IVD; as measured from vascular centres) were taken in addition to recording leaf width, leaf length, leaf number (1 being the oldest) and plant age. When analysing vascular patterning it is important to normalise for these latter measurements so that, for example, the width of the leaf is taken into consideration when comparing vein number. Figure 5.9 shows box plot representations of a key set of Kranz variables for bHLH-OE, LRR-RLK-OE, EV and WT lines. bHLH-OE and LRR-RLK-OE lines show no increase in vein number (Figure 5.9A) or vein number over either leaf width (Figure 5.9B) or leaf length (Figure 5.9C). Leaf dimensions (Figure 5.9D), IVD (Figure 5.9E) and M cell number (Figure 5.9G) also remained unchanged in bHLH-OE and LRR-RLK-OE lines compared to EV and WT lines. A significant difference was identified when IVD was normalised for leaf width, however, as WT plants exhibited significantly lower IVD per leaf width (μm mm^{-1}) than bHLH-OE, LRR-RLK-OE or EV plants ($p < 0.05$), which all displayed similar values (Figure 5.9F). Considering that there were no differences in vein density (Figure 5.9B) nor absolute IVD (Figure 5.9E) between the different lines, this result is likely artefactual. In conclusion, ubiquitous overexpression of either the bHLH-like or LRR-RLK candidate does not induce Kranz-type characteristics in rice.

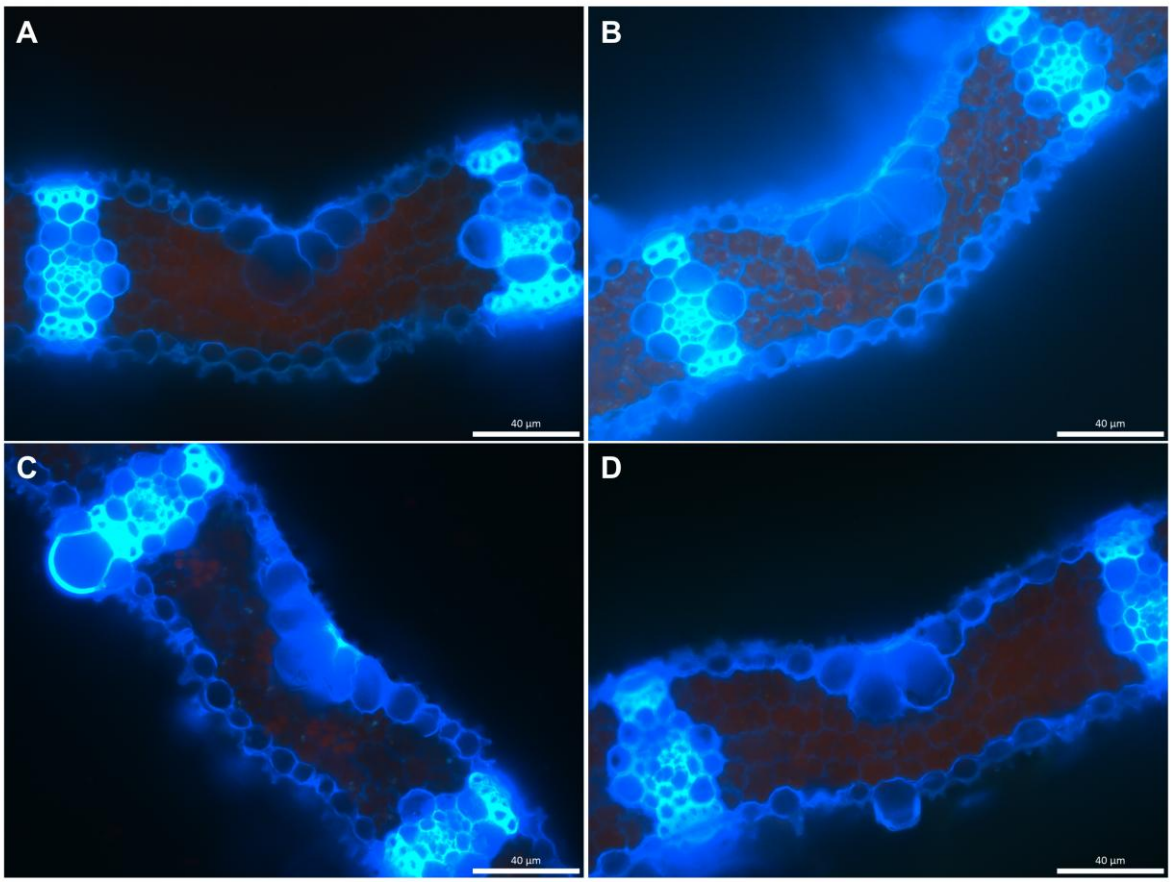


Figure 5.8. **Leaf sections of T1 rice lines and control plants.** Fresh leaf sections were dissected by hand and imaged using light microscopy with a UV filter. Pairs of intermediate veins were located between the 2nd and 3rd outermost lateral veins of 6th-8th leaves. **A)** WT **B)** R1 B6 18 1.9 **C)** R2 B6 34 1.1 **D)** EV B6 7 1.3**.**

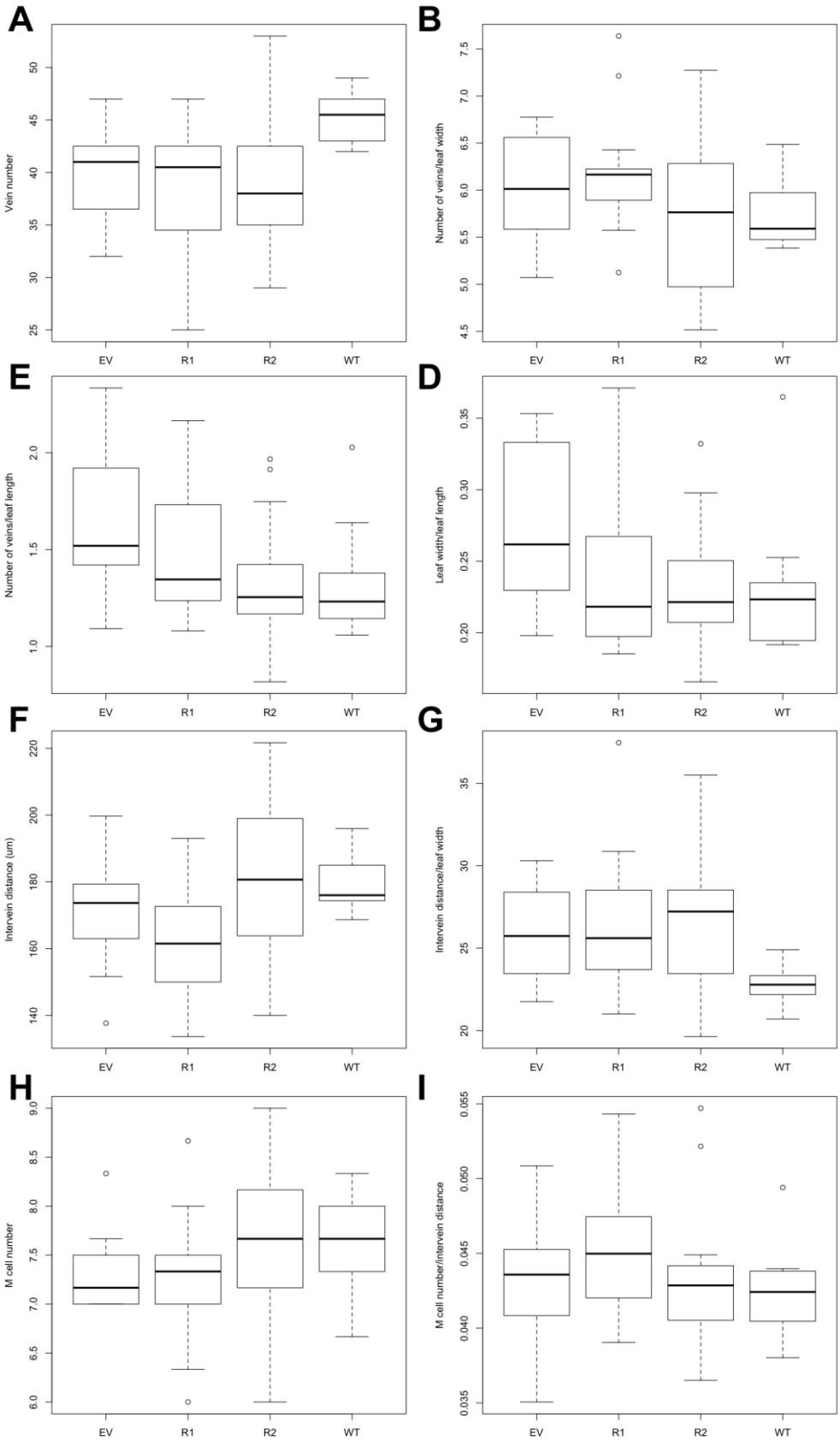


Figure 5.9. Boxplot diagrams of quantified leaf traits in T1 transformed rice lines. Figure legend is on opposite page.

Figure 5.9. Boxplot diagrams of quantified leaf traits in T1 transformed rice lines. Boxes indicate the 25th and 75th percentiles, black lines indicate the median values, whiskers indicate the extreme values still within 1.5 interquartile range of the upper and lower quartiles, open circles represent outlier values outside 1.5 interquartile range. **A)** vein number, **B)** vein density (veins mm^{-1} of leaf width) **C)** veins per leaf length (veins mm^{-1} leaf length) **D)** leaf dimensions (mm mm^{-1}) **E)** inter-vein distance (IVD) (µm) **F)** IVD per leaf width (µm mm^{-1}) **G)** M cell number **H)** M cells per IVD (M cell μm^{-1}). EV = Empty Vector, R1 = bHLH-OE, R2 = LRR-RLK-OE, WT = wild type.

5.2.4 *Setaria viridis* transformation led to low transformant regeneration rates

5.2.4.1 Construct design

Setaria viridis was used as a model to test the effects of reduction in gene expression in a Kranz species due to its increased tractability, short generation time and transformation potential compared to maize (Brutnell et al., 2010). *Setaria* orthologues of maize candidates were identified based on phylogenetic analyses of the genome sequence of *Setaria italica,* the domesticated form of *S. viridis*. To confirm that the target gene sequences were likely to be conserved between *S. italica* and *S. viridis,* orthologue sequences from five diverse grass species (*S. italica*, maize, sorghum, rice and *Brachypodium distachyon*) were aligned (data not shown). High levels of sequence conservation between these five species imply that the design of constructs to target *S. italica* gene expression by RNAi would have a similar effect in *S. viridis*. As such, approximately 300bp fragments were identified without extended sequence overlap with other *Setaria* genes (data not shown). These were cloned from *S. viridis* cDNA for insertion into the RNAi vector pANDA, which has been shown to reduce gene expression in grass species (Miki and Shimamoto, 2004).

5.2.4.2 *Setaria* transformation

Similar to rice, *Agrobacterium* mediated inoculation of embryogenic callus induced from germinating seed (A10 variety) was used as a means for *Setaria* transformation. In *Setaria*, however, transformation is a more recent and less well established technique than in rice (Brutnell et al., 2010). Despite extensive optimisation efforts during repeated transformation attempts, the need for further development of *Setaria* transformation methods is reflected in the much lower transformation efficiencies obtained in *Setaria* (Table 5.1). bHLH-RNAi, LRR-RLK-RNAi and MYB-RNAi transformation efficiencies were all similar to each other. The relatively high transformation efficiency of *Setaria* EV lines may be explained by a decrease in callus viability as a consequence of RNAi in the candidate gene lines, a selective advantage conferred by the presence of the Gateway cassette in pANDA-EV (this includes a chloramphenicol resistance gene) or, more likely, stochastic callus quality.

T0 transformation was confirmed by genomic PCR using primers designed to amplify the linker sequence of the RNAi hairpin loop. This approach was also used in the T1 generation, as well as an additional PCR approach using primers designed to amplify a stretch of sequence from the linker sequence to the post-hairpin *NOS* terminator sequence (Figure 5.10). This only identified a subset of previously determined transformants, possibly as a consequence of the difficulty of amplifying the RNAi hairpin secondary structure, or of generic difficulties associated with *Setaria* genomic PCR (e.g. non-specific amplification, PCR inhibitors). However the results of a hygromycin based germination screen, using seed from the same inflorescence shoots as were

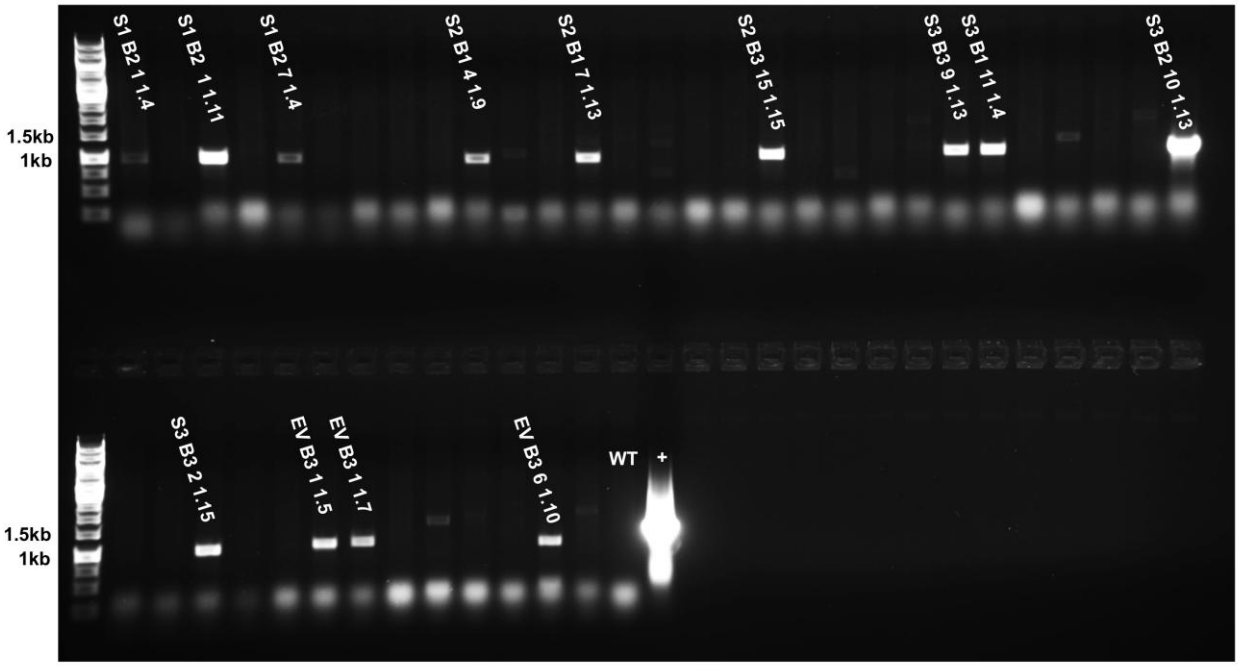


Figure 5.10. **Genomic PCR of T1 *Setaria* lines**. PCR from genomic DNA using primers designed to amplify from pANDA hairpin linker sequence (forward primer sequence 5'- CATGAAGATGCGGACTTACG-3') to *NOSt* sequence (reverse primer sequence 5'-ATTGCCAAATGTTTGAACGA-3'). Vector control (+) represents PCR from original binary vector, WT from untransformed DNA.

used for PCR screening, support the results of the linker primer screen (data not shown). As with rice, technical difficulties meant that Southern blot analyses failed to validate the results of the PCR screen.

Analysis of T1 transformation rates in *Setaria* also suggest chimeric T0 tissue ontogeny (Table 5.2). In *Setaria* this is supported by the observation that different *Setaria* T0 inflorescence shoots produced seed with variable rates of resistance to hygromycin (data not shown). Therefore as with the rice lines, following preliminary phenotypic analyses of three PCR-positive T0 lines per construct, phenotypic characterisation was focused on T1 plants.

5.2.4.3 Expression analyses of T1 *Setaria* lines

Analysis of gene knockdown in *Setaria* was more complex than quantification of transgene expression in rice. As Kranz candidates were selected based on high levels of gene expression in meristems and young leaf primordia relative to other tissues (Figure 5.1), analysis of gene expression required isolation of shoot apices. Despite the lack of obvious inflorescence emergence when sampled 33 days after planting, on removal of outer leaves it became clear that the majority of plants had undergone a transition to floral growth and young, but differentiated, floral tissue was apparent. Despite the lack of vegetative meristem tissue in the majority of shoots, it was still possible to sample young and expanding leaf primordia. Therefore from each plant, depending on the progression of floral growth, young vegetative or inflorescence meristems, young inflorescences and expanding leaf primordia

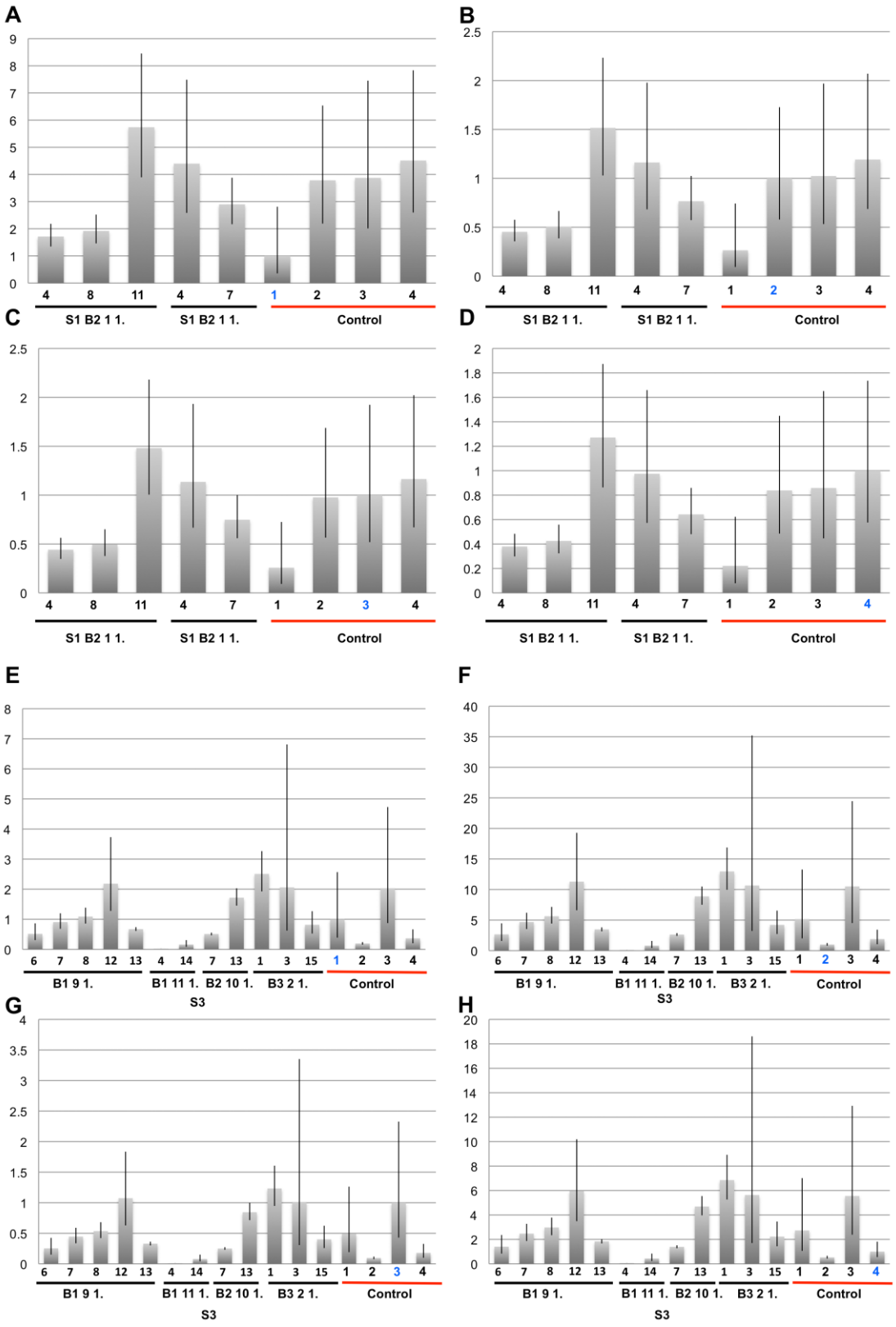


Figure 5.11. qRT-PCR analysis of transcript levels in *Setaria.* Figure legend is on opposite page.

Figure 5.11. qRT-PCR analysis of transcript levels in *Setaria.* Y-axis shows fold change of target gene expression in analysed line over control, normalised for endogenous *UBQ* (Si037467) expression (as calculated by ΔΔCt method). Average ΔΔCt value over three technical replicates is shown. **A-D)** bHLH-RNAi lines, **E-H)** MYB-RNAi lines. Four control lines are shown furthest right in all panels underlined in red. In each case the control used for the fold change measurement in that panel is highlighted in blue (so in all cases fold change = 1 for this control). Control 1 = EV B3 1 1.5, Control 2 = EV B3 6 1.10, Control 3 = Wild type individual 3, Control 4 = Wild type individual 5. (A, E) Fold change is calculated using EV B3 1 1.5 as a control, (B, F) Fold change is calculated using EV B3 6 1.10 as a control, (C, G) Fold change is calculated using WT 3 as a control, (D, H) Fold change is calculated using WT 5 as a control. Error bars represent pooled standard deviation of target gene and *UBQ* expression in the line used for analysis.

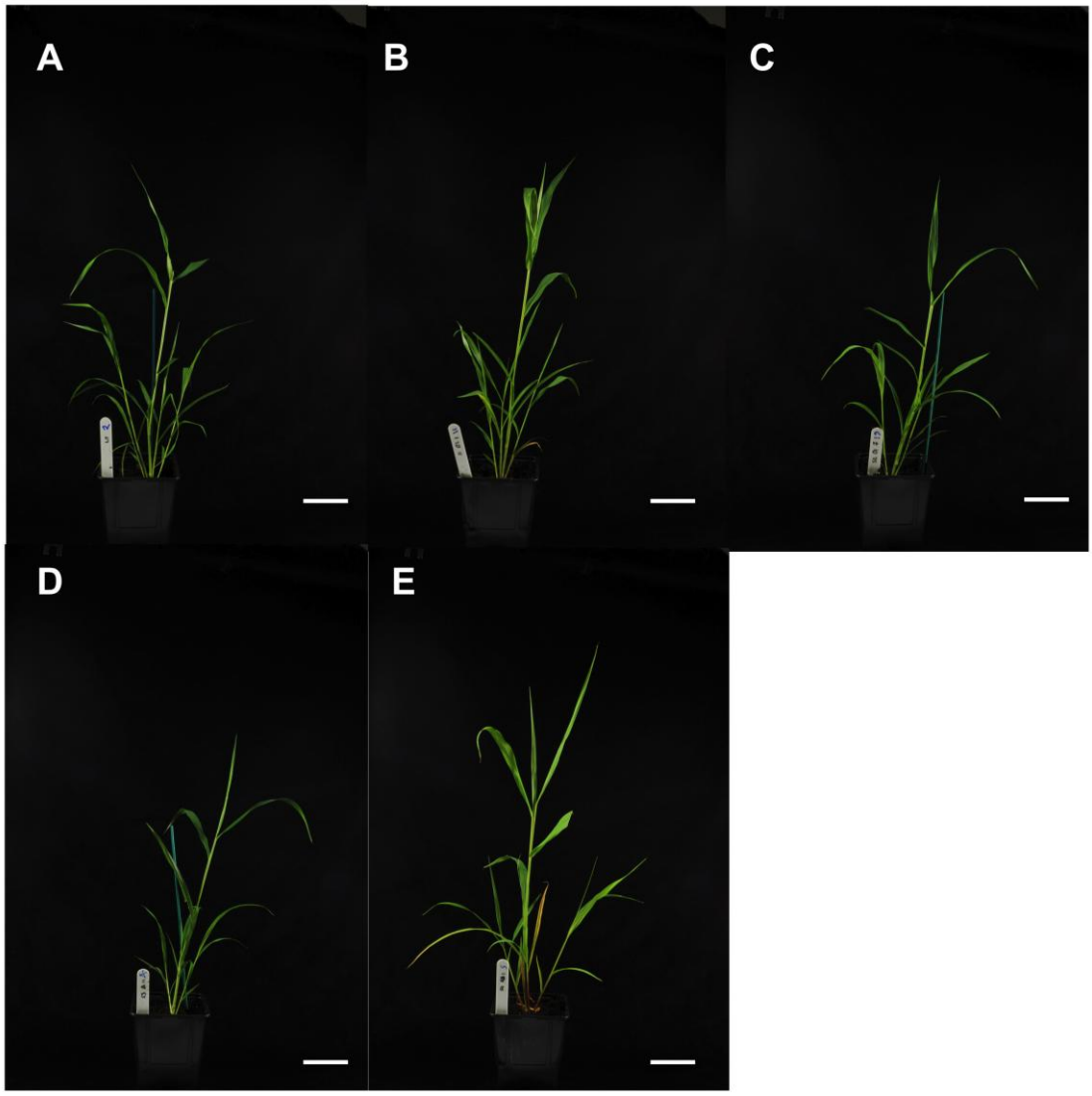


Figure 5.12. **Gross morphology of transformed T1 *Setaria* lines and control plants.** Photos taken 33 days after planting**. A)** WT **B)** S1 B2 1 1.11 **C)** S2 B1 7 1.13 **D)** S3 B1 11 1.4 **E)** EV B3 1 1.5 White scale bar = 5cm.

were pooled from three axillary shoots per transformed individual. As the LRR-RLK candidate is expressed almost exclusively in meristematic tissue, however, it was not possible to quantify candidate gene expression in LRR-RLK-RNAi lines. Due to the variance in developmental state of the T1 and wild type (WT) population sampled, four negative WT (2) and EV (2) controls were employed to compare gene expression in transformed lines against that captured the range of developmental stages in the T1 plants (bHLH-RNAi - Figure 5.11A-D, MYB-RNAi – Figure 5.11E-H). If gene expression was truly reduced in a transformed line, expression was expected to be lower than in all four controls. Endogenous candidate gene expression was normalised to a *UBQ* control and compared to each negative control using the delta-delta Ct method (Livak and Schmittgen, 2001). Figure 5.11A-D shows that endogenous gene expression is not reduced in any of the bHLH-RNAi lines below the level of all four negative controls. The expression of the bHLH-like candidate is significantly lower ($p < 0.05$) in two siblings, S1 B2 1 1.4 and S1 B2 1 1.8, than three of the negative controls (EV B3 6 1.10, WT 3 and WT 5) that exhibit similar expression levels. However, of all the bHLH-RNAi samples and controls, bHLH-like expression is lowest in the EV B3 1 1.5 control line. Control expression patterns are different when endogenous MYB candidate expression is tested, in that EV B3 1 1.5 and WT 3 expression is highly similar and EV B3 6 1.10 and WT 5 expression is similar. In the MYB-RNAi lines, however, S3 B1 11 1.4 MYB expression is reduced significantly below that of all control lines ($p < 0.0001$). MYB expression in S3 B1 11 1.4 sibling S3 B1 11 1.14 is lower than in all the control lines, however, it is only significantly lower ($p < 0.05$) in three out of 4 controls. In summary, while pVec8 led to

strong transgene expression in all transformed rice lines, pANDA induced gene knockdown in *Setaria* was less successful with conclusive evidence for a reduction in gene expression in only a single individual. A lack of developmental uniformity across the T1 population is likely to have weakened this analysis.

5.2.5 Knockdown of the MYB candidate in *Setaria* does not perturb Kranz development

To determine whether reduction in gene expression perturbs Kranz development, similar phenotypic analyses to those described in rice were carried out in *Setaria*. Figure 5.12 shows that the gross morphology of bHLH-RNAi, LRR-RLK-RNAi and MYB-RNAi lines is no different to that of EV or WT *Setaria* lines. Likewise, there also appeared to be no difference between hand cut leaf sections of bHLH-RNAi, LRR-RLK-RNAi, MYB-RNAi, EV and WT *Setaria* (Figure 5.13). Leaf development was quantified as described in rice, however, M cell number was not included in phenotypic analyses of *Setaria* as no deviation from two M cells was recorded in any of the lines. Due to the increased ease of identification under UV light relative to rice, counts of *Setaria* abaxial and adaxial stomata were also recorded (Figure 5.14). Because epidermal patterning genes were identified as upregulated concurrently with vascular elaboration in the transcriptomic analyses described in the Chapter 4, and the MYB candidate shows homology to *MIXTA*-like epidermal regulators (Figure 5.4), it is possible that epidermal

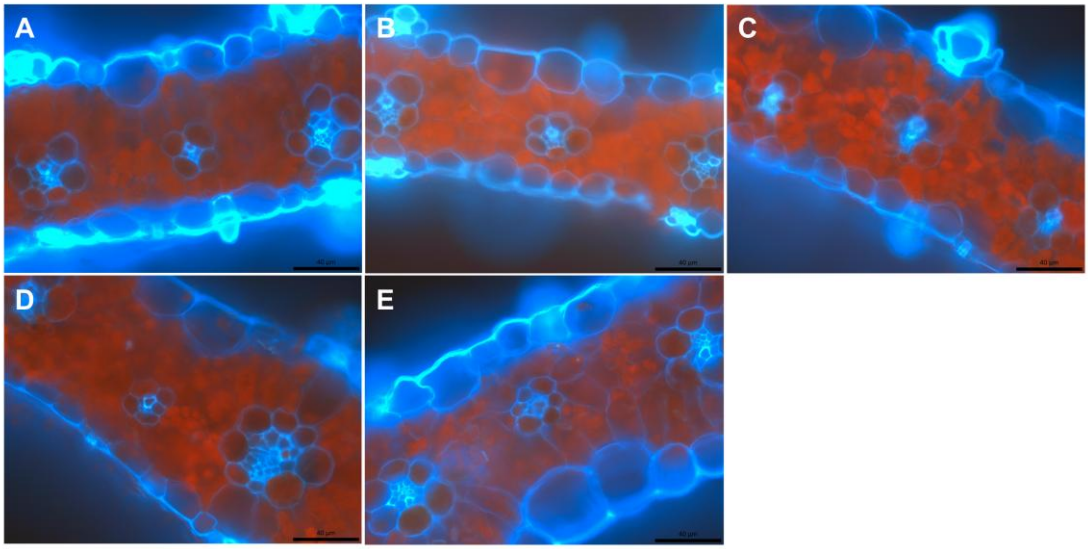


Figure 5.13. Leaf sections of T1 *Setaria* lines and control plants. Fresh leaf sections were dissected by hand and imaged using light microscopy with a UV filter. Pairs of intermediate veins were located between the 2nd and 3rd outermost lateral veins of 6th-9th leaves. **A)** WT **B)** S1 B2 1.11 **C)** S2 B1 7 1.13 **D)** S3 B1 11 1.4 **E)** EV B3 1 1.5.

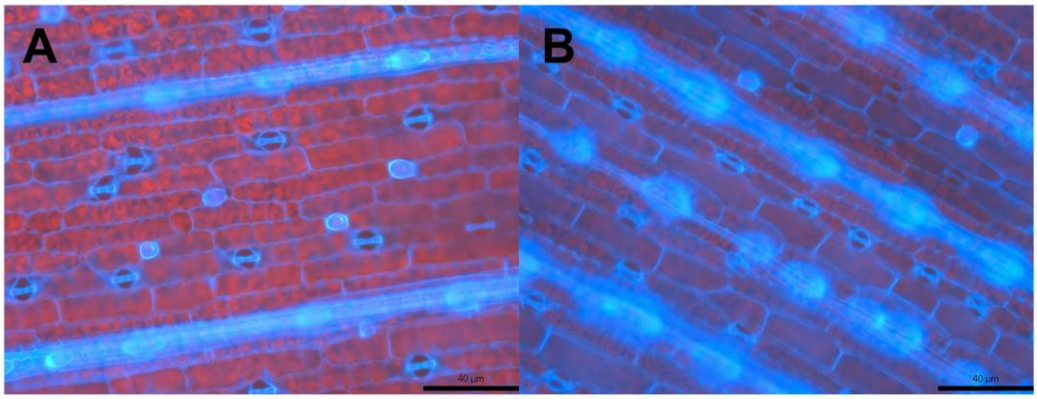


Figure 5.14. **Wild type *Setaria* stomata patterning.** Fresh 6th-9th leaf sections were dissected by hand and imaged using light microscopy with a UV filter. **A)** Abaxial leaf surface, **B**) Adaxial leaf surface.

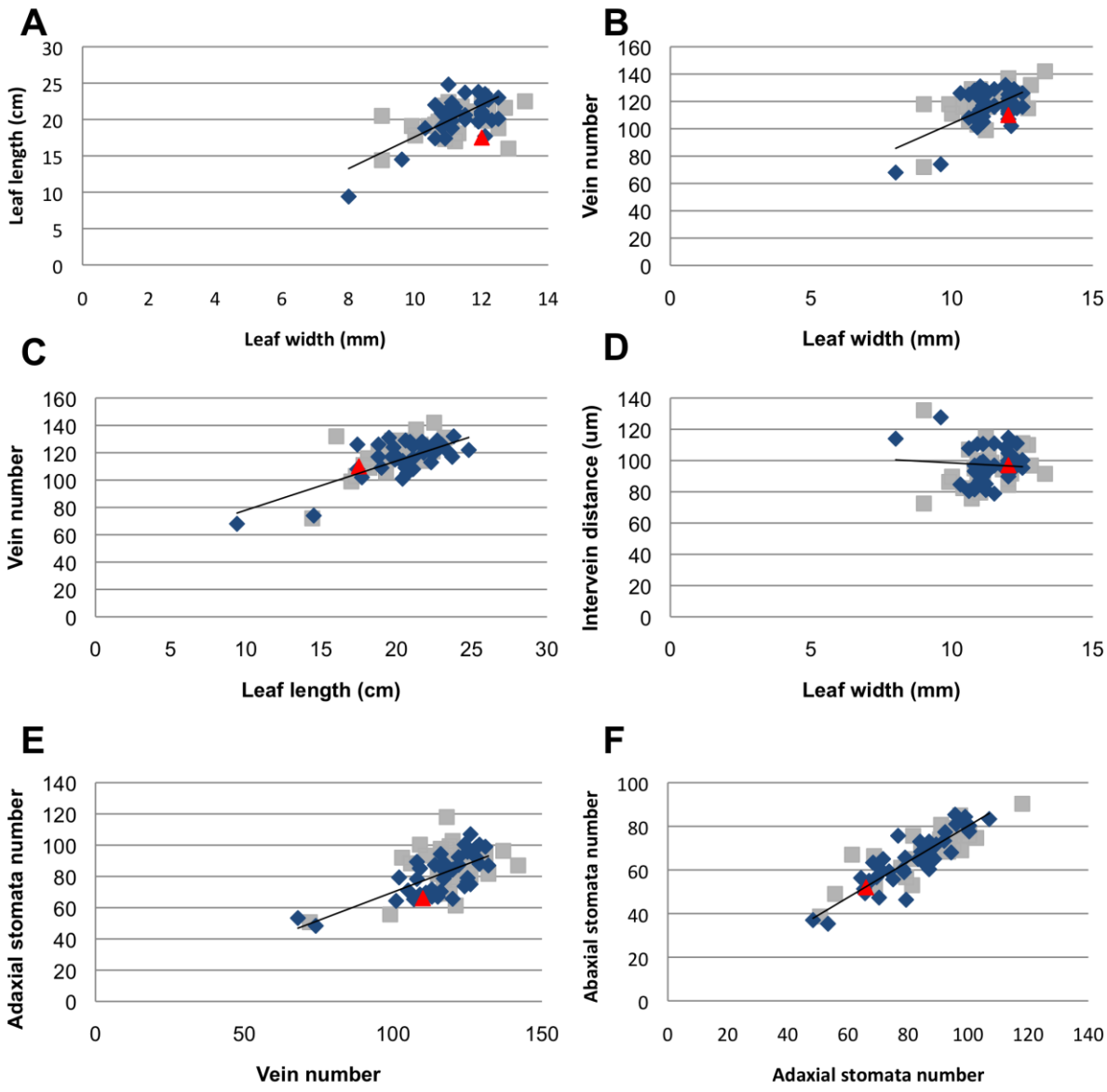


Figure 5.15. Correlations of *Setaria* leaf traits. A) leaf length and leaf width, **B)** vein number and leaf width, **C)** vein number and leaf length, **D) i**nter-vein distance (IVD) and leaf width, **E)** adaxial stomata and vein number, **F)** abaxial stomata and adaxial stomata. Blue diamonds – pooled wild type (WT) and empty vector (EV) lines, grey squares- pooled transformed S1, S2 and S3 lines without significant reduction in expression, red triangle – S3 B1 11 1.4. Trendlines represent WT and EV data.

patterning may be perturbed as a consequence of candidate gene misexpression. Due to the conclusive identification of only a single individual with reduced candidate expression, the use of boxplots to represent the measurements taken of *Setaria* leaf data for each line would be uninformative. Instead, Figure 5.15 shows correlations between different key leaf developmental traits for WT/EV plants, transformed plants with no strong evidence for reduced expression and line S3 B1 11 1.4, in which reduced Myb expression has been demonstrated. If leaf development is perturbed in individual plants, it would be expected that these plants deviate from the highly significant WT/EV correlations as shown by the trendlines in Figure 5.15A,B,C,E,F (Pearson's $r^2 > 0.28$, $p < 0.00001$ in all cases). There was no significant correlation for IVD and leaf width in *Setaria* (Figure 5.14D). Correlations of S3 B1 11 1.4 leaf length and leaf width (Figure 5.15A), vein number and leaf width (Figure 5.15B), vein number and leaf length (Figure 5.15C), adaxial stomata number and vein number (Figure 5.15E) and adaxial and abaxial stomata number (Figure 5.15F) appear similar to those measured in WT/EV plants, and within the range of WT/EV data. This suggests that knockdown of the MYB candidate in *Setaria* does not affect leaf development, however, with data for only a single individual subtle defects may be hidden. The distribution of trait data for WT/EV and transformed plants without evidence for knockdown of gene expression are also overlapping, suggesting that the presence of an RNAi construct is not perturbing development in these lines (Figure 5.15).

5.2.6 Pooled leaf trait measurements reveal rice and *Setaria* growth parameters

As the bHLH-OE, LRR-RLK-OE, bHLH-RNAi, LRR-RLK-RNAi and MYB-RNAi lines described above do not appear different to WT and EV lines (Figure 5.9, 5.15), it was possible to pool the quantified leaf data for all T1 samples to provide an extensive outline of the parameters of leaf growth for the two species. In addition, the size of this dataset means that leaf growth in future transgenic Kranz lines can be compared to WT patterns of growth with strong statistical support.

To first provide insight into the developmental programmes that operate in rice and *Setaria* leaves, the covariances between different leaf traits were analysed. Table 5.3 shows the Pearson correlation coefficients for all developmental traits measure in rice, significant correlations ($p < 0.05$) are shown in yellow. As expected, leaf width and length correlate strongly with vein number, M cell number and IVD, suggesting that as leaves expand more veins initiate, and that veins also become more spread out with increased leaf size. Data on plant age show that the increase in vascular distance operates secondarily to vascular initiation. As plant age correlates with IVD, but not with leaf width or vein number, it implies that leaf width and vascular patterning are determined early in development, but that as plants mature IVD starts to expand. The key roles of leaf size and plant age in determining IVD, rather than the absolute number of veins, is shown by the lack of correlation between vein number and IVD. A strong correlation between M cell number

Trait 1	Trait 2	R^2 value	*P* value
Plant age	Leaf number	0.017	0.174830000000
Plant age	Leaf width	0.024	0.135390000000
Plant age	Leaf length	0.007	0.272610000000
Plant age	Vein number	0.005	0.305020000000
Plant age	M cell number	0.028	0.113580000000
Plant age	Inter-BS	0.070	0.027704000000
Plant age	Inter-VD	0.055	0.045025000000
Leaf number	Leaf width	0.422	0.000000071536
Leaf number	Leaf length	0.409	0.000000124680
Leaf number	Vein number	0.457	0.000000014131
Leaf number	M cell number	0.032	0.099707000000
Leaf number	Inter-BS	0.080	0.020147000000
Leaf number	Inter-VD	0.096	0.011838000000
Leaf width	Leaf length	0.451	0.000000018933
Leaf width	Vein number	0.628	0.000000000002
Leaf width	M cell number	0.065	0.032270000000
Leaf width	Inter-BS	0.410	0.000000122560
Leaf width	Inter-VD	0.453	0.000000017166
Leaf length	Vein number	0.242	0.000090949000
Leaf length	M cell number	0.102	0.009922800000
Leaf length	Inter-BS	0.310	0.000007580600
Leaf length	Inter-VD	0.307	0.000008502500
Vein number	M cell number	0.000	0.533800000000
Vein number	Inter-BS	0.026	0.125940000000
Vein number	Inter-VD	0.033	0.097292000000
M cell number	Inter-BS	0.331	0.000003327600
M cell number	Inter-VD	0.336	0.000002718000
Inter-BS	Inter-VD	0.976	0.000000000002

Table 5.3. **Correlations between rice leaf traits.** Data pooled from WT, EV, R1 and R2 lines. *P* value calculated using Pearson's correlation coefficient. Yellow highlighting shows significant correlations ($p < 0.05$). Inter-BS = distance between BS cell edges of neighbouring veins, inter-VD = distance between centre of vasculature of neighbouring veins.

Trait 1	Trait 2	R^2 value	*P* value	Trait 1	Trait 2	R^2 value	*P* value
Plant height	Tiller number	0.46567074	0.000000000016	Leaf number	Leaf length	0.291288972	0.000000416670
Plant height	Plant age	0.135648589	0.999330000000	Leaf number	Vein number	0.353622349	0.000000014505
Plant height	Leaf number	0.025177977	0.910010000000	Leaf number	Inter-BS	0.259719342	1.000000000000
Plant height	Leaf width	0.004867368	0.721230000000	Leaf number	Inter-VD	0.258831464	1.000000000000
Plant height	Leaf length	0.036876917	0.051804000000	Leaf number	Adaxial stomata	0.287879333	0.000000496750
Plant height	Vein number	0.003207262	0.317070000000	Leaf number	Abaxial stomata	0.368021644	0.000000006400
Plant height	Inter-BS	0.038958695	0.952920000000	Leaf width	Leaf length	0.2758386	0.000000918620
Plant height	Inter-VD	0.012427546	0.826130000000	Leaf width	Vein number	0.332144283	0.000000047652
Plant height	Adaxial stomata	0.016966293	0.136020000000	Leaf width	Inter-BS	0.000469366	0.572180000000
Plant height	Abaxial stomata	0.00342409	0.311450000000	Leaf width	Inter-VD	0.000645563	0.415520000000
Tiller number	Plant age	0.095143144	0.996030000000	Leaf width	Adaxial stomata	0.000947045	0.601970000000
Tiller number	Leaf number	0.006980056	0.241110000000	Leaf width	Abaxial stomata	0.007815254	0.771480000000
Tiller number	Leaf width	0.003795053	0.302320000000	Leaf length	Vein number	0.458098394	0.000000000026
Tiller number	Leaf length	0.038975373	0.047048000000	Leaf length	Inter-BS	0.197332859	0.999960000000
Tiller number	Vein number	0.012518937	0.172980000000	Leaf length	Inter-VD	0.13052437	0.999160000000
Tiller number	Inter-BS	0.025467399	0.911270000000	Leaf length	Adaxial stomata	0.058451624	0.019666000000
Tiller number	Inter-VD	0.003380238	0.687430000000	Leaf length	Abaxial stomata	0.115442944	0.001636200000
Tiller number	Adaxial stomata	0.015138331	0.149860000000	Vein number	Inter-BS	0.529276539	1.000000000000
Tiller number	Abaxial stomata	0.008961204	0.212830000000	Vein number	Inter-VD	0.465148104	1.000000000000
Plant age	Leaf number	0.172299594	0.000130340000	Vein number	Adaxial stomata	0.356270433	0.000000012495
Plant age	Leaf width	0.038863767	0.047288000000	Vein number	Abaxial stomata	0.426519771	0.000000000192
Plant age	Leaf length	0.072004321	0.010856000000	Inter-BS	Inter-VD	0.902307617	0.000000000002
Plant age	Vein number	0.213181736	0.000019641000	Inter-BS	Adaxial stomata	0.507003543	1.000000000000
Plant age	Inter-BS	0.116522821	0.998440000000	Inter-BS	Abaxial stomata	0.609237737	1.000000000000
Plant age	Inter-VD	0.13164406	0.999200000000	Inter-VD	Adaxial stomata	0.483140775	1.000000000000
Plant age	Adaxial stomata	0.138322063	0.000597910000	Inter-VD	Abaxial stomata	0.576788478	1.000000000000
Plant age	Abaxial stomata	0.2117564	0.000021007000	Adaxial stomata	Abaxial stomata	0.76153374	0.000000000002
Leaf number	Leaf width	0.057141185	0.020837000000				

Table 5.4. **Correlations between *Setaria* leaf traits.** Data pooled from WT, EV, S1, S2 and S3 lines (excluding S3 B1 11 1.4). *P* value calculated using Pearson's correlation coefficient. Yellow highlighting shows significant correlations ($p < 0.05$). Inter-BS = distance between BS cell edges of neighbouring veins, inter-VD = distance between centre of vasculature of neighbouring veins.

and IVD conflicts with an earlier report that showed no significant correlation between M cell number and IVD (Smillie et al., 2012), and suggests that M cell division may be a key component of IVD. Table 5.4 highlights the significant correlations that occur between different aspects of *Setaria* development. Plant height and tiller number were measured prior to leaf characterisation and, except for tiller number and leaf length, appear poor indicators of leaf characteristics. Conversely to rice, *Setaria* plant age correlates with leaf width and vein number, suggesting that *Setaria* leaves expand for longer and/or later than rice leaves (leaves were sampled at equivalent periods: rice leaves between 34-55 days after planting and *Setaria* leaves between 35-52 days after planting). Also unlike rice, there is no correlation between either plant age or leaf width with IVD, suggesting that there is little space for expansion between *Setaria* veins. *Setaria* leaves are effectively saturated with veins. As such and similarly to rice, wider (and longer) leaves also exhibit more veins. Stomata numbers increase significantly with leaf length (perhaps reflecting the basipetal gradient of grass stomata formation (Croxdale, 2000)) but not leaf width. However they correlate most strongly with vein number, highlighting the importance of the underlying vascular pattern in determining stomatal numbers.

To determine the extent to which individual traits explain the variance within the leaf datasets, a principal component analysis (PCA) was carried out using data from traits specifically related to leaf morphology (Figure 5.16). Figure 5.16A and B show that a single component can explain the vast majority of the variance with the rice (0.87) and *Setaria* (0.76) datasets respectively.

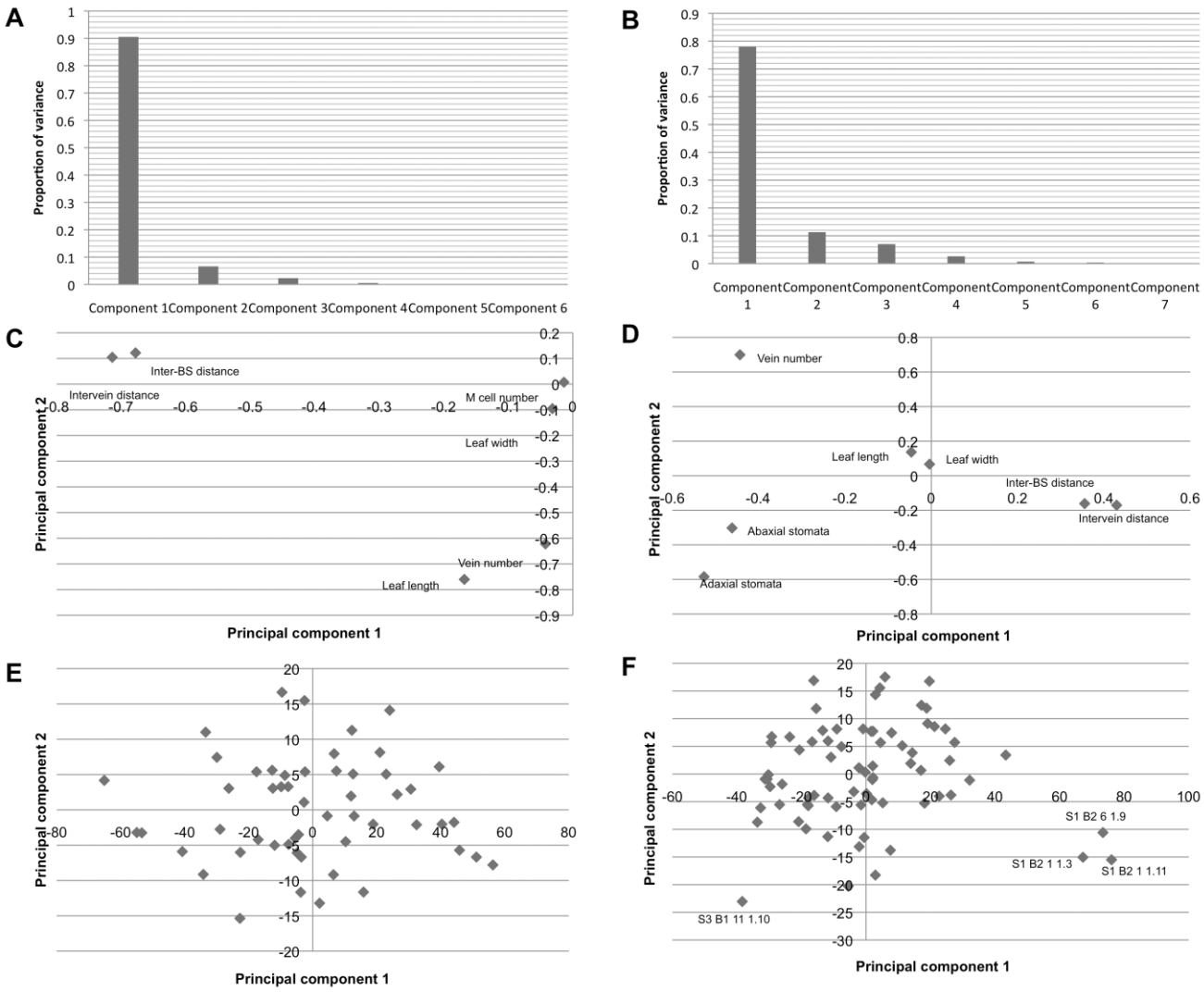


Figure 5.16. Principal component analysis of rice and *Setaria* leaf growth parameters. A, B) Proportion of variance explained by individual components, **C, D)** Loadings diagram of contribution of individual traits for principal component 1 and principal component 2, **E, F)** Score diagram of individual sample contributions to principal component 1 and principal component 2. (A, C, E) rice, (B, D, F) *Setaria.*

Comparison with principal component 2, which explains more than 5% of the variance in both rice and *Setaria*, reveals no positive rice trait effects within principal component 1 (Figure 5.16C). In *Setaria*, however, much of the variance in principal component 1 is explained by changes in IVD (Figure 5.16D). The clustering of traits within the PCA loading diagrams shown in Figure 5.16C and D show that leaf width variance is more similar to leaf length variance in *Setaria* than in rice, and that vein number is more closely associated with leaf length than leaf width in rice. The PCA score diagrams shown in Figure 5.16E and F, which reflect associations between individuals rather than traits, reinforce the similarity of the different transgenic lines in both rice and *Setaria*. Of the three bHLH-RNAi individuals highlighted as forming a distinct group in Figure 5.16F, two are non-transgenic siblings (S1 B2 1 1.3 and S1 B2 1 1.11), suggesting that the unique characteristics of these individuals are not related to transgenic effects. No distinct clusters appear to exist in the rice data (Figure 5.16E).

5.2.7 T-DNA insertions in *Arabidopsis* candidate gene orthologues do not reveal novel vascular patterning defects

To determine whether the candidate genes function in vascular development in *Arabidopsis*, T-DNA lines of candidate orthologues were ordered from the Nottingham *Arabidopsis* Stock Centre (NASC). 40 lines were ordered for orthologues from the preliminary candidate gene list based on gene expression across 6 samples, of which 8 were also on the putative positive candidate gene list described in Chapter 4 (including the MYB candidate

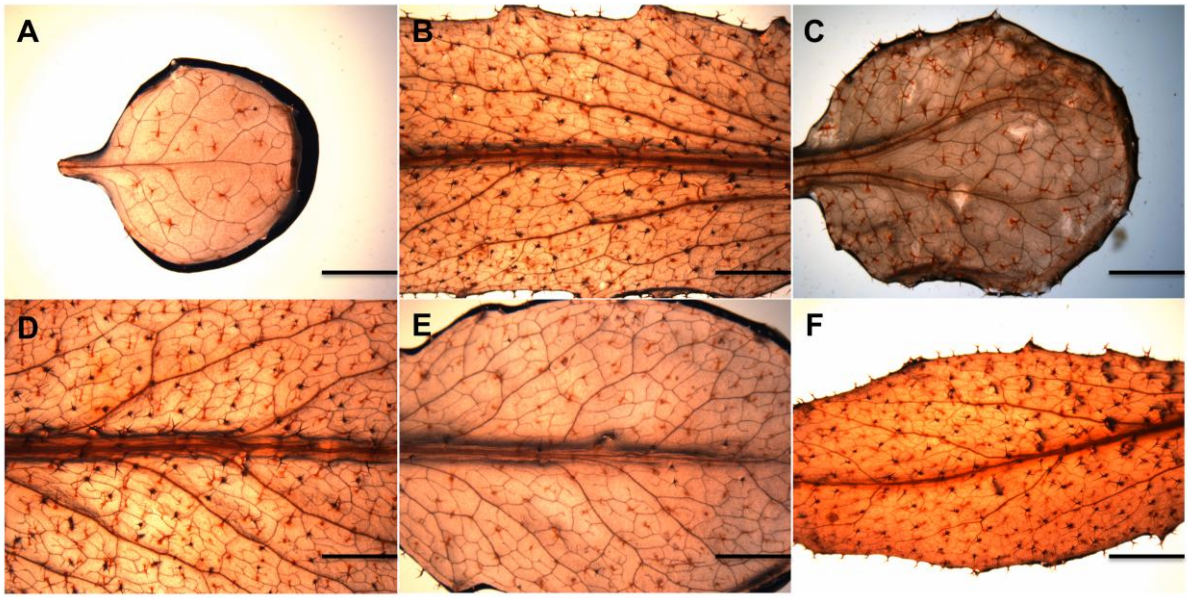


Figure 5.17. Light microscopy of ethanol cleared *Arabidopsis* leaves of homozygous T-DNA insertion lines for Kranz candidate orthologues. A) WT cotyledon **B)** WT juvenile leaf **C)** GRMZM2G150011 orthologue (C2H2 ZnF, *dot5*) **D)** GRMZM2G162020 orthologue (IQ domain protein, At4g00820) **E)** GRMZM2G097275 orthologue (SBP domain, At1g27370) **F)** GRMZM2G040924 (MYB, At3g61250), black scale bars = 2mm.

tested here). Of the 40 lines, 25 were annotated as homozygous, 11 as segregating and 4 were not annotated. A single juvenile leaf was isolated from 3 individuals and cleared using ethanol before analysis by light microscopy. Figure 5.17 shows leaves from 4 homozygous lines of orthologues on the putative positive regulator list. Of the 40 lines tested, the only line to exhibit vascular patterning defects was a C2H2 ZnF orthologue, which displayed a previously described cotyledon split-midvein phenotype (Figure 5.17C) (Petricka et al., 2008).

Maize Ac/Ds element collections were also screened to check for phenotypes in lines in which candidate gene function was perturbed. Nine lines were identified with an insertion in a candidate gene from the preliminary candidate list based on expression across 6 samples, of which 4 were also in the final putative positive candidate list. Insertions were confirmed by PCR and positive individuals selfed or crossed to positive siblings. Vein number and spacing were checked for 10 offspring in each line by light microscopy, however, none appeared different to WT in terms of Kranz patterning (data not shown).

5.2.8 Rice tissue culture generated a chlorophyll enriched mutant

During the screening of T1 rice plants, two of the three LRR-RLK-OE lines exhibited highly stunted individuals with short internodes and dark green short/squat leaves (Figure 5.18A). Analysis of leaf sections revealed distantly spaced veins (Figure 5.18B) and chlorophyll measurements showed

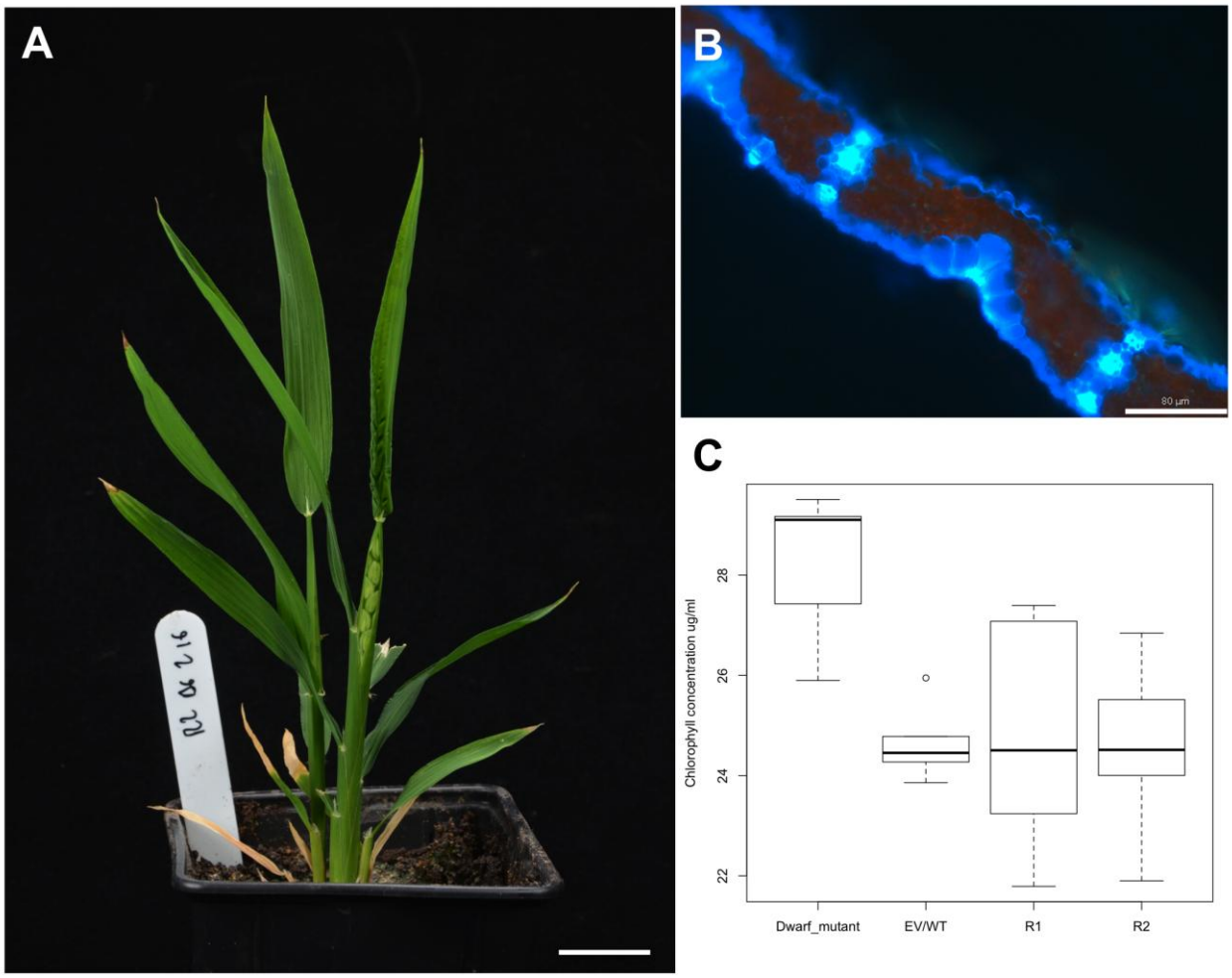


Figure 5.18. Tissue culture induced rice dwarf mutant. A) Gross plant morphology 64 days after planting (white scale bar = 2cm) **B)** Hand cut fresh leaf section imaged using light microscopy with a UV filter. **C)** Chlorophyll concentrations of dwarf mutants are significantly higher than empty vector/wild type, R1 and R2 lines ($p < 0.01$).

significantly ($p < 0.01$) higher chlorophyll concentrations than in other LRR-RLK-OE plants, bHLH-OE or WT/EV lines (Figure 5.16C). 5/20 individuals displayed this phenotype in one line and 3/20 individuals in the other. PCR screening for both the maize candidate sequence and the *HPT* gene showed that two of the dwarf plants were non-transgenic (data not shown), suggesting the phenotype was not a consequence of the transgene, but instead was a (recessive) mutation that occurred during tissue culture. As the phenotype was identical in the two lines in which it occurred, it is likely that these two lines were derived from the same piece of callus during plant regeneration.

5.3 Discussion

5.3.1 A combined approach of overexpression and knockdown is necessary for effective elucidation of Kranz candidate function

As Fatboy Slim famously sang on his classic 90's big beat album You've Come A Long Way Baby: 'build it up, tear it down' (Cook, 1998). While the inspiration behind his lyrics remains unclear, his words have strong resonance with the approach taken to determine the genetic regulation of Kranz anatomy in this study. Despite the lack of insight into Kranz regulation that the results presented in this chapter provide, a combined strategy of 'building' Kranz anatomy in rice and 'tearing' it down in *Setaria* offers the best opportunity to decipher the Kranz enigma (Fouracre, Ando and Langdale *In prep*). Without a fuller understanding of the genetic networks that regulate Kranz development, it is difficult to predict whether an overexpression approach in a C_3 species or

knockdown in a C_4 species is more likely to elucidate candidate gene function (assuming positive regulation of Kranz development). An overexpression-only approach suffers from the potential problem that individual genes may be necessary, but not sufficient, for aspects of Kranz development and that therefore overexpression of a single gene may not alter leaf phenotype. Furthermore, to reveal candidate function, heterologous expression may need to act at a specific level or within tight spatiotemporal restrictions and therefore ubiquitous overexpression may be uninformative. A strategy of overexpression also assumes transcriptional regulation of Kranz development. However, as one of the reasons for elucidating Kranz regulation is to introduce Kranz patterning into rice leaves (Hibberd et al., 2008; von Caemmerer et al., 2012), determining the effects of Kranz candidate expression in rice is of key importance. An approach restricted to reduction in gene expression is hindered by the potential problem of genetic redundancy, although mutant alleles of a single gene have been shown to affect Kranz development in maize (Slewinski et al., 2012). Knockout phenotypes in one species may also not reveal what a gene may potentially do in another, for example maize loss of function *knotted 1* (*kn1*) mutants display embryonic shoot defects (Vollbrecht et al., 2000) whereas overexpression of *KN1* in tobacco leads to altered leaf phenotypes (Sinha et al., 1993). An additional potential problem with knockout of candidates identified in maize in *Setaria* is that *Setaria* has an independent origin of Kranz anatomy to maize (Christin et al., 2009a). It is thus possible that the genes that regulate Kranz development in *Setaria* are distinct from those that operate in maize. However, the high similarity between the leaf structures of the two species, shared biochemical

pathways, repeated Kranz evolution within the grasses, close phylogenetic distance and shared descent from the same anatomical preconditioning event make it likely that the same regulatory networks have been co-opted in both lineages (Christin et al., 2007; Christin et al., 2009a; Christin et al., 2013). Taken together, the weaknesses of overexpression or knockdown approaches carried out in isolation make a combined and complementary strategy essential for effective functional characterisation of Kranz candidate genes.

5.3.2 Alternative expression strategies may provide further insights into Kranz development

The strategy of ubiquitous overexpression employed here acts effectively as a 'first pass'. It is an initial test to determine whether, when expressed in a rice leaf, a maize gene has any potential to modify form. Where maize genes do have a noticeable effect on rice leaves it is envisaged that these genes would then be expressed under more specific promoters that act within a narrower developmental range. For example, expression using primordia specific promoters, such as those of *TEOSINTEBRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR* (*TCP*) genes that were identified in the FA2 cluster (Figure 4.4B) and function in organ outrgrowth (Martin-Trillo and Cubas, 2010), or inducible promoters (that can be induced during young primordia growth) would provide insights into the effects of gene expression at specific developmental stages. Where primordia-specific or inducible promoters are shown to work (and lead to persistent phenotypes throughout leaf development) then these may be

preferable to ubiquitous expression for initial functional tests in rice. However, until these become available, ubiquitous overexpression is still a reasonable and practical strategy for the verification of the potential of maize gene activity in the context of a rice leaf. In the context of the development of a C_4 rice variety, much work is currently being done to identify M and BS specific promoters (Kajala et al., 2011) in order to create a functional C_4 pathway. As the candidate genes identified in this study act prior to clear BS and M cell elaboration, and were identified based on their potential to affect vein spacing, these would be of less use in this instance (although a case could be made for the expression of candidate genes with BS or M cell-specific expression patterns (Appendix 4.34) under such promoters).

Ubiquitous knockdown in *Setaria* remains a legitimate approach but, due to the limited reduction in expression observed in the bHLH-RNAi and MYB-RNAi lines (Figure 5.11), the pANDA vector may not be the most effective method of achieving this. Due to the non-uniformity of the tissues sampled in the *Setaria* T1 qPCR analysis, it is difficult to determine the extent of pANDA functionality. However, it is clear that it does not uniformly lead to strong reduction in expression. Expression analyses would be made easier by sampling a homozygous T2 population, in which multiple primary shoot apices could be harvested early in development without the need to maintain viable individuals. A generational delay would, however, add several months to the screening process. Alternatively, based on the insights this study has provided, the sampling of individual axillary tillers from positive T1 transformants much earlier in development would remove the problem of

reproductive differentiation and provide sufficient RNA for meaningful comparisons (M. Schuler personal communication). In terms of an alternative RNAi vector, the pANIC vector system is Gateway compatible and has been shown to function in monocots as diverse as switchgrass and rice (Mann et al., 2012).

5.3.3 Complementary expression strategies require effective transformation pipelines in multiple species

For a combined strategy of ectopic expression and knockdown to work, efficient transformation pipelines must operate in both the systems used for expression analyses. The data presented here suggest that while this may exist for overexpression in rice, rates of *Setaria* transformation are low enough to limit progress in testing the effects of reduction in gene expression (Table 5.1). Aside from the high rate of EV transformation, an average T0 transformation rate of ~3% is similar to that achieved by other members of the Langdale lab, and by those also working on *Setaria* transformation internationally (T. Brutnell, R. Furbank personal communication). Problems with breaking seed dormancy and the difficulties in genomic DNA amplification described above are also shared by the international C_4 *Setaria* community (T. Brutnell, R. Furbank personal communication). Thus, despite the quick generation time, small plant size, high seed set and increasing genetic resources (Brutnell et al., 2010; Li and Brutnell, 2011; Bennetzen et al., 2012; Jia et al., 2013; Xu et al., 2013), the effectiveness of *Setaria* as a C_4 (and bioenergy) model is restricted by its transformation potential. Higher

rates of transformation are possible, as demonstrated here (Table 5.1 EV line) and by others in the Langdale lab (>50% efficiency in one transformation batch, M. Schuler personal communication), the key to the success of such efficient batches most likely being the quality and type of callus selected. *Setaria* callus is much more heterogeneous in appearance than rice callus. With international efforts increasingly fervent, the refinement of callus selection, in addition to the optimisation of culturing conditions and callus induction, will likely lead to increased rates of transformation efficiency in the coming years. A significant breakthrough may come from the adoption of different *Setaria* lines for callus induction (T. Brutnell personal communication). To date, the lab strain A10 has been used almost ubiquitously in transformation experiments. Alternative accessions may prove more amenable to transformation, much as different rice varieties have been shown to exhibit different capacities for transformation (Nishimura et al., 2006; Hiei and Komari, 2008). Despite the relatively high efficiency of *Agrobacterium* mediated transformation reported for some maize genotypes, the low rates of transformation recorded in most lab varieties (Ishida et al., 2007) and the practical difficulties associated with growing large populations of maize plants, mean that *Setaria* is likely to remain the C_4 grass model of choice for genetic manipulation, regardless of the rate of transformation improvement.

5.3.3.1 Transformation pipelines may provide unexpected phenotypes

The identification here of a dwarf mutant that arose during callus regeneration (Figure 5.18) supports earlier work that demonstrates the mutagenic effects of

tissue culture. The regeneration of plants from undifferentiated rice and *Arabidopsis* cells has been shown to increase spontaneous mutation rates by over 50 fold, with most mutations arising from SNPs and short indels (Jiang et al., 2011; Miyao et al., 2012; Wang et al., 2013c). The dwarf, reduced internode, dark green phenotype identified here is highly reminiscent of rice giberrellic acid (GA) sensitivity (Ikeda et al., 2001) and biosynthesis (Sakamoto et al., 2004) mutants. As the phenotype was found in ~25% of individuals in affected lines in the T1 generation, it is likely that a recessive loss of function mutation gene in either the GA synthesis or signalling pathways has occurred. As a dwarf individual sprayed twice with 0.1mM GA failed to respond (data not shown), it is perhaps more likely that this mutation has occurred in the signalling pathway. However, as the tested individual was fully mature when sprayed, the spray treatment may have been too late to reveal the effects of GA application. The comparison of dwarf plant and unaffected sibling genome sequences would reveal insights into the location and type of mutation.

5.3.4 Insertion libraries remain a valuable resource for testing Kranz candidate function

Considering the difficulties in reducing gene expression in *Setaria*, the identification of pre-existing loss of function mutants in maize or in orthologous genes in other species (or equivalent activation tagged lines in rice in terms of overexpression) would reduce the need to generate heterologous lines. In maize, however, the relatively sparse coverage of Ds

insertions in the genome limits the contribution of this strategy. Maize Mutator lines (Mu) offer an increased number of genes with insertions, and a Mu line has been shown to affect Kranz patterning (Slewinski et al., 2012), although high insertion number and transposition rates make these lines less amenable to work with than Ds lines. The extensive number of *Arabidopsis* T-DNA insertion lines available at NASC and elsewhere make it possible to analyse loss of candidate orthologue function through a relatively high throughput screen based on leaf clearing. The analysis described above showed no novel effects on vascular patterning in candidate *Arabidopsis* orthologue T-DNA lines (Figure 5.17), although this analysis was only preliminary (insertion events were not validated for example) and lacked quantification of vascular patterning. As only 3 individuals were sampled and ~25% of the screened lines were segregating it is also possible that subtle homozygous phenotypes were missed in these cases. The phylogenetic distance of *Arabidopsis* to maize and the differences in leaf form between the two species mean that the absence of a mutant phenotype is not necessarily indicative of a lack of a role in vascular patterning. However, perturbed vascular development in an *Arabidopsis* candidate orthologue T-DNA line would give strong support for further investigation of gene function in the grasses. The verification of the *dot5* mutant phenotype (Figure 5.17C) provides justification for this approach. As new insertion lines are developed, or as the appearance of new data refines Kranz candidate lists, screening of insertion libraries should be actively maintained.

5.3.5 The three Kranz candidates tested here are low priorities for further functional characterisation

The data presented here show that overexpression of the bHLH-like and LRR-RLK candidate in rice (Figure 5.6) and knockdown of the MYB candidate in *Setaria* (Figure 5.11) have no effect on plant form or Kranz development (Figure 5.7-5.9, 5.12, 5.13, 5.15). Considering the arguments outlined above for the importance of combined overexpression and knockdown analyses to determine candidate gene function, the lack of such complementary data for all three candidates limits the extent to which conclusions can be drawn regarding their potential roles in plant development. This is especially the case for the MYB protein. The important developmental function that can be inferred from the likely lethality of overexpression in rice (Table 5.1) suggests that the lack of a *Setaria* knockdown phenotype (based on a single individual) is not indicative of the MYB's endogenous role. As such, it may be worth expressing the MYB under a primordia specific or inducible promoter as discussed above. However, considering that 283 genes have been identified as positive candidate regulators (Appendix 4.31), of which 48 are TFs (Table 4.4), and that plant transformation, regeneration and phenotypic analysis is a time consuming and lengthy process, efforts would be better focused on the characterisation of alternative candidate genes.

5.3.6 The quantification of rice and *Setaria* leaf growth parameters will improve future characterisation of transgenic candidate lines

Without understanding the fundamental patterns of growth that occur in rice and *Setaria* leaves, the assessment of the effects of candidate gene misexpression will be highly limited. The data presented here on the correlations between different aspects of leaf growth (Table 5.3, 5.4, Figure 5.16) provide a baseline picture of WT leaf development to which subsequent functional analyses can be compared. Deviations from the range of WT values identified here will suggest distinct programmes of leaf growth in mutant lines. While detailed observations of rice leaf development and vascular patterning have been made (Itoh et al., 2005; Smillie et al., 2012), until now such a framework has been lacking in *Setaria*. The observations made in this study show that core components of rice and *Setaria* leaf growth are shared – as leaves in both species grow wider they also become longer and initiate more veins. However, a key distinction between the two leaf types is that as rice leaves widen, IVD also increases, whereas this is not the case in *Setaria*. This is presumably a consequence of the increased rates of vascular initiation to facilitiate an effective C_4 pathway in *Setaria*. As rice veins spread out gaps are filled with more M cells (Table 5.3), in *Setaria* they are effectively filled with more veins.

The observation that IVD correlates significantly with M cell number in rice ($p < 0.0001$) conflicts with an earlier report that found no significant correlation between the two traits (Smillie et al., 2012). A significant correlation was found between IVD and M cell area, suggesting that cell expansion rather than cell division explained most of the variation in IVD (Smillie et al., 2012). It is unlikely that the differences between the two results are explained by leaf

sampling strategy as fully expanded 6^{th} (Smillie et al.) or 6^{th}-8^{th} (this study) leaves were sampled around 30 days post germination (Smillie et al.) or 34-55 days post sowing (this study). It is possible that minor vein pair selection varied between the studies. In this study IVD measurements were carried out between 3 pairs of neighbouring minor veins located between the second and third lateral vein from the leaf margin. The location of minor vein pairs used for IVD and M cell calculations is not specified by Smillie et al. As IVD and M cell number become more variable adjacent to the leaf margins and midrib (data not shown) distinct sampling across the mediolateral axis of the leaf may explain the discrepancy. Conversely, Kitaake rice was used in this study whereas IR64 was used by Smillie et al. Smillie et al found an average WT vein density of 4.8 veins mm^{-1} leaf width while the average WT vein density observed in this study was 5.8 veins mm^{-1} leaf width. The differences in M cell properties identified between the two reports may therefore reflect fundamental differences in developmental programmes that exist between rice accessions. This underlines the importance of the selection of appropriate rice germplasm for C_4 variety development. Having outlined the functional characterisation of novel regulators of C_4 growth, in the next chapter I will discuss work carried out on TFs already known to play a role in C_4 patterning, *GOLDEN2-LIKE* genes.

Chapter 6: The role of *GLK* genes in subcellular Kranz patterning

6.1 Introduction

Prior to the work presented in the preceding chapters, very few genes had been implicated in the regulation of Kranz development. Until the recent description of a maize *scarecrow* (*scr*) mutant (Slewinski et al., 2012), the only genes shown to have specific effects on Kranz anatomy have been involved in maize chloroplast development. Maize leaves develop dimorphic chloroplasts. BS cell chloroplasts lack granal stacks and PSII and contain Rubisco, while M cell chloroplasts are granal, have functional PSI and PSII but lack Rubisco. The localisation of Rubisco to BS cell chloroplasts is a common feature of all C_4 species, however, the dimorphic granal distribution is thought to be particularly important for NADP-ME type C_4 species such as maize (Dengler and Nelson, 1999). The maize gene *GOLDEN2* (*G2*), which regulates the differentiation of chloroplasts in maize BS cells, was the first gene shown to specifically regulate C_4 type chloroplast development (Langdale and Kidner, 1994; Hall et al., 1998). Loss of function *g2* mutants show perturbed differentiation of BS chloroplasts while M cell chloroplasts remain unaffected. In addition to the BS cell specific phenotype of loss of *g2* function, a cell specific role for *G2* and the paralagous gene *GOLDEN2-LIKE1*

(*GLK1*) is reinforced by their expression patterns: *G2* transcripts accumulate in BS cells while *GLK1* transcripts accumulate in M cells (Rossini et al., 2001).

Despite the effects of *g2* loss of function on C_4 development and cell specific expression patterns in maize, *GLK* gene duplicates were additionally found in the C_3 species moss, rice and *Arabidopsis* (Rossini et al., 2001; Fitter et al., 2002; Yasumura et al., 2005), suggesting that *GLK* gene duplication is not a C_4 specific phenomenon. Unlike in maize, however, expression of both duplicates occurs in all photosynthetic cell types of these species (Rossini et al., 2001; Fitter et al., 2002; Yasumura et al., 2005; Waters et al., 2008; Nakamura et al., 2009). It is possible, therefore, that *GLK* genes redundantly regulate photosynthetic development in all cell types in C_3 species, but have been co-opted into specialised roles in C_4 plants (at least in the maize lineage) (Rossini et al., 2001). To better understand the possible role for *GLK* gene duplication in C_4 evolution, and the generality of *GLK* function in C_4 development, the work presented in this chapter provides an extensive *GLK* gene phylogeny, tests cell specific expression of a *GLK* gene pair in an additional C_4 species and seeks to identify *cis*-regulatory regions that may explain cell-specific *GLK* expression.

6.2 Results

6.2.1 Two *GLK* gene duplicates exist in all sequenced C_4 species

To investigate the hypothesis that *GLK* gene duplication is an important precursor for the evolution of C_4 photosynthesis, *GLK* gene copy number was originally surveyed in *Sorghum bicolor*, which shares a common evolutionary origin of C_4 with maize (Christin et al., 2007). Two *GLK* duplicates were identified in sorghum by Southern blot analysis (Figure 6.1), which confirmed the sequence data in the first release of the sorghum genome (Paterson et al., 2009).

With an increase in the number of publically available plant genome sequences, the survey of *GLK* genes was extended to include all species with completed genome sequences available on the Phytozome resource (v8.0, www.phytozome.net) plus the potato genome sequence (www.potatogenomics.plantbiology.msu.edu). To determine the *GLK* gene phylogeny, annotated plant genomes were searched using *ZmGLK1* as a query sequence. *GLK* genes are distinguished from other members of the GARP superfamily (Riechmann et al., 2000) by the presence of a C terminal GCT-box and by an AREAEAA motif (consensus sequence) at the C terminal of the DNA-binding domain (Fitter et al., 2002). 57 GLK genes were identified (Appendix 6.1). To confirm that *GLK* genes were not overlooked during manual searching, an alignment of a subset of *GLK* genes was used as a template for an iterative HMMer search of the 31 genomes originally surveyed (Kelly et al., 2011). Phylogenetic analyses showed that 56 of the 57 identified *GLK* genes form a monophyletic clade that is a sister group to the pseudo-response regulator (PRR) group of GARP family genes (data not shown). The single *Selaginella moellendorffii GLK* gene clustered with the

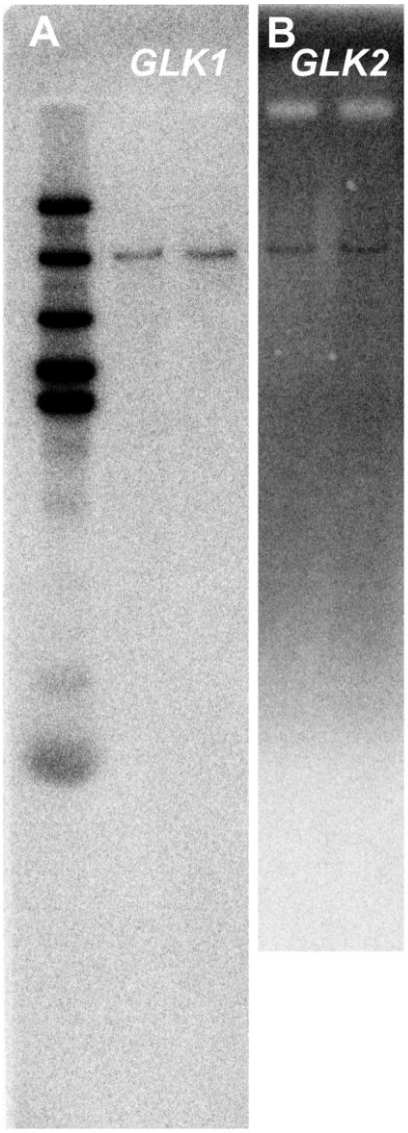


Figure 6.1. Southern blot analysis of sorghum *GLK* genes. A) Sorghum genomic DNA blot (*EcoRI* restriction digest) hybridised with a gene specific fragment of *SbGLK1*. **B)** Sorghum genomic DNA blot (*XbaI* restriction digest) hybridised with a gene specific fragment of *SbGLK2*. Similar results were seen with *HindIII* (*SbGLK1*) and *Sac1* (*SbGLK2*) digestions. Each blot shows two technical replicates.

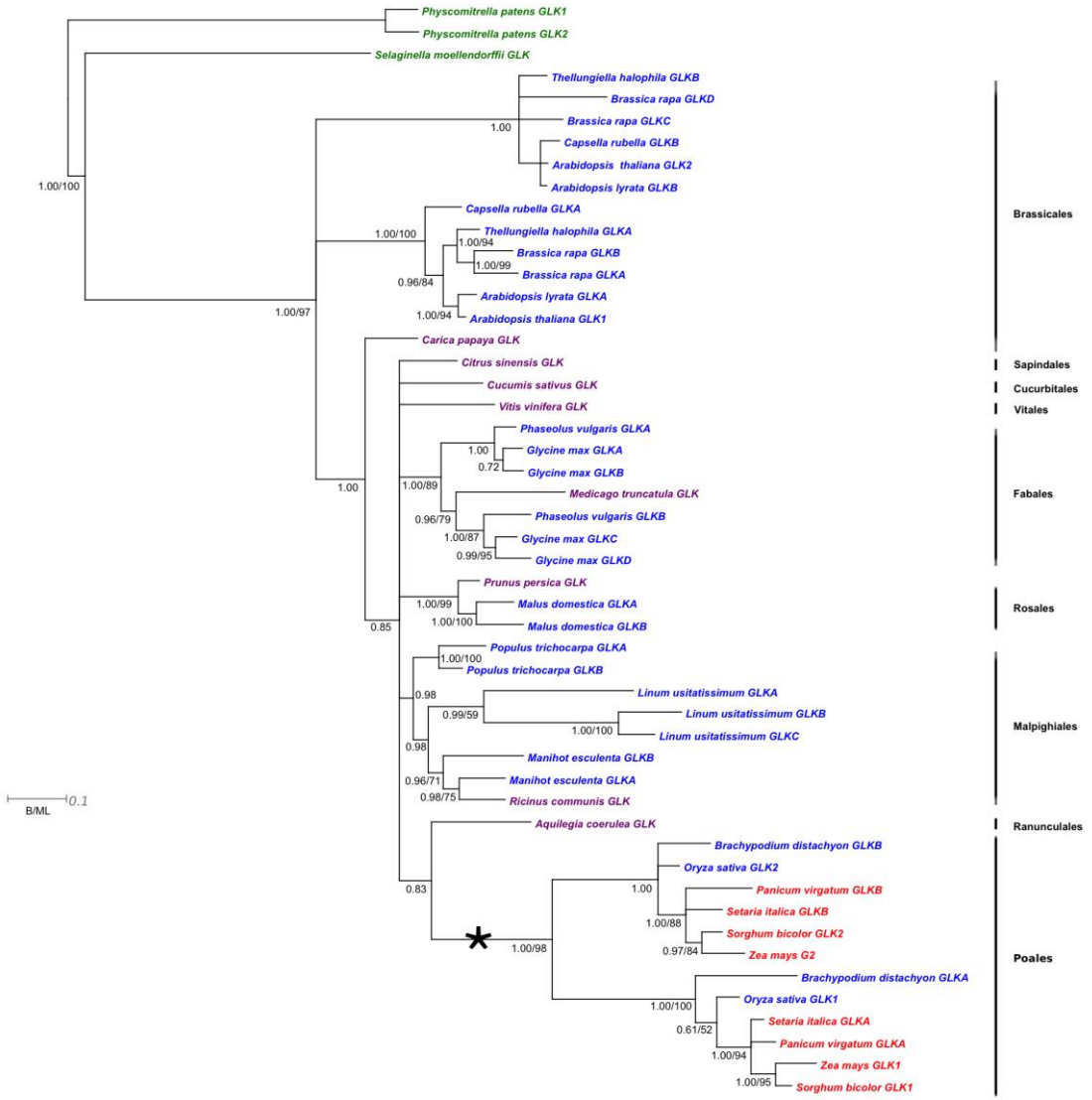


Figure 6.2. Bayesian phylogenetic tree of *GLK* genes. Posterior probabilities (B) and, where appropriate, bootstrap support values (ML), shown at branch nodes. Sequences highlighted in green are non-angiosperm, in purple are C_3 species with a single *GLK* gene, in blue are C_3 species with *GLK* duplicates and in red are C_4 species with *GLK* duplicates. Characterised *GLK* genes are annotated with numbers (e.g. *GLK1*), those that have not been previously described are annotated with letters (e.g. *GLKA*). *GLK* sequences correspond to the gene accession shown in Appendix 6.1. Asterix indicates gene duplication in the Poales.

PRR genes due to the additional presence of a pseudo-response regulator receiver domain in the *S. moellendorffii GLK* gene sequence. Crucially, no new *GLK* genes were identified. Phylogenetic trees of the 57 *GLK* gene sequences were generated using Bayesian and maximum likelihood methods. Preliminary phylogenetic analyses suggested long-branch attraction in the *Eucalyptus*, *Mimulus* and potato sequences and thus they were removed from subsequent analyses. The tree based on the remaining 50 *GLK* genes (Figure 6.2) demonstrates two key points. First, all four C_4 species in the dataset have two *GLK* genes (colored red). Second, some C_3 species have a single *GLK* gene (colored purple), whereas others have two or more *GLK* genes (colored blue). These data are consistent with the suggestion that the last common ancestor of flowering plants had a single *GLK* gene and that gene duplication occurred in specific lineages.

6.2.2 Cell-specific *GLK* gene expression occurs in both maize and sorghum

To determine whether cell specific expression of *GLK* genes is a general feature of C_4 biology rather than a maize-specific phenomenon, expression of the sorghum orthologues of *ZmGLK1* and *G2* (Fig 6.2) were tested by RNA blot analysis. To achieve this objective, sorghum BS and M cells were separated mechanically (BS) and enzymatically (M), in a process adapted from the protocols developed in maize (Sheen and Bogorad, 1985; Westhoff et al., 1991). Successful optimisation of this procedure for sorghum took considerable time, particularly the need to account for smaller sorghum cell

size during M filtration and altered osmotic potential. However, pure preparations of BS and M cells were eventually obtained, as confirmed by light microscopy (Figure 6.3A, B). Cell preparation purity was further confirmed by both western and northern blot analyses, which identified the presence of enzymes and transcripts known to be localised to specific cell types only in the appropriate cell type (i.e. Rubisco LSU and *RbcS* in BS; malate dehyrogenase (MDH) and *PEPC* in M (Sheen, 1999) (Fig 6.3C, D).

Examination of sorghum *GLK* accumulation by RNA blot analysis showed that sorghum exhibits a similar pattern of cell specific *GLK* expression to maize (Fig 6.3D): the *ZmGLK1* orthologue *SbGLK1* accumulates preferentially in M cells and the *G2* orthologue *SbGLK2* accumulates preferentially in BS cells.

6.2.3 C_4 grasses contain unique DNA binding motifs in regions upstream of *GLK* coding sequences

To determine whether cell specific *GLK* expression in maize and sorghum could be a consequence of unique *cis*-regulatory sequences not found in C_3 grass species, upstream regions of *ZmGLK1* and *G2* orthologues in maize, sorghum, *Brachypodium distachyon* and rice were surveyed for DNA sequence motifs present only in maize and sorghum. Due to the redundant expression patterns of *OsGLK* genes (Rossini et al., 2001), the inclusion of rice in this analysis is crucial. Sequences 2kb 5' upstream of the start codons of the four *GLK1* or *G2* orthologues were compiled and DNA motifs found only

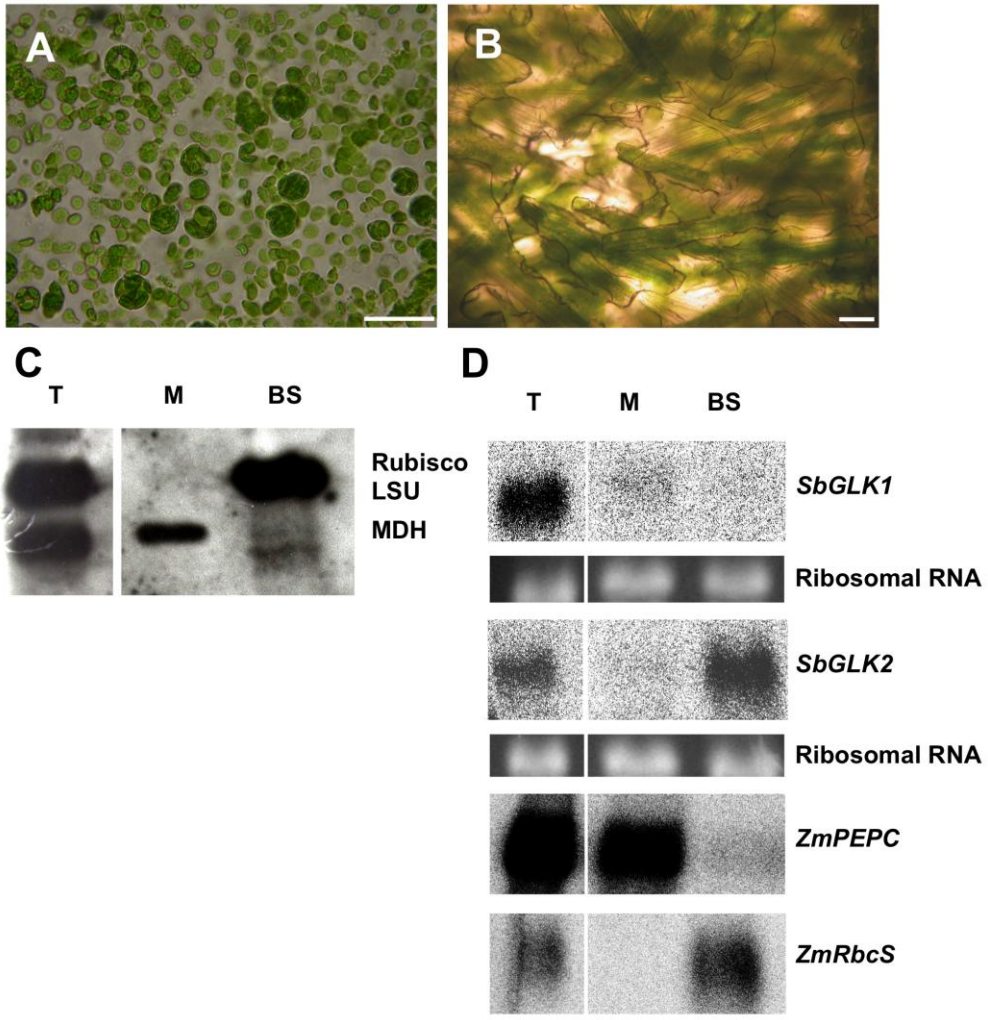
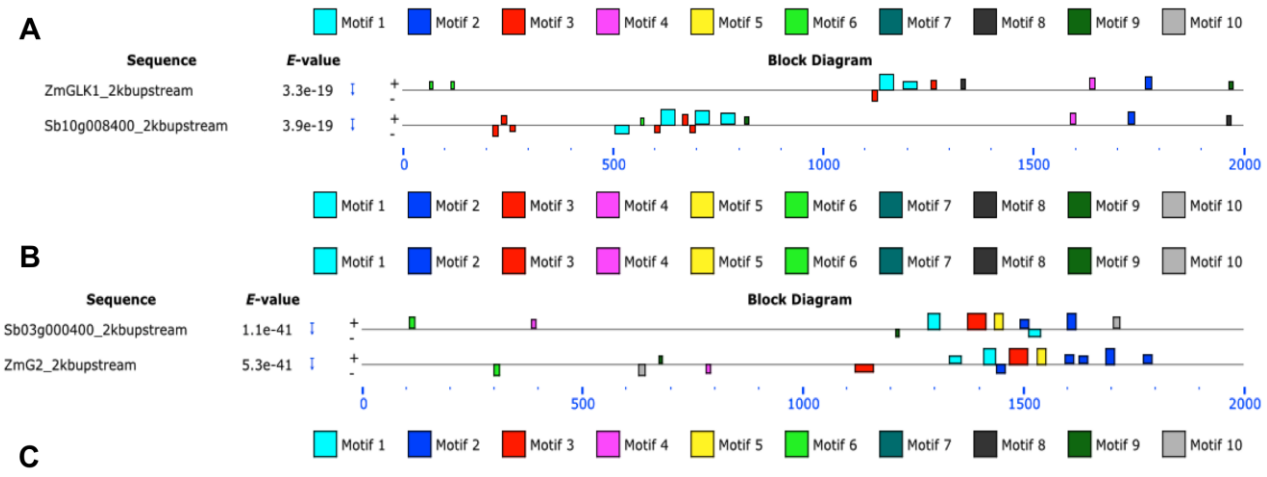


Figure 6.3. **Sorghum mesophyll and bundle sheath extracts reveal cell specific *GLK* expression. A, B)** Verification of cell extract purity by light microscopy of mesophyll (A) and bundle sheath (B) cell extracts. White scale bar = 40 µm**. C)** Western blot analysis of proteins purified from cell extracts shown in (A) and (B) T - total leaf, M – mesophyll, BS – bundle sheath. The blot was probed with antibodies raised against maize Rubisco LSU (BS marker) and MDH (M marker). **D)** Northern blot analysis of 10ug total RNA extracted from cells shown in (A) and (B). The gel was hybridised with ~300bp sequences of *SbGLK1* and *SbGLK2*, *ZmRbcS* (BS specific) and *ZmPEPC* (M specific) sequences were used to confirm cell extract purity. Ethidium bromide stained ribosomal RNA bands are shown as loading controls.

in the maize and sorghum sequences were identified using MEME suite software (Bailey et al., 2009) (options set to identify the 10 most significantly enriched motifs, minimum length of 6bp, maximum length of 50bp, 10). Figure 6.4 shows the distribution and number of non-overlapping repeats of the identified motifs for *ZmGLK1*/*SbGLK1* (Figure 6.4A) and *G2*/*SbGLK2* (Figure 6.4B) along the 2kb query regions. The MEME suite sub-programme MAST (Bailey et al., 2009) was subsequently used to filter redundant motifs. Non-redundant motifs, motif significance and repeat number are shown in Figure 6.4C. Consensus block diagrams for each non-redundant motif are shown in Appendix 6.2 for *ZmGLK1*/*SbGLK1* and Appendix 6.3 for *G2*/*SbGLK2.*

To clarify whether the motifs identified in sorghum and maize *GLK* upstream regions are a common feature in other C_4 grasses, similar regions were also investigated for *S. italica* and *Panicum virgatum GLK* orthologues (*ZmGLK1*: *SiGLKA/PvGLKA*; *G2*: *SiGLKB*/*PvGLKB* (Figure 6.2)). This was not possible for *PvGLKB*, however, in which small genome contig size meant that only 300bp of upstream sequence was available. Of the five motifs identified upstream of *ZmGLK1*/*SbGLK1* two are present 2kb upstream of *SiGLKA* and three upstream of *PvGLKA*. Five of the seven motifs identified upstream of *G2*/*SbGLK2* are present in the 2kb upstream of *SiGLKB*. No C_4 specific motifs were found in the 300bp upstream of *PvGLKB* (Figure 6.4C).

To determine whether any of the DNA sequence motifs identified as C_4 specific are binding sites for transcription factors, which could regulate the cell



	GLK1					*G2*						
Motif	1	2	4	6	8	1	2	3	4	5	6	10
Width (bp)	31	13	11	7	10	26	18	39	9	18	12	13
Maize p-value	7.32E-18	1.05E-08	1.92E-07	3.84E-05	8.22E-07	3.48E-16	3.80E-11	3.60E-23	1.20E-05	7.36E-12	2.04E-07	5.03E-08
Sorghum p-value	2.12E-18	1.05E-08	1.92E-07	8.21E-05	1.76E-06	1.42E-16	1.55E-11	1.47E-23	1.20E-05	7.36E-12	2.04E-07	1.10E-07
Maize repeat number	2	1	1	2	1	2	5	2	1	1	1	1
Sorghum repeat number	4	1	1	1	1	2	2	1	1	1	1	1
S. italica 2kb upstream	0	1	0	0	1	1	2	1	0	1	0	1
P. virgatum 2kb (*GLK1*) or 0.3kb (*G2*) upstream	0	1	2	0	1	0	0	0	0	0	0	0

Figure 6.4. DNA motifs identified in 5' regions of *GLK* genes in C_4 but not C_3 grasses. A, B) 10 most statistically significant motifs identified in 2kb regions 5' of the translational start of *SbGLK1*/*ZmGLK1* that are not present 2kb 5' of *OsGLK1*/*BdGLKA* (A) or 2kb regions 5' of the translational start site of both *SbGLK2*/*ZmG2* that are not present 2kb 5' of *OsGLK2*/*BdGLKB* (B). Coloured motif logos are ordered relative to motif appearance along the 2kb sequence (0 most 5', 2000 adjacent to the translational start site). Motif height represents statistical significance (tallest motifs are the most significant). Where motifs appear in the coloured block schematics but not on the DNA sequence representation this is due to the removal of overlapping motifs. **C)** Features of non-redundant motifs identified within (A) and (B). *P*-values refer to the statistical significance of the most significance motif appearance if the motif appears more than once. Motif repeat number for the same motifs identified in *S. italica* and *P. virgatum* refer to repeats where $p < 0.0001$.

-specific expression of *ZmGLK1*/*SbGLK1* and *G2*/*SbGLK2*, the DNA motifs were checked against the JASPAR-Plants database (Bryne et al., 2008) using the TOMTOM software package (Gupta et al., 2007). JASPAR-Plants computes interaction-likelihoods between DNA motifs and a collection of 21 non-redundant experimentally validated plant transcription factors. The analysis showed significant binding scores ($p < 0.05$) for many of the 21 JASPAR transcription factor profiles with the DNA motifs upstream of both *ZmGLK1/SbGLK1* and *G2/SbGLK2* identified by MEME (an annotated spreadsheet of all significantly ($p < 0.05$) interacting profiles can be found in Appendix 6.4). However, despite the statistical significance of these binding interactions, high *E*-values (expected false positive rate) and *q*-values (*p*-value adjusted for the false discovery rate) for the majority of these interactions make it likely that most are artefactual (Appendix 6.4). Among the transcription factor profiles with the lowest *E*-value scores, INDETERMINATE GROWTH 1 (ID1) (GRMZM2G011357) was suggested to bind motif 2 of the *ZmGLK1/SbGLK1* upstream region ($p = 0.026$, $E = 0.54$, $q = 0.537$) and DOF2 (GRMZM2G009406) ($p = 0.005$, $E = 0.10$, $q = 0.073$) was suggested to bind motif 1 of the upstream region of *G2*/*SbGLK2*. Schematics of the ID1 and DOF2 motif binding interactions are shown in Figure 6.5.

To enhance the analysis of putative *GLK* gene regulatory regions, an alternative method of DNA motif and TF binding prediction was employed. Transfac ® combines both motif and binding prediction (Wingender et al., 2000), however, it is limited by a lack of experimental validation of predicted

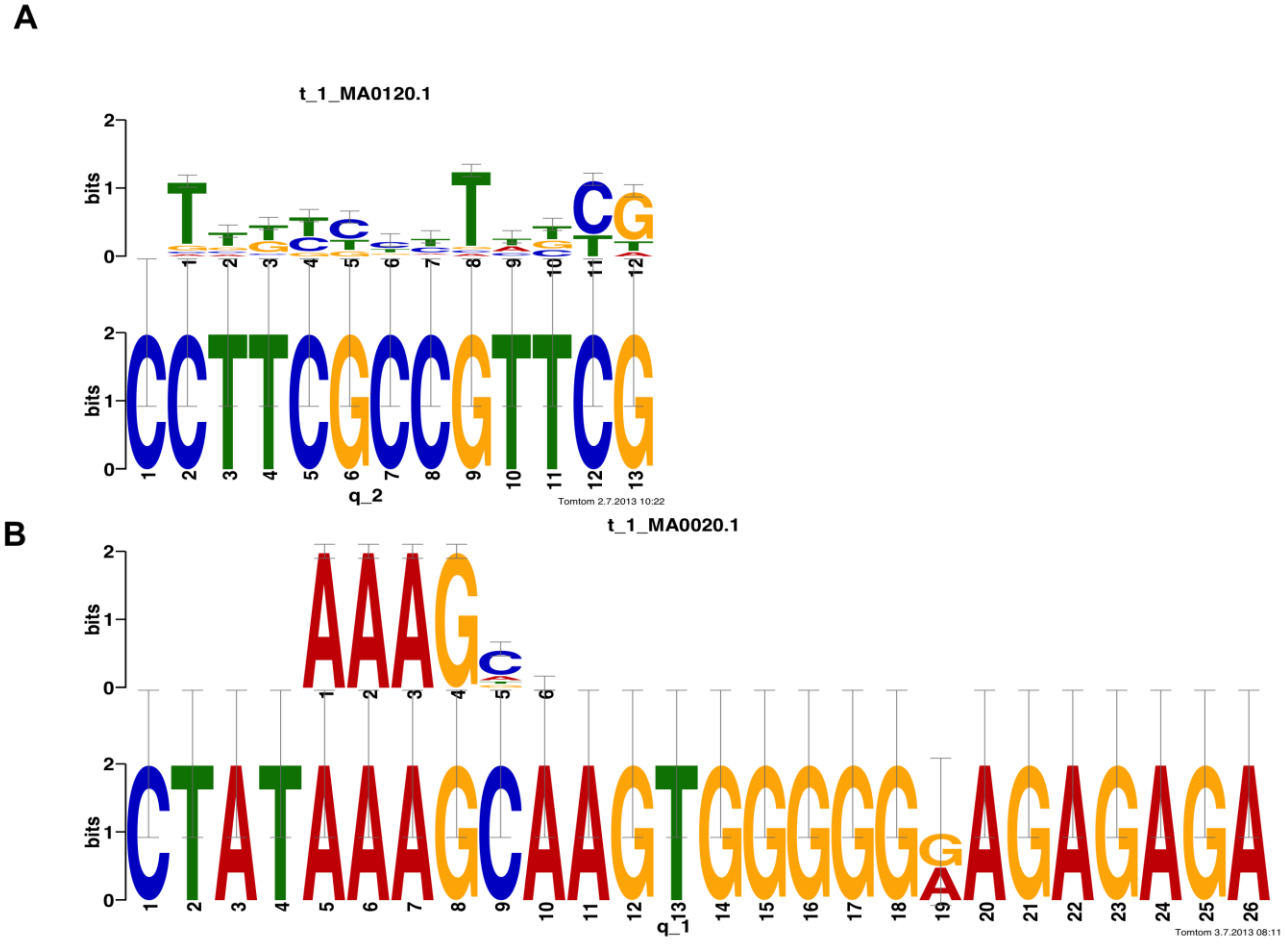


Figure 6.5. Putative *GLK* regulatory motif and transcription factor binding site interactions. Top rows show transcription factor consensus binding motif (JASPAR motif), bottom rows show *GLK* 5' DNA motif (MEME annotated). **A)** ID1 binding site overlaying *SbGLK1*/*ZmGLK1* motif 2. **B)** DOF2 binding site overlaying *SbGLK2*/*ZmG2* motif 1.

DNA/TF binding interactions and the redundancy of its TF profile database. The Transfac ® Match algorithm was used with the 'minimise false positive' option selected. The vast majority of transcription factor binding sites identified were generic, however, DOF2 sites were also identified 2kb upstream of *SbGLK2* and *G2* in addition to *SiGLKB* (x2) as well as *ZmGLK1*. No DOF2 sites were found 2kb upstream of the rice and *Brachypodium GLK* genes.

6.2.4 *GLK* gene duplicates exist in C_4 species outside the grasses

To date there is a paucity of completed C_4 genome sequences, with the only C_4 species to be fully sequenced existing in the grasses. This is reflected in the phylogenetic analysis presented in Figure 6.2. To extend the analysis of *GLK* duplication to C_4 species outside the grasses, sequences for the two *GLK* genes identified in the C_4 species *Cleome gynandra* were obtained from Dr. Julian Hibberd (Cambridge University). *C. gynandra* is one of the most closely related C_4 species to the model C_3 plant *Arabidopsis*, which has previously been shown to exhibit redundant patterns of *GLK* gene expression and of *GLK* gene function (Fitter et al., 2002; Waters et al., 2008). *C. gynandra* is thus positioned at an important phylogenetic juncture when considering the evolution of *GLK* gene function. Figure 6.6 shows that the *GLK* gene duplication at the base of the Brassicaceae demonstrated in Figure 6.2 predates the split between the *Cleome* and *Arabidopsis* lineages. This implies that the two *Cleome gynandra GLK* genes are orthologous to *AtGLK1* and *AtGLK2.*

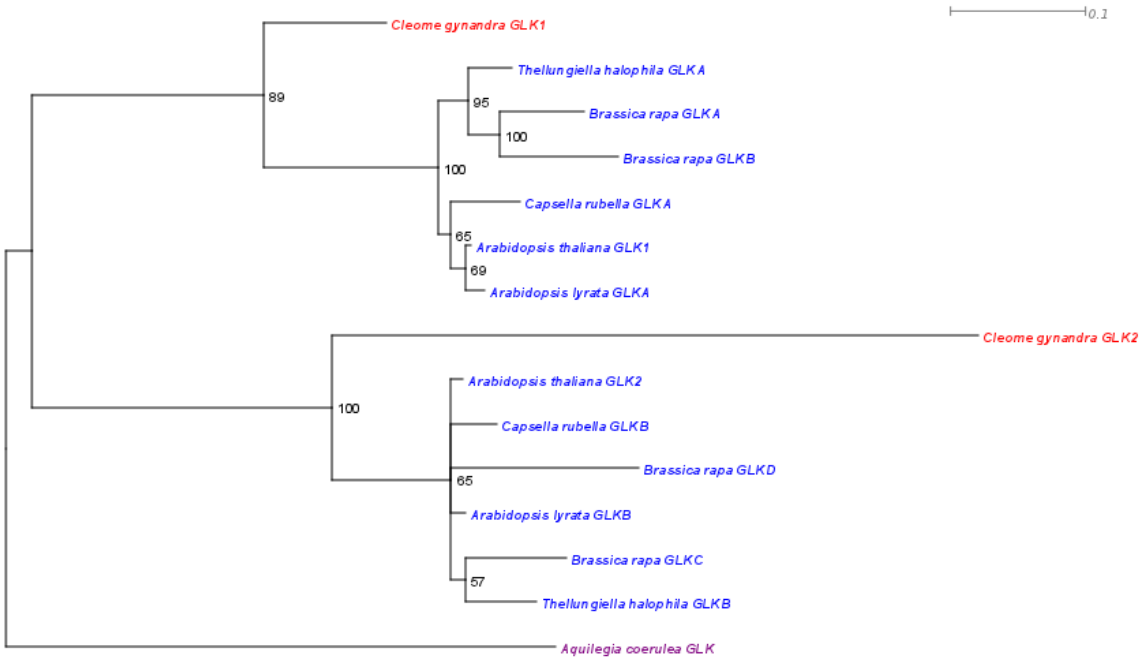


Figure 6.6. ***GLK* gene duplication in the Brassicales predates the split between *Cleome* and *Arabidopsis*.** Bootstrapped maximum likelihood phylogenetic tree of a subset of *GLK* genes from the Brassicales with the *Aquilegia* (Ranunculales) gene used as the outgroup. Genes are coloured as in Figure 6.2.

6.3 Discussion

6.3.1 *GLK* gene duplication may be an important precondition for the evolution of C_4 photosynthesis

The phylogenetic analyses presented here reveal a complex history of the *GLK* gene lineage across the land plants. During plant evolution, the GARP superfamily of transcription factors, which contains the *GLK* lineage (Riechmann et al., 2000), has expanded from 4 members in the extant algal species *Volvox carteri* and *Chlamydomonas reinhardtii* to 54 and 98 in flowering plants such as *Arabidopsis* and maize respectively (http://planttfdb.cbi.edu.cn/family.php?fam=G2-like), suggesting multiple rounds of GARP gene duplication during land plant evolution. One to four *GLK* gene copies are found in land plant genomes, however, none exist in sequenced algal genomes (data not shown). This suggests that the *GLK* lineage arose from an ancestral GARP sequence concomitant with, or prior to, the colonisation of land by plants.

Based on current preliminary data from basal land lineages, it is most likely that a single *GLK* copy was retained during early land plant evolution. Interrogation of draft genome sequences of the hornwort species *Anthoceros punctatus* (E. Frangedakis, S. Kelly, J. Fouracre and J.A. Langdale, unpublished data) and the liverwort *Marchantia polymorpha* (K. Ishizaki, Kyoto University, personal communication) reveal single *GLK* copies in each. Two *GLK* copies exist in the moss *Physcomitrella patens* (Yasumura et al., 2005),

however, this is a likely consequence of a recent genome duplication rather than a specific *GLK* gene duplication (Rensing et al., 2008) (Figure 6.2). The lycophyte *S. moellendorffii* also retains the proposed ancestral single gene state. Unfortunately, the lack of additional non-seed plant genomes precludes further speculation on the timing of *GLK* gene duplication prior to the emergence of the angiosperms.

The multiple genome-wide duplications (GWD) that have occurred within the angiosperms (Soltis et al., 2009) are reflected in the topologies of the *GLK* phylogenies shown in Figure 6.2 and Figure 6.6. The phylogenies exhibit complex patterns of gene duplication, but these can be rationalised as follows. Within the eudicot lineage, as orthologous relationships cannot be demonstrated between *GLK* genes of members of the Rosids and the Asterids, it is likely that all *GLK* gene duplications occurred following the ancient hexaploidisation event that preceded the divergence of these two groups (Jaillon et al., 2007). In the Rosales, the *GLK* gene duplication identified in *M. domestica* can be explained by a GWD in the Maleae tribe (Velasco et al., 2010). Within the Fabales, two GWD events occurred in the Fabaceae: one approximately 54 million years ago before the divergence of *Medicago* from soybean and common bean, followed by a subsequent event with soybean around 13 million years ago (Cannon et al., 2010; Schmutz et al., 2010). These patterns of genome duplication explain the presence of two *GLK* genes found in *P. vulgaris* and the four found in *G. max*. The presence of a single *GLK* gene in *M. truncatula* implies gene loss following the first Fabaceae GWD. In the Malpighiales, the two *P. trichocarpa GLK* genes are a

likely consequence of a Salicaceae specific GWD (Tuskan et al., 2006) and the three *GLK* genes identified in *L. usitatissimum* are likely products of within-species duplications. The two *GLK* duplicates in *M. esculenta* and single *GLK* gene in *R. communis* suggest a *GLK* duplication within the Euphorbiaceae followed by a loss in the *R. communis* lineage.

In the Brassicales, which includes the C_4 species *C. gynandra*, the presence of a single gene in *Carica papaya*, two genes in four of the five Brassicaeae genera sampled (*Arabidopsis, Capsella, Thelungiella* and *Cleome*) and four genes in the fifth, *Brassica,* suggest an initial GWD at the base of the Brassicaceae (following divergence from the Capparaceae) and a second in the *Brassica* lineage (Blanc et al., 2003). Thus *GLK* duplication occurred prior to the evolution of C_4 photosynthesis in the *Cleome* lineage (Figure 6.6). The GWD duplication at the base of the Poales (Soltis et al., 2009) is reflected in the topology of the grass *GLK* phylogeny (Figure 6.6). As the orthology of C_3 and C_4 grass *GLK* genes is so clearly resolved, it is likely that *GLK* gene duplication preceded the evolution of C_4 photosynthesis in the Poales as well.

There is a striking correlation between species that contain single *GLK* genes (*C. sinensis, C. sativus, V. vinifera, P. persic*a and *A. coerulea*), and orders in which C_4 photosynthesis has not evolved (Sapindales, Curcurbitales, Vitales, Rosales and Ranunculales respectively) (Sage et al., 2011a). This is despite the independent evolution of C_4 photosynthesis on over 60 occasions across 9 different orders of flowering plants (of approximately 60 orders in total). A notable exception is in the Malpighiales, which contains a C_4 *Euphorbia*

lineage and where *R. communis* only has a single *GLK* copy. However, as discussed above this is a likely consequence of gene loss. Taken together, it is thus tempting to speculate that *GLK* duplication was a pre-requisite for the evolution of C_4. Genome duplication has long been considered as a possible vital pre-condition for the evolution of C_4 (Monson, 2003; Williams et al., 2012). Indeed, the *GLK* phylogenies and rationalisation of *GLK* duplication patterns presented here may be evidence for the importance of GWD in preconditioning C_4 evolution, rather than for the importance of *GLK* gene duplication *per se*. More genome sampling, particularly of C_4 species, will be required before it is possible to fully determine the significance of *GLK* gene duplication in facilitating C_4 evolution.

6.3.2 Cell-specific expression of *GLK* genes is not a maize specific phenomenon

The premise for the importance of *GLK* gene duplication in the evolution of C_4 is the *GLK* cell-specific expression pattern and BS-specific *g2* loss of function phenotype observed in maize. The finding that *GLK* genes are also expressed in a similar manner in sorghum (Figure 6.3D) supports this hypothesis as it shows that compartmentalisation of *GLK* expression is not a maize-specific phenomenon. Furthermore, RNA-seq of mechanically isolated BS and M cells from *S. viridis* has also revealed differential expression of *Setaria GLK* genes (C. John, S. Kelly, J. Hibberd unpublished data). As in maize and sorghum, the *Setaria* orthologue of *G2* accumulates preferentially in BS cells and the *Setaria* orthologue of *ZmGLK1* accumulates in M cells. This is despite the

independent evolution of C_4 in the *Setaria* and Andropogoneae lineages (Christin et al., 2009a).

The procedure for the isolation of M and BS cells used in this study resulted in highly pure cell extracts (Figure 6.3A, 6.3B, 6.3C, 6.3D) that enabled visualisation of gene expression by northern blot analysis. The compartmentalisation of *SbGLK* expression shown here was subsequently confirmed by a study that employed RNA-seq of sorghum BS and M cell extracts purified in a similar manner (Wang et al., 2013b). Despite being used successfully for 30 years, the methods for manual isolation of BS and M cells as used in this study and that of Wang et al (2013b) have been somewhat superseded by microdissection techniques such as LCM. While cell separation by LCM may be more precise, the technique has not yielded results that conflict with results produced by 'cruder' methods, and have confirmed cell-specific expression of *GLK* genes in maize (Li et al., 2010). As such, and with the continued refinement of manual isolation methods (Covshoff et al., 2013), enzymatic and mechanical isolation of BS and M remains a viable technique, particularly in the grasses.

Utilisation of LCM in the C_4 eudicot *C. gynandra* has revealed that *GLK* gene expression is not compartmentalised in all C_4 species (Wang et al., 2013b). *CgGLK1* and *CgGLK2* are both expressed in M and BS cells, although both accumulate to a significantly higher level in M cells, and *CgGLK1* is expressed tenfold higher than *CgGLK2* in both cell types (Wang et al., 2013b). These data suggest that cell-specific *GLK* expression is not necessary for C_4

chloroplast development in all C_4 species. It is possible that posttranscriptional regulation may lead to compartmentalised *GLK* function in *Cleome*, however, the disparity between the expression patterns of *GLK* genes in Andropogoneae/*Setaria* and *Cleome* may be associated with the different C_4 subtypes that exist in the two groups. Maize, sorghum and *Setaria* are NADP-ME subtype C_4 species (Christin et al., 2009a), which exhibit dimorphic chloroplasts with reduced granal stacks and PSII in the BS, while *Cleome gynandra* is a NAD-ME subtype which displays granal stacks in both cell types (Marshall et al., 2007). In *Arabidopsis*, *GLK* genes act redundantly to regulate a cohort of genes involved in chlorophyll biosynthesis and light harvesting (Waters et al., 2008; Waters et al., 2009). Consequently, *GLK* activity affects thylakoid aggregation and photosystem assembly. Thus it is plausible that compartmentalisation of *GLK* gene function in the PACMAD clade of grasses enabled dimorphic chloroplast formation, whereas lack of *GLK* compartmentalisation in *Cleome* has led to a different C_4 trajectory. Additional expression analyses in a wider variety of C_4 species will provide further insight into whether an association between *GLK* compartmentalisation and dimorphic chloroplast development is a more general phenomenon. Such chloroplast dimorphism is found in both monocot and eudicot C_4 species (Laetsch, 1974). Given that NADP-ME type C_4 photosynthesis has been identified as the most feasible C_4 type for engineering into C_3 plants (Weber and von Caemmerer, 2010; Kajala et al., 2011), if cell-specific expression of *GLK* genes is found to be associated with dimorphic chloroplast development in multiple grass lineages, compartmentalisation of *GLK* expression will need to be given careful consideration in the context of C_4 rice development.

Figure 6.2 shows the maintenance of *GLK* duplicates in a wide variety of species that have recently undergone GWD. Following GWD, the usual evolutionary trajectory is that of diploidisation and gene loss (Wolfe, 2001). Thus the maintenance of so many *GLK* gene duplicates is intriguing, especially given the apparent redundancy of expression and function across photosynthetic cell types identified in a wide range of species (*Arabidopsis* (Fitter et al., 2002), moss (Yasumura et al., 2005), rice and *Cleome* (Wang et al., 2013b)). Given that *GLK* genes have been proposed to function in the coordination of the light and dark reactions of photosynthesis (Waters and Langdale, 2009), sub-functionalisation of *GLK* activity in C_3 species would not be surprising (Wang et al., 2013b). Indeed a degree of *GLK* specialisation has been identified in a number of recent studies in *Arabidopsis* that show differential responses of *GLK* genes to nitrogen (Gutierez et al., 2008), altered plastid import pathways (Kakizaki et al., 2009) and cytokinin (Kobayashi et al., 2012). Developmental specialisation has also been shown to occur in both *Arabidopsis* and tomato, in which only a single *GLK* copy is expressed during fruit development (Fitter et al., 2002; Powell et al., 2012). Thus, specialisation of *GLK* gene function in response to environmental and developmental signals is important in both C_3 and C_4 species. The *g2* loss of function phenotype in maize, which also shows perturbed development in non-C4 cell types (Langdale and Kidner, 1994), may suggest that such specialisation evolves from a default, redundant, state.

6.3.3 Promoter analyses may provide insights into the cell-specific accumulation of *GLK* genes in C_4 grasses

Despite the orthology of maize, sorghum and rice *GLK* genes (Figure 6.2), cell-specific *GLK* accumulation does not occur in rice (Wang et al., 2013b). It is likely, therefore, that altered transcriptional regulation has evolved since the divergence of the BEP and PACMAD clades. While the complexity of genetic regulation at the transcriptional level continues to be revealed, 5' regions adjacent to coding sequences remain centrally important to *cis*-regulation of transcription (Alberts et al., 2008). Using DNA motif identification software, a number of motifs were identified that are present in 5' regions of sorghum and maize *GLK* genes that are absent in rice and *Brachypodium* orthologues (Figure 6.4A, 6.4B). Several of these motifs exhibit conserved syntenic arrangements between sorghum and maize, particularly towards the translational start sites of *ZmGLK1*/*SbGLK1* and *G2*/*SbGLK2* (Figure 6.4A, 6.4B). The maintenance of motif synteny, despite approximately 12 million years of divergence between the maize and sorghum lineages (Swigonova et al., 2004), suggests that the chromosomal arrangement of motifs may be an important component of *GLK* transcriptional regulation. Of the motifs identified in maize and sorghum but not rice and *Brachypodium*, a subset were also identified upstream of the *S. italica* and *P. virgatum* orthologues (Figure 6.4C). *S. italica* and *P. virgatum* orthologues were not included in the original discriminatory analysis due to a lack of *GLK* expression data in these species and the fact that *P. virgatum* is a PEPCK subtype C_4 species (Christin et al., 2009a). This preliminary analysis suggests that there may be unique *cis*-

regulatory motifs upstream of C_4 *GLK1* and *G2* orthologues in the grasses. However, this analysis is limited by the low sample size (particularly in the use of only two C_3 control sequences) and the phylogenetic sampling (conserved motifs between sorghum and maize may be phylogenetic artefacts rather than novel regulatory elements). As additional genome sequences become available, further sampling of a wider range of C_3 and C_4 grass promoter regions will provide greater resolution. If C_4 *GLK* *cis*-regulatory regions do exist, comparisons between NADP-ME and other C_4 subtypes will provide insight into the molecular evolution and function of *GLK* compartmentalisation.

Of the DNA motifs in *GLK* gene promoters identified as potential transcription factor binding sites, few have strong enough statistical support to merit further discussion (Appendix 6.4). Based on the TOMTOM interrogation of the JASPAR TF binding profile database, there is limited support for ID1 binding to *ZmGLK1*/*SbGLK1* motif 2 (p = 0.026, E = 0.54, q = 0.537) (Figure 6.4A, Appendices 6.2, 6.4). Given the proposed roles of the *ID1*-like genes *ZmJAY1* and *ZmRAVEN1* in specifying BS and M cell type in the model of Kranz evolution presented in Chapter 7, the upregulation of *ID1* in the FP5 and FI samples (which exhibit the highest proplastid-plastid transition rates), and the observation that BS specific C_4 genes *NADP-ME*, *RbcS* and *Rubisco activase* are upregulated in an *id1* mutant (Coneva et al., 2007), it is tempting to speculate that *ID1*-like genes may also regulate cell-specific chloroplast differentiation as well. The observations that the putative ID1 binding motif is also present upstream of *SiGLKA* (Figure 6.4C), and that the *Setaria ID1* orthologoue is significantly upregulated in *Setaria* BS cells (C John, S. Kelly,

J. Hibberd unpublished data) are consistent with this scenario. However, analysis of the consensus DNA motif and TF binding profile sequences (Figure 6.5A) suggests only a weak interaction.

There is strong support for an interaction between DOF2 and motif 1 of *G2*/*SbGLK2* (two repeats upstream of both coding sequences) (p = 0.005, E = 0.10, q = 0.073, Figure 6.4B). This motif was also identified upstream of *SiGLKB* (Figure 6.4C) and a TF binding analysis using Transfac ® software identified DOF2 sites 2kb upstream of *SbGLK2, G2* and *ZmGLK1*. *DOF2* is closely related to the *ZmDLK1* and *ZmDLK2* genes identified in Chapter 7 as potential M and BS cell specifying factors. Further, *DOF2* has been shown to bind the promoters of the maize C_4 *PEPC* and the cystolic *PPDK* isoforms (Yanagisawa and Sheen, 1998; Yanagisawa, 2002). Taken together, it is plausible to suggest that shared DOF2 binding elements in *cis*-regulatory regions of *GLK* and C_4 pathway genes may lead to the cell-specific and concomitant initiation of chloroplast maturation and the C_4 pathway. The presence of DOF2 binding sites in both *G2* and *ZmGLK1* upstream regions can be explained by the proposed combinatorial activity of *DOF2*, which is suggested to act as both a transcriptional activator and repressor in different contexts (Yanagisawa, 2002). However, despite the statistical support for the interaction between DOF2 and the motif identified here, the interaction consists of only 5 motif bases. Again, further support for putative DNA motif-TF binding site interactions would come from a wider sample size, however, the predictions made here could be directly tested.

Transcriptional regulation is known to play an important role in cell-specific expression of many of the C_4 pathway genes. For example *cis*-elements have been identified that are critical to compartmentalisation of *Flaveria PEPC* and of maize *PPDK* and *PEPC* expression (reviewed in Hibberd and Covshoff, 2010; Williams et al., 2012). It is therefore reasonable to hypothesise that C_4-specific *cis*-elements exist in the promoter regions of *GLK* genes that exhibit compartmentalised expression. However, regulation of genes that operate in the C_4 pathway operates at many levels. For example, posttranscriptional regulation of cell-specific accumulation has been reported that is dependent on sequences in the UTRs or coding sequences of a number of genes (e.g. Patel et al., 2004; Patel et al., 2006; Brown et al., 2011; Kajala et al., 2012) and posttranslational regulation of C_4 enzyme activity leading to cell specific function has long been known (e.g. Meierhoff and Westhoff, 1993; Nimmo, 2003). In addition, different sets of *trans*-factors may be functional in BS and M cells (Brown et al., 2011; Kajala et al., 2012). Cell-specific accumulation of *GLK* genes therefore may not be dependent on *cis*-elements (or on *cis*-elements located within 2kb 5' of the start codon) but on a wider variety of regulatory factors. As such, the preliminary analysis of putative *GLK* regulatory regions presented here is by no means exhaustive, but may provide some insight into the compartmentalisation of *GLK* gene expression.

Chapter 7: General discussion

7.1 Overview

Kranz was identified as a distinct form of leaf anatomy over a century ago and its importance in facilitating the most productive forms of photosynthesis has been recognised for nearly 50 years. Further, Kranz is now known to have evolved on over 60 occasions independently within the angiosperms. Despite the significance of Kranz in angiosperm evolution and ecology, our understanding of the genetic regulators of Kranz development has remained highly limited. The emergence of systems biology, and in particular next-generation DNA sequencing (NGS) technologies, has enabled previously unparalleled levels of insight into the genetic control of developmental processes in a wide variety of study systems. The work in this thesis aimed to provide insights into the genetic regulation of Kranz anatomy through the analysis of a next-generation RNA-seq dataset of maize leaf development. Following initial validation of the sequencing data, this study has identified candidate Kranz regulators based on patterns of gene expression across a maturation series of two different types of maize leaf. Three of these candidates were functionally characterised in rice and *Setaria*. Finally, analyses of gene expression and phylogenetic diversification were extended for a small transcription factor family already known to play a role in Kranz patterning in maize. Considering the rapid progress in Kranz candidate identification that systems biology approaches have led to, this discussion will

focus on two aspects of systems biology. Firstly, how future systems biology driven analyses could enhance understanding of Kranz development. Secondly, available NGS data will be combined to propose a model for the regulation of Kranz development. As this study was initiated in the context of a collaborative global effort to produce a C_4 rice variety, this discussion will conclude with a summary of the contribution this work has made towards that goal.

7.2 Defining the Kranz-ome using additional 'omics' studies

The data presented in this study extends previous work that identified transcriptional components of both C_4 regulatory networks and early maize leaf development using NGS. These previous studies have described genome-wide transcriptome changes in C_4 versus C_3 species (Brautigam et al., 2011; Gowik et al., 2011), along the developmental gradient of a single maize leaf (Li et al., 2010; Pick et al., 2011), between BS and M cells (Li et al., 2010; Chang et al., 2012), between maize shoot and axillary meristems (Takacs et al., 2012) and following maize seed imbibition (Liu et al., 2013). Despite the insights these studies provided into C_4-type leaf development, the developmental stages used for analysis were either too old, too young or too diluted to provide extensive information on potential Kranz regulators. Through comparison of genome-wide expression changes that occur during key stages of primordia growth, this study has presented a detailed outline of a Kranz associated transcriptome. By doing so, it has allowed the

identification of 283 putative positive candidate Kranz regulators using stringent filters based on changes in gene expression between different stages of maize leaf development.

As more data become available, however, this candidate gene list is liable to change in size, although it is unclear in which direction, as the results of current and potential sequencing projects may extend or refine this list. A key comparative study that would likely reduce the number of candidate genes is a transcriptome analysis of equivalent rice leaf primordia. Rice does not develop husk leaves, but comparison of gene expression within the FP, FP3/4 and FP5 samples would be likely to identify conserved components of leaf regulatory networks between the two species, particularly epidermal regulators, chloroplast biogenesis genes, blade specification factors and regulators of vascular element differentiation. As such, the appearance of orthologues in both sets of primordia (to similar levels of expression) would merit deprioritisation as a candidate in maize. Work has already begun on the comparison of an equivalent rice leaf developmental gradient to that described in maize (Li et al., 2010, T. Brutnell, personal communication). The results of this study are also likely to reduce the candidate list if orthologous genes are found to be expressed at an equivalent level or higher at the base of a rice leaf compared to the maize leaf base. It is less clear how the forthcoming results of the 1000 plants (1KP) project, that is aiming to sequence 1000 plant transcriptomes including those of a number of closely related C_3 and C_4 species, will affect the maize candidate gene list identified here (www.onekp.com). As the samples used for the 1KP transcriptomic

analysis were fully differentiated, observations of changes in gene expression between C_3 and C_4 species are unlikely to provide insights into Kranz regulation. However, information that the transcriptomic data has already provided has revealed a surprising degree of parallel sequence evolution, enriched in C_4 species, for a large number of genes (S. Kelly, personal communication). It has previously been demonstrated that certain amino acid substitutions have evolved independently in a small number of C_4 pathway enzymes in C_4 species (e.g. Christin et al., 2007; Christin et al., 2009b). The identification of C_4-type molecular signatures in a far greater range of genes may lead to prioritisation of candidate genes already identified, or to functional characterisation for genes that have narrowly failed the stringent candidate filtration process. An area of transcript accumulation that is yet to be investigated is that of the contribution of small noncoding RNA, such as microRNAs (miRNA), to Kranz regulation. miRNAs are known to be crucial components of regulatory networks that, for example, control processes as diverse as vegetative phase change (Wu and Poethig, 2006) and root vascular differentiation (Carlsbecker et al., 2010). However, miRNAs were not identified as part of the RNA-seq dataset presented here. While informative, investigation into miRNA expression would require additional dissection of further leaf primordia samples. A technically simpler and more urgent test of gene expression in the primordia samples would be to analyse candidate transcript accumulation by *in situ* hybridisation. This will both further validate the RNA-seq data, and provide insight into the accumulation patterns of candidate transcripts within individual primordium samples. If candidate

transcripts were found to localise to epidermal cells types, or be distributed ubiquitously for example, this would merit candidate deprioritisation.

In addition to the benefits that further targeted transcriptomic studies might provide, alternative systems approaches that test distinct components of cellular function, such as epigenetics, proteomics and metabolomics, may also prove insightful. As a common code of histone modification has been found in the promoter regions of C_4 pathway genes in independent C_4 lineages (Heimann et al., 2013), epigenomic analyses of different young Kranz tissues from a variety of C_4 species could provide an additional layer of information about possible conserved elements of genetic regulation of Kranz development. The potential significance of transcriptomic and epigenomic analyses assumes that Kranz development is transcriptionally regulated. Although systems based evidence for transcriptional regulation of C_4 enzymes (Li et al., 2010; Brautigam et al., 2011; Gowik et al., 2011; Pick et al., 2011) supports a model of Kranz regulation at the level of gene expression, it is possible that Kranz regulators are not differentially expressed between Kranz and non-Kranz tissue types and that Kranz development is a consequence of post-transcriptional regulatory mechanisms. Additional proteomic analyses of maize leaf primordia would thus be required. Such analyses have already revealed maize BS and M chloroplast-specific protein content (Majeran et al., 2005; Majeran et al., 2008; Friso et al., 2010) and proteomic changes along the developmental gradient of a maize leaf (Majeran et al., 2010). Although proteomic comparisons of maize leaf primordia would require extensive downstream processing relative to the equivalent transcriptomic comparisons

described here. It is also possible that Kranz is not largely genetically regulated, and is instead a consequence of unidentified metabolic or hormonal signals, in which case metabolomic comparisons of leaf primordia would be worthwhile. For example, if the proposed Kranz inducing signal (Langdale and Nelson, 1991) was not protein based, then metabolomes of isolated cell types may prove informative. However, the rigid Kranz developmental programme that exists in foliar leaves regardless of environmental or nutrient conditions suggests that Kranz development is not likely to be a consequence of large-scale changes in the metabolic content of the leaf.

7.3 Short roots and leaf wreaths: a radial patterning model for Kranz development

Based on the transcriptome data presented here and in a study of separated maize BS and M cells (Li et al., 2010), the genes shown in Table 7.1 are proposed to function as key components of the regulation of the three stages of Kranz patterning in maize 1) procambium initiation, 2) BS and M cell-specification and, to a lesser extent, 3) cell-specific accumulation of C_4 pathway enzymes.

Accession	Gene family	Gene subfamily	Gene name	Foliar profile	Husk profile	Kranz candidate list	BS expression [a]	M expression [a]	Ratio Foliar P3/4: Husk P3/4
GRMZM2G045883	bHLH	SPCH clade	*ZmSPCH-like 1 (ZmSPL1)*	FA2	HN	Yes	0.0	0.5	3:1
GRMZM2G132794	GRAS	SHR clade	*ZmSHR1*	FA2	HN	Yes	11.9	0.0	4:1
GRMZM2G172657	GRAS	SHR clade	*ZmSHR2*	FA2	HN	Yes	0.0	0.5	3:1
GRMZM2G028046	C2H2	MRPI clade	*ZmMRPI-like1 (ZmMPL1)*	FA2	HN	Yes	0.0	1.0	3:1
GRMZM2G136494	C2H2	MRPI clade	*ZmMPL2*	FA2	HN	Yes	0.0	0.0	4:1
GRMZM2G150011	C2H2	DOT5 clade	*ZmDOT1*	FA2	HN	Yes	0.0	0.0	10:1
GRMZM2G039074	MYB	KAN clade	*ZmATS-like 1 (ZmATL1)*	FA2	HN	Yes	0.0	0.0	No HP3/4
GRMZM2G114998	C2C2-Dof	DAG1 clade	*ZmDAG-like 1 (ZmDLK1)*	FA2	HN	No	224.3	22.5	2:1
GRMZM2G171852	C2C2-Dof	DAG1 clade	*ZmDLK2*	FA2	HN	No	279.4	19.1	2:1
GRMZM2G129261	C2H2	ID1 clade	*ZmJAY1*	FA2	HN	No	70.5	402.4	2:1
GRMZM2G143723	C2H2	ID1 clade	*ZmRAVEN1 (ZmRVN1)*	FA2	HN	No	19.0	5.0	5:1
GRMZM2G131516	GRAS	SCR	*ZmSCR1*	FN	HN	Yes	142.2	78.2	2:1
GRMZM2G040278	F-box	SNEEZY-like	not named	FN	HN	Yes	0.0	0.5	12:1
GRMZM2G440543	F-box	SNEEZY-like	not named	FN	HN	Yes	0.0	2.0	5:1
GRMZM2G163724	Kinase	RLK	not named	FN	HN	Yes	0.0	0.0	6:1
GRMZM2G114893	C2H2	-	not named	FN	HN	Yes	52.0	87.5	4:1

Table 7.1. Genes proposed to function in a *SHR* based model of Kranz development. The top tier of genes are present in the FA2 profile and the Kranz candidate list, the middle tier of genes are present in the FA2 profile, and the bottom tier of genes are orthologues of known SHR targets present in the Kranz candidate list. Genes highlighted in green are enriched in BS cells, genes highlighted in blue are enriched in M cells. [a] raw data from Li et al (2010) remapped using RSEM (Li and Dewey, 2011), values in RPKM.

7.3.1. Initiation of procambium

The seven genes shown in the top tier of Table 7.1 were identified in both the FA2 profile and the positive candidate filter list. The FA2 profile was designed to select for genes upregulated during procambial specification, and the list of putative positive candidate regulators was a product of filters designed to

select for upregulated genes in young undifferentiated foliar tissue, compared to older foliar tissue and husk tissue of a range of ages. The presence of these genes in both lists strongly suggests that they have roles in the initiation of Kranz-type close vein spacing observed in maize foliar leaf blades. Further support for the function of these genes in vascular patterning comes from functional data on *Arabidopsis* orthologues. The bHLH gene is closely related to *SPEECHLESS* (*SPCH*) and is annotated here as *ZmSPCH-like 1* (*ZmSPL1*). *SPCH* is known to regulate the spacing of stomata in the epidermis (Lampard et al., 2008) and it is possible that it has been co-opted into a procambial spacing network in the maize leaf. A maize orthologue of *DEFECTIVELY ORGANISED TRIBUTARIES 5* (*DOT5*) is also present in this list, annotated here as *ZmDOT1*. In *Arabidopsis*, *dot5* loss of function mutants display perturbed vascular patterning in cotyledons (Petricka et al., 2008). The *KANADI* gene in Table 7.1 is closely related to the *Arabidopsis* gene *ABERRANT TESTA SHAPE* (*ATS*)*/KANADI4* (Eshed et al., 2001; McAbee et al., 2006) and is annotated as *ZmATS-like 1* (*ZmATL1*). In *Arabidopsis* ATS/*KAN4* contributes to PIN1 localisation (Izhaki and Bowman, 2007), which is necessary for the polarised flow of auxin. As auxin flow precedes vascular differentiation (Scarpella et al., 2010) the presence of *ZmATL1* is intriguing.

The presence of two orthologues of *SHR* (annotated as *ZmSHR1* and *ZmSHR2*) is consistent with a proposed role for the *SHR*/*SCR* regulatory network in regulating Kranz development (Slewinski et al., 2012; Slewinski, 2013; Wang et al., 2013a) and the observation that *SHR* expression domains

precede vascular initiation in *Arabidopsis* (Gardiner et al., 2011). Two further genes identified in this cohort are closely related to maize *MYB-related protein 1-interactor 1* (*ZmMRPI1*), which was isolated on the basis of its interaction with the endosperm transfer cell localised protein ZmMRP (Royo et al., 2009). These two genes have been annotated *ZmMRPI1-like 1* and *2* (*ZmMPL1* and *ZmMPL2*). Support for *ZmMPL1* and *ZmMPL2* in vascular patterning is less clear, however, the presence of a protein-interacting domain in both proteins that is shared with targets of *SHR* in *Arabidopsis* (Levesque et al., 2006; Cui et al., 2011) suggest they may function downstream of *SHR* in vascular patterning. In addition to these seven genes, 40 other genes were identified by both the FA2 profile and the positive Kranz candidate filters (Appendix 7.1). Given the role of auxin in vascular patterning (Scarpella et al., 2010), the presence of an orthologue of an *Arabidopsis* auxin import carrier (*AUX1*) (Peret et al., 2012) is also particularly noteworthy.

Except for the described increase in intervein spacing, husk leaf veins are indistinguishable from foliar veins both during early development (Figure 3.1) and in fully differentiated leaves (Pengelly et al., 2011). As such it is expected that putative regulators of procambium initiation are expressed in both foliar and husk primordia. *ZmSHR1*, *ZmSHR2*, *ZmSPL1*, *ZmDOT1*, *ZmMPL1* and *ZmMPL2* are all expressed in both foliar and husk primordia, but at much lower levels in husk primordia. This suggests that increased vein density in foliar leaves is a consequence of a relative increase in transcript accumulation in foliar leaves compared to husk leaves, rather than the presence or absence

of transcripts for the majority of vascular regulators in foliar and husk leaves respectively. This is consistent with the view that the repeated evolution of C_4 has been facilitated by the co-option of networks that already exist in C_3 plants (Hibberd and Quick, 2002; Sage, 2004), and the observation that vascular formation occurs at a higher rate in C_4 species compared to C_3 (Mckown and Dengler, 2009). An exception is seen for *ZmATL1*, which is hardly expressed in the HP34 or HP5 samples (Appendix 4.31). *ATS/KAN4* is known to regulate *PIN* expression in *Arabidopsis* (Izhaki and Bowman, 2007). As PIN protein localisation determines auxin flow, which precedes vascular initiation (Scarpella et al., 2010), the expression of the *ZmATL1* specifically in foliar primordia samples may be a key factor in the initiation of close vein spacing in foliar leaves. *KAN* genes are also known to regulate radial patterning in stems (Emery et al., 2003), it is therefore possible that *ZmATL1* may function in radial patterning of Kranz type cells around veins rather than in vein spacing. However, as radial patterning is also observed in husk leaves (see section 7.3.3), the lack of *ZmATL1* expression in husk primordia makes this less likely.

7.3.2 Specification of BS and M cell type

In the *Arabidopsis* root, *SHR* is expressed within the stele (Helariutta et al., 2000) while *SCR* is expressed in the endodermal layer immediately surrounding it (Di Laurenzio et al., 1996). The movement of SHR protein to the neighbouring cell layer reinforces *SCR* expression, which leads to sequestration of the SHR protein and the specification of endodermal cell

identity (Helariutta et al., 2000; Nakajima et al., 2001; Heidstra et al., 2004; Cui et al., 2007). Likewise, in the *Arabidopsis* shoot system, expression of *SCR* is limited to a single cell layer around the vasculature: the starch sheath of the stem and the BS of leaves (Wysocka-Diller et al., 2000). Although the root endodermis and BS have been proposed to be equivalent tissues (Esau, 1965; Slewinski, 2013), the expression domains of *SHR* and *SCR* orthologues in maize leaves encompass an additional cell layer (Table 7.1) (Li et al., 2010). *ZmSHR1* is expressed in the BS while *ZmSCR1* is expressed preferentially in BS cells but also in the M. The mechanism for this remains unclear, however, it is possible that expression of *ZmSHR1* in the BS leads to excess quantities of protein that SCR is unable to sequester (Figure 7.1). Given the identification of both *ZmSHR1* and *ZmSCR1* by expression filters designed to identify Kranz regulators (Table 7.1), and the known role of *SHR* in cell layer specification in *Arabidopsis*, the extension of the *ZmSHR1* and *ZmSCR1* expression domains into leaf tissue may be crucial to C_4 BS and M cell type specification.

If ZmSHR1 protein accumulates in both BS and M cell types as proposed, it is unlikely that there will be equivalent protein levels in both cell types due to additional protein migration into the BS from the vascular tissue (Figure 7.1). Considering that SHR has been shown to affect radial patterning in a dose dependent manner in the *Arabidopsis* root (Koizumi et al., 2012), it is possible that BS and M cell specification could be a direct consequence of cellular SHR content, and that SHR is the diffusible Kranz specification factor originally proposed over twenty years ago (Langdale and Nelson, 1991). As

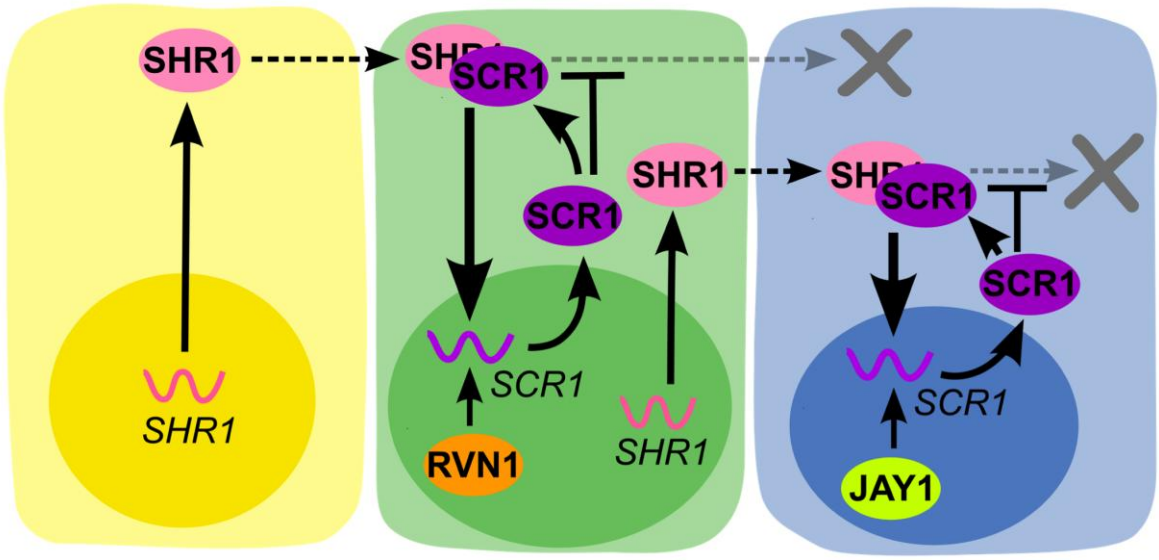


Figure 7.1. Schematic of a proposed model for *SHR* regulated BS and M cell-specification in maize foliar leaves. Distinct from *SHR* regulated root patterning, the model proposes that expression of *SHR* in the BS leads to the migration of *SHR* to an additional cell layer. Cell-specification is proposed to be a consequence of SHR level and cell-specifying factors such as *RVN1* and *JAY1*. Arrows imply promotion of gene expression, lines with 'T' ends imply inhibition of protein movement, dashed lines imply protein movement, waved lines represent mRNA, coloured ovals represent proteins. Yellow depicts vasculature, green depicts BS, and blue depicts M. Figure produced by Sayuri Ando.

proposed in this model, however, additional specification factors must be present for the two cell types to fully differentiate. In *Arabidopsis* root ground tissue, transcript accumulation of the *INDETERMINATE 1* (*ID1*)-*like* genes (Colasanti et al., 1998) *JKD* and *MGP* is strongly reduced in loss of function *shr* mutants while SHR is known to physically interact with both JKD and MGP (Welch et al., 2007). Loss of *jkd* function leads to a supernumerary ground tissue cell layer, however, this phenotype is largely complemented by loss of *mgp* function (Welch et al., 2007). This suggests that *ID1-like* genes are key components of the *SHR* root radial patterning network and that, within this network, *ID1-like* genes exhibit divergent functions. Two *ID1-like* genes were identified in the FA2 profile (and not the equivalent husk profile), one of which is enriched in BS cells (annotated as *ZmRAVEN1* (*ZmRvn1*)) while the other is enriched in M cells (annotated as *ZmJAY1*) (Table 7.1) (Li et al., 2010). It is thus tempting to speculate that BS and M cell identities are determined by the cell specific accumulation of *ID1-like* transcripts, which contribute to divergent *SHR* regulatory networks in the two cell types. Although *SHR* mediated patterning networks could be key to the regulation of Kranz development, they are unlikely to function exclusively. This is supported by the observation that two *DOF AFFECTING GERMINATION* (*DAG*) like genes (annotated here as *ZmDAG-like 1* (*ZmDLK1*) and *ZmDLK2*) are present in the FA2 profile and that transcripts are highly enriched in BS cells (Table 7.1) (Li et al., 2010). *DAG* genes are not known to operate within *SHR* networks, however, genes within the *DOF* clade are suggested to function in vascular patterning in *Arabidopsis*,(Gardiner et al., 2010; Gandotra et al., 2013).

Husk leaves exhibit functional Kranz units with C_4 BS and M cells encircling veins, however, in between these veins M cells accumulate Rubisco and may operate partial C_3 cycles (Langdale et al., 1988b; Pengelly et al., 2011). The model described above proposes that maize C_4 BS and M cell-specification is determined by the accumulation, and possibly level, of ZmSHR1, acting in combination with cell-specific factors including either *ZmRVN1* or *ZmJAY1*. As such, it is proposed that the same networks that operate in foliar Kranz development also operate in husk leaves (Figure 7.2). This is supported by the observation that *ZmSHR1*, *ZmSCR1*, *ZmRVN1*, *ZmJAY1*, *ZmDLK1* and *ZmDLK2* are all expressed in husk primordia samples, although to a lower level than in foliar primordia. The development of non-C_4 type M cells that differentiate when not in contact with a BS cell (Pengelly et al., 2011) is proposed to be a consequence of a lack of ZmSHR1 accumulation (Figure 7.2). It is possible that the proposed C_4 M cell-specification factor *ZmJAY1* may still be expressed in non-C_4 M cells, but as it has been shown that *JKD* and *MGP* expression decreases in the absence of *SHR* in *Arabidopsis* (Welch et al., 2007), this is unlikely.

7.3.3 Integration of the C_4 pathway

Based on profile expression analyses, it was proposed in Chapter 4 that integration of a functional C_4 pathway follows BS and M cell specification. This is supported by anatomical and proteomic analyses that show that BS and M cell identities are determined prior to the onset of distinct metabolic pathways in the two cell types (Majeran et al., 2010). As such, the genes identified in

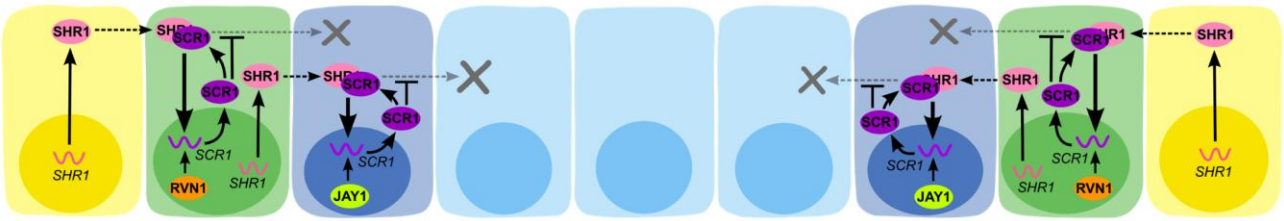


Figure 7.2. Schematic of a proposed model for *SHR* regulated BS and M cell-specification in maize husk leaves. Arrows imply promotion of gene expression, lines with 'T' ends imply inhibition of protein movement, dashed lines imply protein movement, waved lines represent mRNA, coloured ovals represent proteins. Yellow depicts vasculature, green depicts BS, dark blue depicts C_4 M and light blue depicts non-C_4 M. Figure produced by Sayuri Ando.

Table 7.1 may function too early in development to regulate the onset of the C_4 pathway. However, the observation that a suite of photosynthetic genes, including the C_4 pathway genes *NADP-ME* and *RbcS* were misexpressed in a maize *id1* loss of function mutant (Coneva et al., 2007), suggests that *ID1-like* genes may function throughout BS and M cell development. Further insights into putative regulators of C_4 pathway initiation would be provided through modification of the expression filters described here to identify genes expressed in the appropriate developmental stages of maize development. These are likely to be found within the FI signature gene list (Appendix 4.4)

7.3.4 Testable predictions of the model for C_4 development

The assumptions of the model proposed above make several predictions that can be tested experimentally in a variety of ways. Firstly, if the extension of *SHR* expression into the BS is a key facilitator for C_4 evolution it should also be observed in other C_4 lineages. RNA-seq of BS and M cells from the C_4 species *Setaria viridis* (cells isolated mechanically; C. John, S. Kelly, J. Hibberd unpublished data) and *Cleome gynandra* (cells isolated by LCM; S. Aubry, S. Kelly, J. Hibberd unpublished data) shows that *ZmSHR1* orthologues are expressed in BS cells in both species. As in maize, the *C. gynandra ZmSCR1* orthologue is also expressed in both BS and M cells (a missing *SCR* gene model annotation in the *S. italica* genome means that *SCR* expression data is not available for *S. viridis*). Further, *Setaria ZmJAY1*, *ZmRVN1* and *ZmDLK2* orthologues display the same direction of expression gradients as observed in maize (Table 7.1). *ID1-like* gene orthology is less

clear in *Cleome*, however, a *ZmJAY1* co-orthologue also shows enrichment in M cells and a *ZmRVN1* co-orthologue in BS cells. The orthologue of *ZmDLK1* and *ZmDLK2* is similarly enriched in *C. gynandra* BS cells. The striking similarity of the expression patterns of these genes in such phylogenetically distant C_4 species is compelling evidence for the role of *SHR* in the repeated independent evolution of Kranz development.

A second prediction of the model is loss of function mutant phenotypes. In maize, loss of *Zmscr1* function has been shown to produce a variety of perturbations to leaf anatomy, but most notably a loss of M cells between veins, supernumerary BS layers and reduced vein density (Slewinski et al., 2012). It has been shown that loss of *scr* function in *Arabidopsis* leads to the loss of a ground tissue layer in the root (Di Laurenzio et al., 1996), therefore it is possible that the loss of M cells in the *Zmscr1* mutant in an equivalent defect (Figure 7.3). If ZmSHR1 determines BS and M cell identity in a dose dependent manner (Koizumi et al., 2012) as proposed, without sequestration by ZmSCR1 in the *Zmscr1* mutant, the additional free ZmSHR1 may promote ectopic BS development. As such, altered *SHR* and *SCR* expression patterns may explain the variation in BS cell layer numbers observed in the Poales (Edwards and Voznesenskaya, 2011). Further, if BS specific expression of *SHR* is sufficient for BS cell identity and C_4 radial patterning, this model would be consistent with the initiation of C_4 development around BS-like distinctive cells that form without veins in some C_4 grass lineages. As loss of *scr* function affects patterns of cell division (Di Laurenzio et al., 1996) the reduced vein

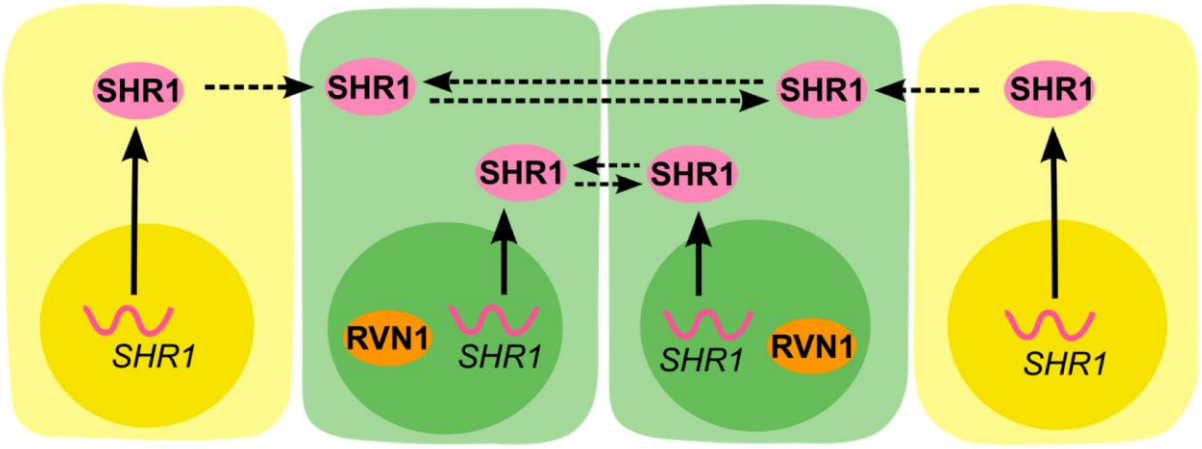


Figure 7.3. **Schematic of a proposed model for *SHR* regulated BS and M cell-specification in a *Zmscr1* mutant.** Cell division defects in a *scr* loss of function mutant may lead to a loss of M cells. Arrows imply promotion of gene expression, dashed lines imply protein movement, waved lines represent mRNA, coloured ovals represent proteins. Yellow depicts vasculature, green depicts BS. Figure produced by Sayuri Ando.

spacing observed in the *Zmscr1* mutant may be a consequence of increased divisions of M cells. Loss of function mutants or knockdown lines for other key Kranz regulatory genes proposed above have not been described in C_4 species. Based on the proposed model, it is expected that loss of function to *Zmshr1*, *Zmshr2*, *Zmdot1*, *Zmspl1*, *Zmatl1 Zmmpl1* and *Zmmpl2* will lead to perturbed procambial initiation and reduced vein spacing. Loss of *Zmshr1* function is also likely to perturb BS development. Similarly, loss of function of the proposed cell-specification factors *Zmrvn1*, *Zmdlk1* and *Zmdlk2* is expected to lead to defective BS development (although likely genetic redundancy may require double *Zmdlk1* and *Zmdlk2* knockouts) while loss of *Zmjay1* function is expected to lead to defective M cell development. This may be easiest to test by RNAi in orthologous *Setaria* genes.

The third prediction of the model is that ectopic expression of these genes in C_3 species will induce elements of Kranz development. For example, ectopic expression of the proposed maize procambial regulators should increase rates of procambial initiation in a C_3 leaf context, while ectopic expression of the proposed cell-specification factors should alter C_3 cell identities. However, in all cases, as these genes are proposed to act in a network, ectopic expression in a C_3 species may require the introduction of a cohort of genes simultaneously to reveal their Kranz potential in a C_3 leaf. Likewise, ectopic expression may need to be restricted to tight spatiotemporal domains in order to reveal an effect, particularly for the proposed cell-specification factors.

Finally, considering the importance of ZmSHR1 migration to the model proposed above, translational fusion or immunolocalisation assays will be of key importance for model validation. It is predicted that in Kranz species, SHR protein will be limited to M cells in contact with BS cells, that M cells will accumulate less SHR than BS cells and that no SHR will migrate to the epidermis (Figure 7.1).

7.4 Contribution towards the development of C_4 rice

This study commenced as part of a collaborative international effort to develop a C_4 rice variety. To generate C_4 rice a suite of developmental and physiological alterations need to be made (Hibberd et al., 2008), and key to this will be the introduction of Kranz-type anatomy. The bioinformatics analyses described here have produced a candidate gene list of a feasible magnitude for functional characterisation. Three candidate genes have been functionally tested and shown to have no effect on Kranz development, reducing the candidate gene list further. As part of these characterisation assays, proof of principle of a knockdown approach in *Setaria viridis* has been provided, which is key to future work. Further, through detailed quantitative analyses of leaf development, parameters of *Setaria* leaf growth have been defined, and distinct components of C_4 growth revealed. In summary, this study has provided a firm foundation and a testable model for the future selection and characterisation of candidate Kranz regulators.

Digital appendix legend

Appendix 2.1. List of primer sequences used in this study.

Appendix 3.1. Raw qPCR data for validation of RNA-seq data.

Appendix 4.1. mRNA abundance estimates and annotations for FP signature genes.

Appendix 4.2. mRNA abundance estimates and annotations for FP3/4 signature genes.

Appendix 4.3. mRNA abundance estimates, annotations and enrichment test results for GO terms, MaizeCyc pathways, Pfam domains and MapMan terms for FP5 signature genes.

Appendix 4.4. mRNA abundance estimates, annotations and enrichment test results for GO terms, MaizeCyc pathways, Pfam domains and MapMan terms for FI signature genes.

Appendix 4.5. mRNA abundance estimates, annotations and enrichment test results for GO terms, MaizeCyc pathways, Pfam domains and MapMan terms for FE signature genes.

Appendix 4.6. mRNA abundance estimates and annotations for HP signature genes.

Appendix 4.7. mRNA abundance estimates and annotations for HP3/4 signature genes.

Appendix 4.8. mRNA abundance estimates, annotations and enrichment test results for GO terms, MaizeCyc pathways, Pfam domains and MapMan terms for HP5 signature genes.

Appendix 4.9. mRNA abundance estimates, annotations and enrichment test results for GO terms, MaizeCyc pathways, Pfam domains and MapMan terms for HI signature genes.

Appendix 4.10. mRNA abundance estimates, annotations and enrichment test results for GO terms, MaizeCyc pathways, Pfam domains and MapMan terms for HE signature genes.

Appendix 4.11. GO term enrichment in the signature genes enriched within leaf type. GO terms in FP/HP (purple segments), FP3/4 & HP/3/4 (orange segments), FP5/HP5 (yellow segments), FI/HI (light green segments) and FE/HE (dark green segments) samples. GO terms that are significantly enriched in the signature gene set of a single sample compared to the other four samples from the same leaf are displayed. Where this occurs in both foliar (F) and husk (H) samples GO terms are presented in black text; foliar only – blue; husk only – red.

Appendix 4.12. Pathway enrichment in the signature genes enriched within leaf type. Metabolic pathways that are significantly enriched in the signature gene set of a single sample compared to the other four samples from the same leaf are displayed. Where this occurs in both foliar (F) and husk (H) samples metabolic pathways are presented in black text; foliar only – blue; husk only – red. FI/HI are depicted as light green segments; FE/HE as dark green segments.

Appendix 4.13. mRNA abundance estimates, annotations and enrichment test results for GO terms, Pfam domains and MapMan terms for FD1 profile genes.

Appendix 4.14. mRNA abundance estimates, annotations and enrichment test results for GO terms and MapMan terms for FD2 profile genes.

Appendix 4.15. mRNA abundance estimates, annotations and enrichment test results for GO terms, MaizeCyc pathways, Pfam domains and MapMan terms for FD3 profile genes.

Appendix 4.16. mRNA abundance estimates, annotations and enrichment test results for GO terms, Pfam domains and MapMan terms for FA1 profile genes.

Appendix 4.17. mRNA abundance estimates, annotations and enrichment test results for GO terms, MaizeCyc pathways, Pfam domains and MapMan terms for FA2 profile genes.

Appendix 4.18. mRNA abundance estimates, annotations and enrichment test results for GO terms, MaizeCyc pathways, Pfam domains and MapMan terms for FA3 profile genes.

Appendix 4.19. mRNA abundance estimates, annotations and enrichment test results for GO terms, MaizeCyc pathways, Pfam domains and MapMan terms for FN profile genes.

Appendix 4.20. mRNA abundance estimates, annotations and enrichment test results for GO terms, Pfam domains and MapMan terms for HD1 profile genes.

Appendix 4.21. mRNA abundance estimates, annotations and enrichment test results for GO terms, Pfam domains and MapMan terms for HD2 profile genes.

Appendix 4.22. mRNA abundance estimates, and annotations for HD3 profile genes.

Appendix 4.23. mRNA abundance estimates, annotations and enrichment test results for GO terms, MaizeCyc pathways, Pfam domains and MapMan terms for HA1 profile genes.

Appendix 4.24. mRNA abundance estimates, annotations and enrichment test results for GO terms, Pfam domains and MapMan terms for HA2 profile genes.

Appendix 4.25. mRNA abundance estimates, annotations and enrichment test results for GO terms, MaizeCyc pathways, Pfam domains and MapMan terms for HA3 profile genes.

Appendix 4.26. mRNA abundance estimates, annotations and enrichment test results for GO terms, Pfam domains and MapMan terms for for HN profile genes.

Appendix 4.27. GO term enrichments for all profiles. Pie charts show significantly enriched GO terms in that profile ($p < 0.05$), size of segment represents number of genes annotated in each profile annotated with a particular GO term. HD3 is not present as no GO terms were enriched in this profile.

Appendix 4.28. Summary data for all detected genes in transcriptomic analysis showing maize ID, EntrezGene classification, RPKM values for replicates of all 10 samples, presence in signature gene set (if relevant) plus affiliation with specific foliar and husk profiles.

Appendix 4.29. Manual annotation of transcription factors in all descending and ascending profiles.

Appendix 4.30. Genes identified as putative positive Kranz regulators or in FA2 profile, and as upregulated in *Arabidopsis* provascular or procambial cells

(Gandotra et al 2013) or in embryonic maize leaves post imbibition (Liu et al 2013).

Appendix 4.31. mRNA abundance estimates, annotations and enrichment test results for GO terms, Pfam domains and MapMan terms for 283 putative positive regulators of Kranz.

Appendix 4.32. mRNA abundance estimates, annotations and enrichment test results for GO terms, Pfam domains and MapMan terms for 142 putative negative regulators of Kranz.

Appendix 4.33. Summary of expression data of maize orthologues of *Arabidopsis* genes implicated in vascular patterning.

Appendix 4.34. Summary of BS and M enrichment of genes identified in expression profiles and as putative positive Kranz regulators. BS and M expression based on studies by Li et al (2010) and Chang et al (2012).

Appendix 4.35. Summary of expression data for orthologues of SHORT-ROOT (SHR) targets in *Arabidopsis* identified in expression profiles and putative positive Kranz candidate list. SHR targets identified in studies by Levesque et al (2006) and Cui et al (2011).

Appendix 5.1. Semi-quantitative PCR validation of transgene expression in rice. 1µg of RNA was used for cDNA production, rice *UBQ* gene Os03g13170 was used as a control. A) bHLH-OE lines B) LRR-RLK-OE lines.

Appendix 6.1. Summary of *GLK* sequences accessions used for phylogeny construction

Appendix 6.2. Schematic diagrams of motifs identified upstream of *ZmGLK1* and *SbGLK1* but not *OsGLK1* or *BdGLKA.*

Appendix 6.3. Schematic diagrams of motifs identified upstream of *ZmG2* and *SbGLK2* but not *OsGLK2* or *BdGLKB.*

Appendix 6.4. Summary of *GLK* upstream motif profiles and statistical support for transcription factor motif binding.

References

Aharoni, A., Dixit, S., Jetter, R., Thoenes, E., van Arkel, G., and Pereira, A. (2004). The *SHINE* clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in *Arabidopsis*. Plant Cell **16,** 2463-2480.

Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., and Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biology **12,** R18.

Akhani, H., Barroca, J., Koteeva, N., Voznesenskaya, E., Franceschi, V., Edwards, G., Ghaffari, S., and Ziegler, H. (2005). *Bienertia sinuspersici* (Chenopodiaceae): A new species from southwest Asia and discovery of a third terrestrial C_4 plant without Kranz anatomy. Systematic Botany **30,** 290-301.

Akyildiz, M., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., and Westhoff, P. (2007). Evolution and function of a cis-regulatory module for mesophyll-specific gene expression in the C_4 dicot *Flaveria trinervia*. Plant Cell **19,** 3391-3402.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2008). Molecular biology of the cell. (New York: Garland Science).

Alexandrov, N.N., Brover, V.V., Freidin, S., Troukhan, M.E., Tatarinova, T.V., Zhang, H., Swaller, T.J., Lu, Y.-P., Bouck, J., Flavell, R.B., and

Feldmann, K.A. (2009). Insights into corn genes derived from large-scale cDNA sequencing. Plant Mol Biol **69,** 179-194.

Alves, S., Worland, B., Thole, V., Snape, J., Bevan, M., and Vain, P. (2009). A protocol for *Agrobacterium*-mediated transformation of *Brachypodium distachyon* community standard line Bd21. Nat Protoc **4,** 638-649.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biology **11,** R106.

Arnon, D.I. (1949). Copper enzymes in isolated chloroplasts. Polyphenoloxidase in *Beta vulgaris*. Plant Physiology **24,** 1-15.

Aubry, S., Brown, N.J., and Hibberd, J.M. (2011). The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway. J Exp Bot **62,** 3049-3059.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J.Y., Li, W.W., and Noble, W.S. (2009). MEME Suite: tools for motif discovery and searching. Nucleic Acids Research **37,** W202-W208.

Ballesteros, M.L., Bolle, C., Lois, L.M., Moore, J.M., Vielle-Calzada, J.P., Grossniklaus, U., and Chua, N.H. (2001). *LAF1*, a MYB transcription activator for phytochrome A signaling. Gene Dev **15,** 2613-2625.

Bassel, G.W., Gaudinier, A., Brady, S.M., Hennig, L., Rhee, S.Y., and De Smet, I. (2012). Systems analysis of plant functional, transcriptional, physical interaction, and metabolic networks. Plant Cell **24,** 3859-3875.

Bassham, J.A., Barker, S.A., Calvin, M., and Quarck, U.C. (1956). Intermediates in the photosynthetic cycle. Biochim Biophys Acta **21,** 376-377.

Baumbusch, L.O., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R., Schulz, I., Reuter, G., and Aalen, R.B. (2001). The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Research **29,** 4319-4333.

Bauwe, H. (2011). Photorespiration: The Bridge to C_4 Photosynthesis. In C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms, A.S. Raghavendra and R. Sage, eds (Dordrecht: Springer), pp. 81-108.

Benfey, P.N., Linstead, P.J., Roberts, K., Schiefelbein, J.W., Hauser, M.T., and Aeschbacher, R.A. (1993). Root development in *Arabidopsis* - 4 mutants with dramatically altered root morphogenesis. Development **119,** 57-70.

Bennetzen, J.L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A.C., Estep, M., Feng, L., Vaughn, J.N., Grimwood, J., Jenkins, J., Barry, K., Lindquist, E., Hellsten, U., Deshpande, S., Wang, X.W., Wu, X.M., Mitros, T., Triplett, J., Yang, X.H., Ye, C.Y., Mauro-Herrera, M., Wang, L., Li, P.H., Sharma, M., Sharma, R., Ronald, P.C., Panaud, O., Kellogg, E.A., Brutnell, T.P., Doust, A.N., Tuskan, G.A., Rokhsar, D., and Devos, K.M. (2012). Reference genome sequence of the model plant *Setaria*. Nat Biotechnol **30,** 555-561.

Blanc, G., Hokamp, K., and Wolfe, K.H. (2003). A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. Genome Research **13,** 137-144.

Blasing, O., Westhoff, P., and Svensson, P. (2000). Evolution of C_4 phosphoenolpyruvate carboxylase in *Flaveria*, a conserved serine residue in the carboxyl-terminal part of the enzyme is a major determinant for C_4-specific characteristics. Journal of Biological Chemistry **275,** 27917-27923.

Bomblies, K., Wang, R.L., Ambrose, B.A., Schmidt, R.J., Meeley, R.B., and Doebley, J. (2003). Duplicate *FLORICAULA*/*LEAFY* homologs *zfl1* and *zfl2* control inflorescence architecture and flower patterning in maize. Development **130,** 2385-2395.

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

Bosabalidis, A.M., Evert, R.F., and Russin, W.A. (1994). Ontogeny of the vascular bundles and contiguous tissues in the maize leaf blade. American Journal of Botany **81,** 745-752.

Bowes, G. (2011). Single-Cell C_4 Photosynthesis in Aquatic Plants. In C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms, A.S. Raghavendra and R. Sage, eds (Dordrecht: Springer), pp. 63-80.

Brautigam, A., Hoffmann-Benning, S., and Weber, A. (2008). Comparative proteomics of chloroplast envelopes from C_3 and C_4 plants reveals specific adaptations of the plastid envelope to C_4 photosynthesis and

candidate proteins required for maintaining C_4 metabolite fluxes. Plant Physiology **148,** 568-579.

Brautigam, A., Kajala, K., Wullenweber, J., Sommer, M., Gagneul, D., Weber, K.L., Carr, K.M., Gowik, U., Mass, J., Lercher, M.J., Westhoff, P., Hibberd, J.M., and Weber, A.P.M. (2011). An mRNA blueprint for C_4 photosynthesis derived from comparative transcriptomics of closely related C_3 and C_4 species. Plant Physiology **155,** 142-156.

Brown, N.J., Newell, C.A., Stanley, S., Chen, J.E., Perrin, A.J., Kajala, K., and Hibberd, J.M. (2011). Independent and parallel recruitment of preexisting mechanisms underlying C_4 photosynthesis. Science **331,** 1436-1439.

Brown, W.V. (1975). Variations in anatomy, associations, and origins of Kranz tissue. American Journal of Botany **62,** 395-402.

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

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

Bryne, J.C., Valen, E., Tang, M.H.E., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B., and Sandelin, A. (2008). JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Research **36,** D102-D106.

Candela, H., Johnston, R., Gerhold, A., Foster, T., and Hake, S. (2008). The *milkweed pod1* gene encodes a KANADI protein that is required for abaxial/adaxial patterning in maize leaves. The Plant cell **20,** 2073-2087.

Cannon, S.B., Ilut, D., Farmer, A.D., Maki, S.L., May, G.D., Singer, S.R., and Doyle, J.J. (2010). Polyploidy did not predate the evolution of nodulation in all legumes. PLoS ONE **5,** e11630.

Carles, C.C., Choffnes-Inada, D., Reville, K., Lertpiriyapong, K., and Fletcher, J.C. (2005). *ULTRAPETALA1* encodes a SAND domain putative transcriptional regulator that controls shoot and floral meristem activity in *Arabidopsis*. Development **132,** 897-911.

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

Cerling, T., Harris, J., MacFadden, B., Leakey, M., Quade, J., Eisenmann, V., and Ehleringer, J. (1997). Global vegetation change through the Miocene/Pliocene boundary. Nature **389,** 153-158.

Chang, Y.M., Liu, W.Y., Shih, A.C.C., Shen, M.N., Lu, C.H., Lu, M.Y.J., Yang, H.W., Wang, T.Y., Chen, S.C.C., Chen, S.M., Li, W.H., and Ku, M.S.B. (2012). Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. Plant Physiology **160,** 165-177.

Cheung, F., Haas, B.J., Goldberg, S.M., May, G.D., Xiao, Y., and Town, C.D. (2006). Sequencing *Medicago truncatula* expressed sequenced tags using 454 Life Sciences technology. BMC Genomics **7,** 272.

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

Christin, P.-A., Petitpierre, B., Salamin, N., Buchi, L., and Besnard, G. (2009b). Evolution of C_4 phospho*enol*pyruvate carboxykinase in grasses, from genotype to phenotype. Molecular Biology and Evolution **26,** 357-365.

Christin, P.-A., Salamin, N., Muasya, A.M., Roalson, E.H., Russier, F., and Besnard, G. (2008a). Evolutionary switch and genetic convergence on *rbcL* following the evolution of C_4 photosynthesis. Molecular Biology and Evolution **25,** 2361-2368.

Christin, P.-A., Sage, T.L., Edwards, E.J., Ogburn, R.M., Khoshravesh, R., and Sage, R.F. (2011). Complex evolutionary transitions and the significance of C_3-C_4 intermediate forms of photosynthesis in Molluginaceae. Evolution **65,** 643-660.

Christin, P.A., Freckleton, R.P., and Osborne, C.P. (2010a). Can phylogenetics identify C_4 origins and reversals? Trends Ecol Evol **25,** 403-409.

Christin, P.A., Salamin, N., Savolainen, V., Duvall, M.R., and Besnard, G. (2007). C_4 photosynthesis evolved in grasses via parallel adaptive genetic changes. Current biology **17,** 1241-1247.

Christin, P.A., Samaritani, E., Petitpierre, B., Salamin, N., and Besnard, G. (2010b). Evolutionary insights on C_4 photosynthetic subtypes in grasses from genomics and phylogenetics. Genome Biology and Evolution **1,** 221-230.

Christin, P.A., Besnard, G., Samaritani, E., Duvall, M.R., Hodkinson, T.R., Savolainen, V., and Salamin, N. (2008b). Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. Current Biology **18,** 37-43.

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

Chuang, C.F., Running, M.P., Williams, R.W., and Meyerowitz, E.M. (1999). The *PERIANTHIA* gene encodes a bZIP protein involved in the determination of floral organ number in *Arabidopsis thaliana*. Gene Dev **13,** 334-344.

Chuck, G., Meeley, R., and Hake, S. (2008). Floral meristem initiation and meristem cell fate are regulated by the maize AP2 genes ids1 and sid1. Development **135,** 3013-3019.

Colasanti, J., Yuan, Z., and Sundaresan, V. (1998). The *indeterminate* gene encodes a zinc finger protein and regulates a leaf-generated signal required for the transition to flowering in maize. Cell **93,** 593-603.

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

Coneva, V., Zhu, T., and Colasanti, J. (2007). Expression differences between normal and *indeterminate1* maize suggest downstream targets of *ID1*, a floral transition regulator in maize. J Exp Bot **58,** 3679-3693.

Cook, N. (1998). Build it up - Tear it Down. In You've Come A Long Way, Baby (Skint Records).

Covshoff, S., Furbank, R.T., Leegood, R.C., and Hibberd, J.M. (2013). Leaf rolling allows quantification of mRNA abundance in mesophyll cells of sorghum. J Exp Bot **64,** 807-813.

Covshoff, S., Majeran, W., Liu, P., Kolkman, J., van Wijk, K., and Brutnell, T. (2008). Deregulation of maize C_4 photosynthetic development in a mesophyll cell-defective mutant. Plant Physiology **146,** 1469-1481.

Crookston, R.K., and Moss, D. (1973). Variation of C_4 leaf anatomy in *Arundinella hirta* (Gramineae). Plant Physiology **52,** 397-402.

Croxdale, J.L. (2000). Stomatal patterning in angiosperms. American Journal of Botany **87,** 1069-1080.

Cui, H., Hao, Y., Kovtun, M., Stolc, V., Deng, X.W., Sakakibara, H., and Kojima, M. (2011). Genome-wide direct target analysis reveals a role for *SHORT-ROOT* in root vascular patterning through cytokinin homeostasis. Plant Physiology **157,** 1221-1231.

Cui, H.C., Levesque, M.P., Vernoux, T., Jung, J.W., Paquette, A.J., Gallagher, K.L., Wang, J.Y., Blilou, I., Scheres, B., and Benfey, P.N. (2007). An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. Science **316,** 421-425.

Danker, T., Dreesen, B., Offermann, S., Horst, I., and Peterhansel, C. (2008). Developmental information but not promoter activity controls the methylation state of histone H3 lysine 4 on two photosynthetic genes in maize. Plant Journal **53,** 465-474.

Dawe, D. (2007). Agricultural Research, Poverty Alleviation and Key Trends in Asia's Rice Economy. In Charting New Pathways to C_4 Rice, J.E. Sheehy, P.L. Mitchell, and B. Hardy, eds (Los Banos, Philippines: World Scientific), pp. 37-53.

de Oliveira Dal'Molin, C.G., Quek, L.E., Palfreyman, R.W., Brumbley, S.M., and Nielsen, L.K. (2010). C4GEM, a genome-scale metabolic model to study C_4 plant metabolism. Plant Physiology **154,** 1871-1885.

Dengler, N., and Nelson, T. (1999). Leaf Structure and Development in C_4 Plants. In C_4 Plant Biology, R. Sage and R. Monson, eds (San Diego: Academic Press), pp. 133-172.

Dengler, N.G., Dengler, R.E., and Hattersley, P.W. (1985). Differing ontogenetic origins of PCR (Kranz) sheaths in leaf blades of C_4 grasses (Poaceae). American Journal of Botany **72,** 284-302.

Dengler, N.G., Dengler, R.E., and Grenville, D.J. (1990). Comparison of photosynthetic carbon-reduction (Kranz) cells having different ontogenic origins in the C_4 NADP - malic enzyme grass *Arundinella hirta*. Canadian Journal of Botany-Revue Canadienne De Botanique **68,** 1222-1232.

Dengler, N.G., Woodvine, M.A., Donnelly, P.M., and Dengler, R.E. (1997). Formation of vascular pattern in developing leaves of the C_4 grass *Arundinella hirta*. International Journal of Plant Sciences **158,** 1-12.

Dengler, R.E., and Dengler, N.G. (1990). Leaf vascular architecture in the atypical C_4 NADP - malic enzyme grass *Arundinella hirta*. Canadian Journal of Botany-Revue Canadienne De Botanique **68,** 1208-1221.

Dhonukshe, P., Weits, D.A., Cruz-Ramirez, A., Deinum, E.E., Tindemans, S.H., Kakar, K., Prasad, K., Mahonen, A.P., Ambrose, C., Sasabe, M., Wachsmann, G., Luijten, M., Bennett, T., Machida, Y., Heidstra, R., Wasteneys, G., Mulder, B.M., and Scheres, B. (2012). A *PLETHORA*-auxin transcription module controls cell division plane rotation through MAP65 and CLASP. Cell **149,** 383-396.

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

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

Downs, G.S., Bi, Y.M., Colasanti, J., Wu, W.Q., Chen, X., Zhu, T., Rothstein, S.J., and Lukens, L.N. (2013). A developmental transcriptional network for maize defines coexpression modules. Plant Physiology **161,** 1830-1843.

Drincovich, M.F., Lara, M.V., Andreo, C.S., and Maurino, V.G. (2011). C_4 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,

A.S. Raghavendra and R. Sage, eds (Dordrecht: Springer), pp. 277-300.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis*. Trends Plant Sci **15,** 573-581.

Eddy, S.R. (1998). Profile hidden Markov models. Bioinformatics **14,** 755-763.

Edwards, E.J., and Smith, S.A. (2010). Phylogenetic analyses reveal the shady history of C_4 grasses. Proceedings of the National Academy of Sciences of the United States of America **107,** 2532-2537.

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

Ehleringer, J.R., Cerling, T.E., and Helliker, B.R. (1997). C_4 photosynthesis, atmospheric CO_2 and climate. Oecologia **112,** 285-299.

Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L. (2003). Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and KANADI genes. Current Biology **13,** 1768-1774.

Emrich, S.J., Barbazuk, W.B., Li, L., and Schnable, P.S. (2007). Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome research **17,** 69-73.

Esau, K. (1943). Ontogeny of the vascular bundle in Zea mays. Hilgardia **15,** 327-368.

Esau, K. (1965). Plant Anatomy. (New York: John Wiley and Sons).

Eshed, Y., Baum, S.F., Perea, J.V., and Bowman, J.L. (2001). Establishment of polarity in lateral organs of plants. Current Biology **11,** 1251-1260.

Evans, M.M. (2007). The *indeterminate gametophyte1* gene of maize encodes a LOB domain protein required for embryo sac and leaf development. Plant Cell **19,** 46-62.

Feinberg, A.P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem **132,** 6-13.

Ficklin, S.P., and Feltus, F.A. (2011). Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice. Plant Physiology **156,** 1244-1256.

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

Fitter, D., Martin, D., Copley, M., Scotland, R., and Langdale, J. (2002). *GLK* gene pairs regulate chloroplast development in diverse plant species. Plant Journal **31,** 713-727.

Fladung, M. (1994). Genetic variants of *Panicum maximum* (Jacq.) in C_4 photosynthetic traits. Journal of Plant Physiology **143,** 165-172.

Folkers, U., Berger, J., and Hulskamp, M. (1997). Cell morphogenesis of trichomes in *Arabidopsis*: differential control of primary and secondary branching by branch initiation regulators and cell growth. Development **124,** 3779-3786.

Friso, G., Majeran, W., Huang, M., Sun, Q., and van Wijk, K. (2010). Reconstruction of metabolic pathways, protein expression, and homeostasis machineries across maize bundle sheath and mesophyll chloroplasts: large-scale quantitative proteomics using the first maize genome assembly. Plant Physiology **152,** 1219-1250.

Fukaki, H., Wysocka-Diller, J., Kato, T., Fujisawa, H., Benfey, P.N., and Tasaka, M. (1998). Genetic evidence that the endodermis is essential for shoot gravitropism in *Arabidopsis thaliana*. Plant Journal **14,** 425-430.

Furbank, R.T. (2011). Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? J Exp Bot **62,** 3103-3108.

Furumoto, T., Yamaguchi, T., Ohshima-Ichie, Y., Nakamura, M., Tsuchida-Iwata, Y., Shimamura, M., Ohnishi, J., Hata, S., Gowik, U., Westhoff, P., Brautigam, A., Weber, A.P.M., and Izui, K. (2011). A plastidial sodium-dependent pyruvate transporter. Nature **476,** 472-475.

Galego, L., and Almeida, J. (2002). Role of *DIVARICATA* in the control of dorsoventral asymmetry in *Antirrhinum* flowers. Gene Dev **16,** 880-891.

Galtier, N. (2001). Maximum-likelihood phylogenetic analysis under a covarion-like model. Molecular Biology and Evolution **18,** 866-873.

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

Garber, M., Grabherr, M.G., Guttman, M., and Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. Nature Methods **8,** 469-477.

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

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

Ghannoum, O., Evans, J.R., and von Caemmerer, S. (2011). Nitrogen and Water Use Efficiency of C_4 Plants. In C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms, A.S. Raghavendra and R. Sage, eds (Dordrecht: Springer), pp. 129-146.

Gibalova, A., Renak, D., Matczuk, K., Dupl'akova, N., Chab, D., Twell, D., and Honys, D. (2009). *AtbZIP34* is required for *Arabidopsis* pollen wall patterning and the control of several metabolic pathways in developing pollen. Plant Mol Biol **70,** 581-601.

Gilding, E.K., and Marks, M.D. (2010). Analysis of purified *glabra3-shapeshifter* trichomes reveals a role for *NOECK* in regulating early trichome morphogenic events. Plant Journal **64,** 304-317.

Gould, S.J. (1977). Ontogeny and phylogeny. (Cambridge, Mass ; London: Belknap Press of Harvard University Press).

Gowik, U., Engelmann, S., Blasing, O.E., Raghavendra, A.S., and Westhoff, P. (2006). Evolution of C_4 phospho*enol*pyruvate carboxylase

in the genus *Alternanthera*: gene families and the enzymatic characteristics of the C_4 isozyme and its orthologues in C_3 and C_3/C_4 *Alternantheras*. Planta **223,** 359-368.

Gowik, U., Brautigam, A., Weber, K.L., Weber, A.P.M., and Westhoff, P. (2011). Evolution of C_4 photosynthesis in the genus *Flaveria*: how many and which genes does it take to make C_4? Plant Cell **23,** 2087-2105.

Gowik, U., Burscheidt, J., Akyildiz, M., Schlue, U., Koczor, M., Streubel, M., and Westhoff, P. (2004). *cis*-Regulatory elements for mesophyll-specific gene expression in the C_4 plant *Flaveria trinervia*, the promoter of the C_4 phospho*enol*pyruvate carboxylase gene. Plant Cell **16,** 1077-1090.

Gualberti, G., Papi, M., Bellucci, L., Ricci, L., Bouchez, D., Camilleri, C., Costantino, P., and Vittorioso, P. (2002). Mutations in the Dof zinc finger genes *DAG2* and *DAG1* influence with opposite effects the germination of *Arabidopsis* seeds. Plant Cell **14,** 1253-1263.

Guo, H., and Ecker, J.R. (2004). The ethylene signaling pathway: new insights. Curr Opin Plant Biol **7,** 40-49.

Guo, Y., Qin, G.J., Gu, H.Y., and Qu, L.J. (2009). *Dof5.6*/*HCA2*, a Dof transcription factor gene, regulates interfascicular cambium formation and vascular tissue development in *Arabidopsis*. Plant Cell **21,** 3518-3534.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biology **8**.

Gutierez, R.A., Stokes, T.L., Thum, K., Xu, X., Obertello, M., Katari, M.S., Tanurdzic, M., Dean, A., Nero, D.C., McClung, C.R., and Coruzzi, G.M. (2008). Systems approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene *CCA1*. Proceedings of the National Academy of Sciences of the United States of America **105,** 4939-4944.

Haberlandt, G. (1882). Vergleichende Anatomie des assimilatorischen Gewebesystems der Pflanzen. Jahrbuch der. Wissenschaftlichen Botanik **13,** 74-188.

Hall, L.N., Rossini, L., Cribb, L., and Langdale, J.A. (1998). *GOLDEN 2*: a novel transcriptional regulator of cellular differentiation in the maize leaf. Plant Cell **10,** 925-936.

Harris, J.C., Hrmova, M., Lopato, S., and Langridge, P. (2011). Modulation of plant growth by HD-Zip class I and II transcription factors in response to environmental stimuli. New Phytologist **190,** 823-837.

Hatch, M., and Slack, C.R. (1966). Photosynthesis by sugar cane leaves: a new carboxylation reaction and the pathway of sugar formation. Biochemical Journal **101,** 103-111.

Hauser, B.A., Villanueva, J.M., and Gasser, C.S. (1998). *Arabidopsis TSO1* regulates directional processes in cells during floral organogenesis. Genetics **150,** 411-423.

Hay, A., and Tsiantis, M. (2010). *KNOX* genes: versatile regulators of plant development and diversity. Development **137,** 3153-3165.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A.P., and Lercher, M.J. (2013). Predicting C_4 photosynthesis

evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. Cell **153,** 1579-1588.

Heidstra, R., Welch, D., and Scheres, B. (2004). Mosaic analyses using marked activation and deletion clones dissect *Arabidopsis SCARECROW* action in asymmetric cell division. Gene Dev **18,** 1964-1969.

Heimann, L., Horst, I., Perduns, R., Dreesen, B., Offermann, S., and Peterhansel, C. (2013). A common histone modification code on C_4 genes in maize and its conservation in sorghum and *Setaria italica*. Plant Physiology **162,** 456-469.

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

Heuer, S., Hansen, S., Bantin, J., Brettschneider, R., Kranz, E., Lorz, H., and Dresselhaus, T. (2001). The maize MADS box gene *ZmMADS3* affects node number and spikelet development and is co-expressed with *ZmMADS1* during flower development, in egg cells, and early embryogenesis. Plant Physiology **127,** 33-45.

Hibberd, J.M., and Quick, W.P. (2002). Characteristics of C_4 photosynthesis in stems and petioles of C_3 flowering plants. Nature **415,** 451-454.

Hibberd, J.M., and Covshoff, S. (2010). The regulation of gene expression required for C_4 photosynthesis. Annual Review of Plant Biology **61,** 181-207.

Hibberd, J.M., Sheehy, J.E., and Langdale, J.A. (2008). Using C_4 photosynthesis to increase the yield of rice - rationale and feasibility. Curr Opin Plant Biol **11,** 228-231.

Hiei, Y., and Komari, T. (2008). *Agrobacterium*-mediated transformation of rice using immature embryos or calli induced from mature seed. Nat Protoc **3,** 824-834.

Hiei, Y., Ohta, S., Komari, T., and Kumashiro, T. (1994). Efficient transformation of rice (*Oryza-sativa L*) mediated by *Agrobacterium* and sequence-analysis of the boundaries of the T-DNA. Plant Journal **6,** 271-282.

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

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17,** 754-755.

Husbands, A.Y., Chitwood, D.H., Plavskin, Y., and Timmermans, M.C.P. (2009). Signals and prepatterns: new insights into organ polarity in plants. Gene Dev **23,** 1986-1997.

Huson, D.H., Richter, D.C., Rausch, C., Dezulian, T., Franz, M., and Rupp, R. (2007). Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics **8,** 460.

Ikeda, A., Ueguchi-Tanaka, M., Sonoda, Y., Kitano, H., Koshioka, M., Futsuhara, Y., Matsuoka, M., and Yamaguchi, J. (2001). *slender*

rice, a constitutive gibberellin response mutant, is caused by a null mutation of the *SLR1* gene, an ortholog of the height-regulating gene *GAI*/*RGA*/*RHT*/*D8*. Plant Cell **13,** 999-1010.

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

Ishikawa, M., Ohmori, Y., Tanaka, W., Hirabayashi, C., Murai, K., Ogihara, Y., Yamaguchi, T., and Hirano, H.Y. (2009). The spatial expression patterns of *DROOPING LEAF* orthologs suggest a conserved function in grasses. Genes Genet Syst **84,** 137-146.

Itoh, J., Nonomura, K., Ikeda, K., Yamaki, S., Inukai, Y., Yamagishi, H., Kitano, H., and Nagato, Y. (2005). Rice plant development: from zygote to spikelet. Plant Cell Physiol **46,** 23-47.

Izhaki, A., and Bowman, J.L. (2007). KANADI and class III HD-zip gene families regulate embryo patterning and modulate auxin flow during embryogenesis in *Arabidopsis*. Plant Cell **19,** 495-508.

Jackson, D., Veit, B., and Hake, S. (1994). Expression of maize *knotted1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. Development **120,** 405-413.

Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyere, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I.,

Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pe, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.F., Weissenbach, J., Quetier, F., and Wincker, P. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449,** 463-467.

Jakoby, M.J., Falkenhan, D., Mader, M.T., Brininstool, G., Wischnitzki, E., Platz, N., Hudson, A., Lskamp, M.H.R., Larkin, J., and Schnittger, A. (2008). Transcriptional profiling of mature *Arabidopsis* trichomes reveals that *NOECK* encodes the *MIXTA*-like transcriptional regulator *MYB106*. Plant Physiology **148,** 1583-1602.

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

Jia, G., Huang, X., Zhi, H., Zhao, Y., Zhao, Q., Li, W., Chai, Y., Yang, L., Liu, K., Lu, H., Zhu, C., Lu, Y., Zhou, C., Fan, D., Weng, Q., Guo, Y., Huang, T., Zhang, L., Lu, T., Feng, Q., Hao, H., Liu, H., Lu, P., Zhang, N., Li, Y., Guo, E., Wang, S., Wang, S., Liu, J., Zhang, W., Chen, G., Zhang, B., Li, W., Wang, Y., Li, H., Zhao, B., Li, J., Diao, X., and Han, B. (2013). A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). Nature Genetics **45,** 957-961.

Jia, Y., Lisch, D.R., Ohtsu, K., Scanlon, M.J., Nettleton, D., and Schnable, P.S. (2009). Loss of RNA-dependent RNA polymerase 2 (*RDR2*) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. PLoS Genetics **5,** e1000737.

Jiang, C.F., Mithani, A., Gan, X.C., Belfield, E.J., Klingler, J.P., Zhu, J.K., Ragoussis, J., Mott, R., and Harberd, N.P. (2011). Regenerant *Arabidopsis* lineages display a distinct genome-wide spectrum of mutations conferring variant phenotypes. Current Biology **21,** 1385-1390.

Juarez, M.T., Twigg, R.W., and Timmermans, M.C. (2004). Specification of adaxial cell fate during maize leaf development. Development **131,** 4533-4544.

Kadereit, G., Borsch, T., Weising, K., and Freitag, H. (2003). Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C_4 photosynthesis. International Journal of Plant Sciences **164,** 959-986.

Kajala, K., Brown, N.J., Williams, B.P., Borrill, P., Taylor, L.E., and Hibberd, J.M. (2012). Multiple *Arabidopsis* genes primed for recruitment into C_4 photosynthesis. Plant Journal **69,** 47-56.

Kajala, K., Covshoff, S., Karki, S., Woodfield, H., Tolley, B.J., Dionora, M.J.A., Mogul, R.T., Mabilangan, A.E., Danila, F.R., Hibberd, J.M., and Quick, W.P. (2011). Strategies for engineering a two-celled C_4 photosynthetic pathway into rice. J Exp Bot **62,** 3001-3010.

Kakizaki, T., Matsumura, H., Nakayama, K., Che, F.S., Terauchi, R., and Inaba, T. (2009). Coordination of plastid protein import and nuclear

gene expression by plastid-to-nucleus retrograde signaling. Plant Physiology **151,** 1339-1353.

Kanai, R., Edwards, G., Sage, R., and Monson, R. (1999). The Biochemistry of C_4 Photosynthesis. In C_4 Plant Biology, R. Sage and R. Monson, eds (San Diego: Academic Press), pp. 49-80.

Kang, J., and Dengler, N. (2004). Vein pattern development in adult leaves of *Arabidopsis thaliana*. International Journal of Plant Sciences **165,** 231-242.

Kanno, T., Mette, M.F., Kreil, D.P., Aufsatz, W., Matzke, M., and Matzke, A.J. (2004). Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. Curr Biol **14,** 801-805.

Karpilov, Y. (1960). The distribution of radioactive carbon 14 amongst the products of photosynthesis in maize. Trudy Kazansk Sel'shokoz Institute **41,** 15-24.

Katoh, K., Kuma, K., Miyata, T., and Toh, H. (2005). Improvement in the accuracy of multiple sequence alignment program MAFFT. Genome Inform **16,** 22-33.

Kellogg, E. (1999). Phylogenetic Aspects of the Evolution of C_4 Photosynthesis. In C_4 Plant Biology, R. Sage and R. Monson, eds (San Diego: Academic Press), pp. 411-444.

Kelly, S., and Maini, P.K. (2013). DendroBLAST: approximate phylogenetic trees in the absence of multiple sequence alignments. PLoS ONE **8,** e58537.

Kelly, S., Wickstead, B., and Gull, K. (2011). Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea

and a thaumarchaeal origin for the eukaryotes. Proc Biol Sci **278,** 1009-1018.

Kerstetter, R.A., Bollman, K., Taylor, R.A., Bomblies, K., and Poethig, R.S. (2001). *KANADI* regulates organ polarity in *Arabidopsis*. Nature **411,** 706-709.

Kobayashi, K., Baba, S., Obayashi, T., Sato, M., Toyooka, K., Keranen, M., Aro, E.M., Fukaki, H., Ohta, H., Sugimoto, K., and Masuda, T. (2012). Regulation of root greening by light and auxin/cytokinin signaling in *Arabidopsis*. Plant Cell **24,** 1081-1095.

Koizumi, K., Hayashi, T., Wu, S., and Gallagher, K.L. (2012). The SHORT-ROOT protein acts as a mobile, dose-dependent signal in patterning the ground tissue. Proceedings of the National Academy of Sciences of the United States of America **109,** 13010-13015.

Kortschak, H., Hartt, C., and Burr, G. (1965). Carbon dioxide fixation in sugar cane leaves. Plant Physiology **40,** 209-213.

Kumar, R., Kushalappa, K., Godt, D., Pidkowich, M.S., Pastorelli, S., Hepworth, S.R., and Haughn, G.W. (2007). The *Arabidopsis BEL1-LIKE HOMEODOMAIN* proteins SAW1 and SAW2 act redundantly to regulate *KNOX* expression spatially in leaf margins. Plant Cell **19,** 2719-2735.

Laetsch, W.M. (1974). C_4 syndrome - Structural analysis. Annu Rev Plant Phys **25,** 27-52.

Lampard, G.R., MacAlister, C.A., and Bergmann, D.C. (2008). *Arabidopsis* stomatal initiation is controlled by MAPK-mediated regulation of the bHLH *SPEECHLESS*. Science **322,** 1113-1116.

Langdale, J. (1994). *In situ* hybridization. In The Maize Handbook, M. Freeling and V. Walbot, eds (Heidelberg: Springer), pp. 165-179.

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

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

Langdale, J., Metzler, M., and Nelson, T. (1987). The *argentia* mutation delays normal development of photosynthetic cell-types in *Zea mays*. Developmental Biology **122,** 243-255.

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

Langdale, J.A. (2011). C_4 cycles: past, present, and future research on C_4 photosynthesis. Plant Cell **23,** 3879-3892.

Langdale, J.A., Rothermel, B., and Nelson, T. (1988a). Cellular pattern of photosynthetic gene expression in developing maize leaves. Gene Dev **2,** 106.

Langdale, J.A., Taylor, W.C., and Nelson, T. (1991). Cell-specific accumulation of maize phospho*enol*pyruvate carboxylase is correlated with demethylation at a specific site greater than 3 kb upstream of the gene. Mol Gen Genet **225,** 49-55.

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

Le, S.Q., and Gascuel, O. (2008). An improved general amino acid replacement matrix. Molecular Biology and Evolution **25,** 1307-1320.

Lee, I., Ambaru, B., Thakkar, P., Marcotte, E.M., and Rhee, S.Y. (2010). Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*. Nat Biotechnol **28,** 149-U114.

Lee, I., Seo, Y.S., Coltrane, D., Hwang, S., Oh, T., Marcotte, E.M., and Ronald, P.C. (2011). Genetic dissection of the biotic stress response using a genome-scale gene network for rice. Proceedings of the National Academy of Sciences of the United States of America **108,** 18548-18553.

Levesque, M.P., Vernoux, T., Busch, W., Cui, H., Wang, J.Y., Blilou, I., Hassan, H., Nakajima, K., Matsumoto, N., Lohmann, J.U., Scheres, B., and Benfey, P.N. (2006). Whole-genome analysis of the *SHORT-ROOT* developmental pathway in *Arabidopsis*. PLoS Biology **4,** e143.

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

Li, H., Johnson, P., Stepanova, A., Alonso, J.M., and Ecker, J.R. (2004). Convergence of signaling pathways in the control of differential cell growth in Arabidopsis. Dev Cell **7,** 193-204.

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

Li, P.H., and Brutnell, T.P. (2011). *Setaria viridis* and *Setaria italica*, model genetic systems for the Panicoid grasses. J Exp Bot **62,** 3031-3037.

Li, R., Yu, C., Li, Y., Lam, T., Yiu, S., Kristiansen, K., and Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics **25,** 1966-1967.

Liljegren, S.J., Roeder, A.H.K., Kempin, S.A., Gremski, K., Ostergaard, L., Guimil, S., Reyes, D.K., and Yanofsky, M.F. (2004). Control of fruit patterning in Arabidopsis by *INDEHISCENT*. Cell **116,** 843-853.

Lim, J., Jung, J.W., Lim, C.E., Lee, M.H., Kim, B.J., Kim, M., Bruce, W.B., and Benfey, P.N. (2005). Conservation and diversification of *SCARECROW* in maize. Plant Mol Biol **59,** 619-630.

Lister, R., Gregory, B.D., and Ecker, J.R. (2009). Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. Curr Opin Plant Biol **12,** 107-118.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. Cell **133,** 523-536.

Liu, P.P., Koizuka, N., Martin, R.C., and Nonogaki, H. (2005). The *BME3* (*Blue Micropylar End 3*) GATA zinc finger transcription factor is a positive regulator of Arabidopsis seed germination. Plant Journal **44,** 960-971.

Liu, T., Ohashi-Ito, K., and Bergmann, D.C. (2009). Orthologs of *Arabidopsis thaliana* stomatal bHLH genes and regulation of stomatal development in grasses. Development **136,** 2265-2276.

Liu, W.Y., Chang, Y.M., Chen, S.C., Lu, C.H., Wu, Y.H., Lu, M.Y., Chen, D.R., Shih, A.C., Sheue, C.R., Huang, H.C., Yu, C.P., Lin, H.H., Shiu, S.H., Sun-Ben Ku, M., and Li, W.H. (2013). Anatomical and transcriptional dynamics of maize embryonic leaves during seed germination. Proceedings of the National Academy of Sciences of the United States of America **110,** 3979-3984.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods **25,** 402-408.

Long, T.A., Brady, S.M., and Benfey, P.N. (2008). Systems approaches to identifying gene regulatory networks in plants. Annu Rev Cell Dev Bi **24,** 81-103.

Lunter, G., and Goodson, M. (2010). Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Research **21,** 936-939.

Majeran, W., Cai, Y., Sun, Q., and van Wijk, K.J. (2005). Functional differentiation of bundle sheath and mesophyll maize chloroplasts determined by comparative proteomics. Plant Cell **17,** 3111-3140.

Majeran, W., Zybailov, B., Ytterberg, A.J., Dunsmore, J., Sun, Q., and Van Wijk, K.J. (2008). Consequences of C_4 differentiation for chloroplast membrane proteomes in maize mesophyll and bundle sheath cells. Molecular & Cellular Proteomics **7,** 1609-1638.

Majeran, W., Friso, G., Ponnala, L., Connolly, B., Huang, M., Reidel, E., Zhang, C., Asakura, Y., Bhuiyan, N.H., Sun, Q., Turgeon, R., and van Wijk, K.J. (2010). Structural and metabolic transitions of C_4 leaf

development and differentiation defined by microscopy and quantitative proteomics in maize. Plant Cell **22,** 3509-3542.

Manfield, I.W., Devlin, P.F., Jen, C.H., Westhead, D.R., and Gilmartin, P.M. (2007). Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the *Arabidopsis* GATA gene family. Plant Physiology **143,** 941-958.

Mann, D.G.J., LaFayette, P.R., Abercrombie, L.L., King, Z.R., Mazarei, M., Halter, M.C., Poovaiah, C.R., Baxter, H., Shen, H., Dixon, R.A., Parrott, W.A., and Stewart, C.N. (2012). Gateway-compatible vectors for high-throughput gene functional analysis in switchgrass (*Panicum virgatum* L.) and other monocot species. Plant Biotechnol J **10,** 226-236.

Mao, C.Z., Ding, W.N., Wu, Y.R., Yu, J., He, X.W., Shou, H.X., and Wu, P. (2007). Overexpression of a NAC-domain protein promotes shoot branching in rice. New Phytologist **176,** 288-298.

Mardis, E.R. (2008). Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics **9,** 387-402.

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

Martin-Trillo, M., and Cubas, P. (2010). TCP genes: a family snapshot ten years later. Trends Plant Sci **15,** 31-39.

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

McKown, A.D., and Dengler, N.G. (2007). Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in *Flaveria* (Asteraceae). American Journal of Botany **94,** 382.

Mckown, A.D., and Dengler, N.G. (2009). Shifts in leaf vein density through accelerated vein formation in C_4 *Flaveria* (Asteraceae). Annals of Botany **104,** 1085-1098.

Meierhoff, K., and Westhoff, P. (1993). Differential biogenesis of Photosystem-II in mesophyll and bundle-sheath cells of monocotyledonous NADP-malic enzyme-type C_4 plants - the nonstoichiometric abundance of the subunits of Photosystem-II in the bundle-sheath chloroplasts and the translational activity of the plastome-encoded genes. Planta **191,** 23-33.

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

Miyao, A., Nakagome, M., Ohnuma, T., Yamagata, H., Kanamori, H., Katayose, Y., Takahashi, A., Matsumoto, T., and Hirochika, H. (2012). Molecular spectrum of somaclonal variation in regenerated rice revealed by whole-genome sequencing. Plant Cell Physiol **53,** 256-264.

Miyao, M., Masumoto, C., Miyazawa, S.I., and Fukayama, H. (2011). Lessons from engineering a single-cell C_4 photosynthetic pathway into rice. J Exp Bot **62,** 3021-3029.

Miyashima, S., Sebastian, J., Lee, J.Y., and Helariutta, Y. (2013). Stem cell function during plant vascular development. EMBO Journal **32,** 178-193.

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

Mockaitis, K., and Estelle, M. (2008). Auxin receptors and plant development: a new signaling paradigm. Annu Rev Cell Dev Bi **24,** 55-80.

Monson, R. (2003). Gene duplication, neofunctionalization, and the evolution of C_4 photosynthesis. International Journal of Plant Sciences **164,** S43-S54.

Monson, R.K., and Moore, B.D. (1989). On the significance of C_3-C_4 intermediate photosynthesis to the evolution of C_4 photosynthesis. Plant Cell Environ **12,** 689-699.

Moreno-Risueno, M.A., Busch, W., and Benfey, P.N. (2010). Omics meet networks - using systems approaches to infer regulatory networks in plants. Curr Opin Plant Biol **13,** 126-131.

Morita, M.T., Sakaguchi, K., Kiyose, S., Taira, K., Kato, T., Nakamura, M., and Tasaka, M. (2006). A C2H2-type zinc finger protein, *SGR5*, is

involved in early events of gravitropism in *Arabidopsis* inflorescence stems. Plant Journal **47,** 619-628.

Mortazavi, A., Williams, B.A., Mccue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods **5,** 621-628.

Muhaidat, R., Sage, T.L., Frohlich, M., Dengler, N.G., and Sage, R.F. (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.

Mukherjee, K., Brocchieri, L., and Burglin, T.R. (2009). A comprehensive classification and evolutionary analysis of plant homeobox genes. Molecular Biology and Evolution **26,** 2775-2794.

Muller, B., and Sheen, J. (2007). Advances in cytokinin signaling. Science **318,** 68-69.

Munster, T., Wingen, L.U., Faigl, W., Werth, S., Saedler, H., and Theissen, G. (2001). Characterization of three *GLOBOSA*-like MADS-box genes from maize: evidence for ancient paralogy in one class of floral homeotic B-function genes of grasses. Gene **262,** 1-13.

Murray, F., Brettell, R., Matthews, P., Bishop, D., and Jacobsen, J. (2004). Comparison of *Agrobacterium*-mediated transformation of four barley cultivars using the *GFP* and *GUS* reporter genes. Plant Cell Rep **22,** 397-402.

Nagano, Y., Furuhashi, H., Inaba, T., and Sasaki, Y. (2001). A novel class of plant-specific zinc-dependent DNA-binding protein that binds to A/T-rich DNA sequences. Nucleic Acids Research **29,** 4097-4105.

Nakajima, K., Sena, G., Nawy, T., and Benfey, P.N. (2001). Intercellular movement of the putative transcription factor SHR in root patterning. Nature **413,** 307-311.

Nakamura, H., Muramatsu, M., Hakata, M., Ueno, O., Nagamura, Y., Hirochika, H., Takano, M., and Ichikawa, H. (2009). Ectopic overexpression of the transcription factor *OsGLK1* induces chloroplast development in non-green rice cells. Plant Cell Physiol **50,** 1933-1949.

Nardmann, J., Zimmermann, R., Durantini, D., Kranz, E., and Werr, W. (2007). *WOX* gene phylogeny in Poaceae: a comparative approach addressing leaf and embryo development. Molecular Biology and Evolution **24,** 2474-2484.

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,** 3039-3048.

Nelson, T., and Dengler, N. (1992). Photosynthetic tissue differentiation in C_4 plants. International Journal of Plant Sciences **153,** S93-S105.

Nelson, T., and Dengler, N. (1997). Leaf vascular pattern formation. Plant Cell **9,** 1121-1135.

Newman, L.J., Perazza, D.E., Juda, L., and Campbell, M.M. (2004). Involvement of the R2R3-MYB, *AtMYB61*, in the ectopic lignification and dark-photomorphogenic components of the *det3* mutant phenotype. Plant Journal **37,** 239-250.

Nimmo, H. (2003). Control of the phosphorylation of phospho*enol*pyruvate carboxylase in higher plants. Archives of Biochemistry and Biophysics **414,** 189-196.

Nishimura, A., Aichi, I., and Matsuoka, M. (2006). A protocol for *Agrobacterium*-mediated transformation in rice. Nat Protoc **1,** 2796-2802.

Offermann, S., Danker, T., Dreymuller, D., Kalamajka, R., Topsch, S., Weyand, K., and Peterhansel, C. (2006). Illumination is necessary and sufficient to induce histone acetylation independent of transcriptional activity at the C_4-specific phospho*enol*pyruvate carboxylase promoter in maize. Plant Physiology **141,** 1078-1088.

Ogasawara, H., Kaimi, R., Colasanti, J., and 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,** 489-499.

Osborne, C.P. (2011). The Geologic History of C_4 Plants. In C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms, A.S. Raghavendra and R. Sage, eds (Dordrecht: Springer), pp. 339-357.

Oshima, Y., Shikata, M., Koyama, T., Ohtsubo, N., Mitsuda, N., and Ohme-Takagi, M. (2013). *MIXTA*-like transcription factors and *WAX INDUCER1*/*SHINE1* coordinately regulate cuticle development in *Arabidopsis* and *Torenia fournieri*. Plant Cell **25,** 1609-1624.

Park, J., Okita, T.W., and Edwards, G.E. (2010). Expression profiling and proteomic analysis of isolated photosynthetic cells of the non-Kranz C_4 species *Bienertia sinuspersici*. Funct Plant Biol **37,** 1-13.

Pastore, J.J., Limpuangthip, A., Yamaguchi, N., Wu, M.F., Sang, Y., Han, S.K., Malaspina, L., Chavdaroff, N., Yamaguchi, A., and Wagner, D.

(2011). LATE MERISTEM IDENTITY2 acts together with LEAFY to activate *APETALA1*. Development **138,** 3189-3198.

Patel, M., Siegel, A., and Berry, J. (2006). Untranslated regions of *FbRbcS1* mRNA mediate bundle sheath cell-specific gene expression in leaves of a C_4 plant. Journal of Biological Chemistry **281,** 25485-25491.

Patel, M., Corey, A., Yin, L.-P., Ali, S., Taylor, W., and Berry, J. (2004). Untranslated regions from C_4 *Amaranth AhRbcs1* mRNAs confer translational enhancement and preferential bundle sheath cell expression in transgenic C_4 *Flaveria bidentis*. Plant Physiology **136,** 3550-3561.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H.B., Wang, X.Y., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L.F., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob-ur-Rahman, Ware, D., Westhoff, P., Mayer, K.F.X., Messing, J., and Rokhsar, D.S. (2009). The *Sorghum bicolor* genome and the diversification of grasses. Nature **457,** 551-556.

Penfield, S., Meissner, R.C., Shoue, D.A., Carpita, N.C., and Bevan, M.W. (2001). *MYB61* is required for mucilage deposition and extrusion in the Arabidopsis seed coat. Plant Cell **13,** 2777-2791.

Pengelly, J.J.L., Kwasny, S., Bala, S., Evans, J.R., Voznesenskaya, E.V., Koteyeva, N.K., Edwards, G.E., Furbank, R.T., and von Caemmerer, S. (2011). Functional analysis of corn husk photosynthesis. Plant Physiology**,** 1-11.

Peret, B., Swarup, K., Ferguson, A., Seth, M., Yang, Y.D., Dhondt, S., James, N., Casimiro, I., Perry, P., Syed, A., Yang, H.B., Reemmer, J., Venison, E., Howells, C., Perez-Amador, M.A., Yun, J.G., Alonso, J., Beemster, G.T.S., Laplaze, L., Murphy, A., Bennett, M.J., Nielsen, E., and Swarup, R. (2012). *AUX*/*LAX* genes encode a family of auxin influx transporters that perform distinct functions during *Arabidopsis* development. Plant Cell **24,** 2874-2885.

Petricka, J.J., Clay, N.K., and Nelson, T.M. (2008). Vein patterning screens and the defectively organized tributaries mutants in *Arabidopsis thaliana*. Plant Journal **56,** 251-263.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research **29,** e45.

Pick, T.R., Brautigam, A., Schluter, U., Denton, A.K., Colmsee, C., Scholz, U., Fahnenstich, H., Pieruschka, R., Rascher, U., Sonnewald, U., and Weber, A.P.M. (2011). Systems analysis of a maize leaf developmental gradient redefines the current C_4 model and provides candidates for regulation. Plant Cell **23,** 4208-4220.

Pires, N., and Dolan, L. (2010). Origin and diversification of basic-helix-loop-helix proteins in plants. Molecular Biology and Evolution **27,** 862-874.

Powell, A.L.T., Nguyen, C.V., Hill, T., Cheng, K.L., Figueroa-Balderas, R., Aktas, H., Ashrafi, H., Pons, C., Fernandez-Munoz, R., Vicente, A.,

Lopez-Baltazar, J., Barry, C.S., Liu, Y.S., Chetelat, R., Granell, A., Van Deynze, A., Giovannoni, J.J., and Bennett, A.B. (2012). *Uniform ripening* encodes a *Golden 2-like* transcription factor regulating tomato fruit chloroplast development. Science **336,** 1711-1715.

Putterill, J., Laurie, R., and Macknight, R. (2004). It's time to flower: the genetic control of flowering time. Bioessays **26,** 363-373.

Rashotte, A.M., Mason, M.G., Hutchison, C.E., Ferreira, F.J., Schaller, G.E., and Kieber, J.J. (2006). A subset of *Arabidopsis* AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway. Proceedings of the National Academy of Sciences of the United States of America **103,** 11081-11085.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin-I, T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S., Yamaguchi, K., Sugano, S., Kohara, Y., Fujiyama, A., Anterola, A., Aoki, S., Ashton, N., Barbazuk, W.B., Barker, E., Bennetzen, J.L., Blankenship, R., Cho, S.H., Dutcher, S.K., Estelle, M., Fawcett, J.A., Gundlach, H., Hanada, K., Heyl, A., Hicks, K.A., Hughes, J., Lohr, M., Mayer, K., Melkozernov, A., Murata, T., Nelson, D.R., Pils, B., Prigge, M., Reiss, B., Renner, T., Rombauts, S., Rushton, P.J., Sanderfoot, A., Schween, G., Shiu, S.H., Stueber, K., Theodoulou, F.L., Tu, H., Van de Peer, Y., Verrier, P.J., Waters, E., Wood, A., Yang, L.X., Cove, D., Cuming, A.C., Hasebe, M., Lucas, S., Mishler, B.D., Reski, R., Grigoriev, I.V., Quatrano, R.S., and Boore, J.L.

(2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science **319,** 64-69.

Reyes, J.C., Muro-Pastor, M.I., and Florencio, F.J. (2004). The GATA family of transcription factors in *Arabidopsis* and rice. Plant Physiology **134,** 1718-1732.

Riechmann, J.L., Heard, J., Martin, G., and Reuber, L. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science **290,** 2105.

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., and Pachter, L. (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biology **12,** R22.

Rossini, L., Cribb, L., Martin, D., and Langdale, J. (2001). The maize *Golden2* gene defines a novel class of transcriptional regulators in plants. Plant Cell **13,** 1231-1244.

Royo, J., Gomez, E., Barrero, C., Muniz, L.M., Sanz, Y., and Hueros, G. (2009). Transcriptional activation of the maize endosperm transfer cell-specific gene *BETL*1 by *ZmMRP-1* is enhanced by two C2H2 zinc finger-containing proteins. Planta **230,** 807-818.

Rutjens, B., Bao, D., van Eck-Stouten, E., Brand, M., Smeekens, S., and Proveniers, M. (2009). Shoot apical meristem function in *Arabidopsis* requires the combined activities of three BEL1-like homeodomain proteins. Plant Journal **58,** 641-654.

Sack, L., and Scoffoni, C. (2013). Leaf venation: structure, function, development, evolution, ecology and applications in the past, present and future. New Phytologist **198,** 983-1000.

Sage, R. (2001). Environmental and evolutionary preconditions for the origin and diversification of the C_4 photosynthetic syndrome. Plant Biology **3,** 202-213.

Sage, R. (2004). The evolution of C_4 photosynthesis. New Phytologist **161,** 341-370.

Sage, R.F., Christin, P.A., and Edwards, E.J. (2011a). The C_4 plant lineages of planet Earth. J Exp Bot **62,** 3155-3169.

Sage, R.F., Sage, T.L., and Kocacinar, F. (2012). Photorespiration and the evolution of C_4 photosynthesis. Annual Review of Plant Biology **63,** 19-47.

Sage, T.L., Sage, R.F., Vogan, P.J., Rahman, B., Johnson, D.C., Oakley, J.C., and Heckel, M.A. (2011b). The occurrence of C*2* photosynthesis in *Euphorbia* subgenus *Chamaesyce* (Euphorbiaceae). J Exp Bot **62,** 3183-3195.

Sakamoto, T., Miura, K., Itoh, H., Tatsumi, T., Ueguchi-Tanaka, M., Ishiyama, K., Kobayashi, M., Agrawal, G.K., Takeda, S., Abe, K., Miyao, A., Hirochika, H., Kitano, H., Ashikari, M., and Matsuoka, M. (2004). An overview of gibberellin metabolism enzyme genes and their related mutants in rice. Plant Physiology **134,** 1642-1653.

Sang, X.C., Li, Y.F., Luo, Z.K., Ren, D.Y., Fang, L.K., Wang, N., Zhao, F.M., Ling, Y.H., Yang, Z.L., Liu, Y.S., and He, G.H. (2012). *CHIMERIC FLORAL ORGANS1*, encoding a monocot-specific MADS box protein, regulates floral organ identity in rice. Plant Physiology **160,** 788-807.

Sato, Y., Antonio, B., Namiki, N., Motoyama, R., Sugimoto, K., Takehisa, H., Minami, H., Kamatsuki, K., Kusaba, M., Hirochika, H., and

Nagamura, Y. (2011). Field transcriptome revealed critical developmental and physiological transitions involved in the expression of growth potential in japonica rice. Bmc Plant Biol **11**.

Satoh-Nagasawa, N., Nagasawa, N., Malcomber, S., Sakai, H., and Jackson, D. (2006). A trehalose metabolic enzyme controls inflorescence architecture in maize. Nature **441,** 227-230.

Scarpella, E., Francis, P., and Berleth, T. (2004). Stage-specific markers define early steps of procambium development in *Arabidopsis* leaves and correlate termination of vein formation with mesophyll differentiation. Development **131,** 3445-3455.

Scarpella, E., Barkoulas, M., and Tsiantis, M. (2010). Control of leaf and vein development by auxin. Cold Spring Harbor Perspectives in Biology **2,** a001511.

Scheres, B., Di Laurenzio, L., Willemsen, V., Hauser, M.T., Janmaat, K., Weisbeek, P., and Benfey, P.N. (1995). Mutations affecting the radial organization of the *Arabidopsis* root display specific defects throughout the embryonic axis. Development **121,** 53-62.

Schlereth, A., Moller, B., Liu, W.L., Kientz, M., Flipse, J., Rademacher, E.H., Schmid, M., Jurgens, G., and Weijers, D. (2010). *MONOPTEROS* controls embryonic root initiation by regulating a mobile transcription factor. Nature **464,** 913-916.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu,

S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.-C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C., and Jackson, S.A. (2010). Genome sequence of the palaeopolyploid soybean. Nature **463,** 178-183.

Schutze, P., Freitag, H., and Weising, K. (2003). An integrated molecular and morphological study of the subfamily Suaedoideae Ulbr. (Chenopodiaceae). Plant Systematics and Evolution **239,** 257-286.

Sharman, B. (1942). Developmental anatomy of the shoot of *Zea mays* L. Annals of Botany **6,** 245-282.

Sheehy, J., Mitchell, P., and Hardy, B. (2007). Charting New Pathways to C_4 Rice. (Los Banos, Philippines: World Scientific).

Sheen, J. (1999). C_4 gene expression. Annu Rev Plant Phys **50,** 187-217.

Sheen, J.Y., and Bogorad, L. (1985). Differential expression of the ribulose bisphosphate carboxylase large subunit gene in bundle sheath and mesophyll cells of developing maize leaves is influenced by light. Plant Physiology **79,** 1072-1076.

Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. Nat Biotechnol **26,** 1135-1145.

Sinha, N.R., Williams, R.E., and Hake, S. (1993). Overexpression of the maize homeobox gene, *knotted-1*, causes a switch from determinate to indeterminate cell fates. Gene Dev **7,** 787-795.

Slewinski, T.L. (2013). Using evolution as a guide to engineer Kranz-type C_4 photosynthesis. Frontiers in plant science **4,** 212.

Slewinski, T.L., Anderson, A.A., Zhang, C., and Turgeon, R. (2012). Scarecrow plays a role in establishing Kranz anatomy in maize leaves. Plant & Cell Physiology **53,** 2030-2037.

Smillie, I.R.A., Pyke, K.A., and Murchie, E.H. (2012). Variation in vein density and mesophyll cell architecture in a rice deletion mutant population. J Exp Bot **63,** 4563-4570.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C.F., Sankoff, D., dePamphilis, C.W., Wall, P.K., and Soltis, P.S. (2009). Polyploidy and angiosperm diversification. American Journal of Botany **96,** 336-348.

Soneson, C., and Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics **14,** 91.

Sorin, C., Salla-Martret, M., Bou-Torrent, J., Roig-Villanova, I., and Martinez-Garcia, J.F. (2009). ATHB4, a regulator of shade avoidance, modulates hormone response in Arabidopsis seedlings. Plant J **59,** 266-277.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22,** 2688-2690.

Sukumaran, J., and Holder, M.T. (2010). DendroPy: a Python library for phylogenetic computing. Bioinformatics **26,** 1569-1571.

Sun, S., Yu, J.P., Chen, F., Zhao, T.J., Fang, X.H., Li, Y.Q., and Sui, S.F. (2008). TINY, a dehydration-responsive element (DRE)-binding protein-like transcription factor connecting the DRE- and ethylene-

responsive element-mediated signaling pathways in *Arabidopsis*. Journal of Biological Chemistry **283,** 6261-6271.

Swigonova, Z., Lai, J.S., Ma, J.X., Ramakrishna, W., Llaca, V., Bennetzen, J.L., and Messing, J. (2004). Close split of sorghum and maize genome progenitors. Genome Research **14,** 1916-1923.

Takacs, E.M., Li, J., Du, C.L., Ponnala, L., Janick-Buckner, D., Yu, J.M., Muehlbauer, G.J., Schnable, P.S., Timmermans, M.C.P., Sun, Q., Nettleton, D., and Scanlon, M.J. (2012). Ontogeny of the maize shoot apical meristem. Plant Cell **24,** 3219-3234.

Tatarinova, T.V., Alexandrov, N.N., Bouck, J.B., and Feldmann, K.A. (2010). GC3 biology in corn, rice, sorghum and other grasses. BMC Genomics **11,** 308.

ten Hove, C.A., Bochdanovits, Z., Jansweijer, V.M.A., Koning, F.G., Berke, L., Sanchez-Perez, G.F., Scheres, B., and Heidstra, R. (2011a). Probing the roles of LRR-RLK genes in *Arabidopsis thaliana* roots using a custom T-DNA insertion set. Plant Mol Biol **76,** 69-83.

ten Hove, C.A., de Jong, M., Lapin, D., Andel, A., Sanchez-Perez, G.F., Tarutani, Y., Suzuki, Y., Heidstra, R., and van den Ackerveken, G. (2011b). Trans-repression of gene activity upstream of T-DNA tagged *RLK902* links *Arabidopsis* root growth inhibition and downy mildew resistance. Plos One **6,** e19028.

Tolley, B.J., Sage, T.L., Langdale, J.A., and Hibberd, J.M. (2012a). Individual maize chromosomes in the C_3 plant oat can increase bundle sheath cell size and vein density. Plant Physiology **159,** 1418-1427.

Tolley, B.J., Woodfield, H., Wanchana, S., Bruskiewich, R., and Hibberd, J.M. (2012b). Light-regulated and cell-specific methylation of the maize *PEPC* promoter. J Exp Bot **63,** 1381-1390.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol **28,** 511-515.

Tsuwamoto, R., Fukuoka, H., and Takahata, Y. (2008). *GASSHO1* and *GASSHO2* encoding a putative leucine-rich repeat transmembrane-type receptor kinase are essential for the normal development of the epidermal surface in *Arabidopsis* embryos. Plant Journal **54,** 30-42.

Tuskan, G.A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.-L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., D √ ©jardin, A., dePamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasj√ § rvi, J., Karlsson, J., Kelleher, C., Kirkpatrick,

R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Lepl√©, J.-C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouz√©, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.-J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer, Y., and Rokhsar, D. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313,** 1596-1604.

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. International Journal of Plant Sciences **156,** 270-289.

Ueno, O., Kawano, Y., Wakayama, M., and Takeda, T. (2006). Leaf vascular systems in C_3 and C_4 grasses: a two-dimensional analysis. Annals of Botany **97,** 611-621.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3: new capabilities and interfaces. Nucleic Acids Research **40,** e115.

Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F.M., Bassel, G.W., Tanimoto, M., Chow, A., Steinhauser, D., Persson, S., and Provart, N.J. (2009). Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. Plant Cell Environ **32,** 1633-1651.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S.K., Troggio, M., Pruss, D., Salvi, S., Pindo, M., Baldi, P., Castelletti, S., Cavaiuolo, M., Coppola, G., Costa, F., Cova, V., Dal Ri, A., Goremykin, V., Komjanc, M., Longhi, S., Magnago, P., Malacarne, G., Malnoy, M., Micheletti, D., Moretto, M., Perazzolli, M., Si-Ammour, A., Vezzulli, S., Zini, E., Eldredge, G., Fitzgerald, L.M., Gutin, N., Lanchbury, J., Macalma, T., Mitchell, J.T., Reid, J., Wardell, B., Kodira, C., Chen, Z., Desany, B., Niazi, F., Palmer, M., Koepke, T., Jiwan, D., Schaeffer, S., Krishnan, V., Wu, C., Chu, V.T., King, S.T., Vick, J., Tao, Q., Mraz, A., Stormo, A., Stormo, K., Bogden, R., Ederle, D., Stella, A., Vecchietti, A., Kater, M.M., Masiero, S., Lasserre, P., Lespinasse, Y., Allan, A.C., Bus, V., Chagne, D., Crowhurst, R.N., Gleave, A.P., Lavezzo, E., Fawcett, J.A., Proost, S., Rouze, P., Sterck, L., Toppo, S., Lazzari, B., Hellens, R.P., Durel, C.E., Gutin, A., Bumgarner, R.E., Gardiner, S.E., Skolnick, M., Egholm, M., Van de Peer, Y., Salamini, F., and Viola, R. (2010). The genome of the domesticated apple (*Malus* x *domestica* Borkh.). Nature Genetics **42,** 833-839.

Verbsky, M.L., and Richards, E.J. (2001). Chromatin remodeling in plants. Curr Opin Plant Biol **4,** 494-500.

Vicentini, A., Barber, J.C., Aliscioni, S.S., Giussani, L.M., and Kellogg, E.A. (2008). The age of the grasses and clusters of origins of C_4 photosynthesis. Global Change Biol **14,** 2963-2977.

Vollbrecht, E., Reiser, L., and Hake, S. (2000). Shoot meristem size is dependent on inbred background and presence of the maize homeobox gene, *knotted1*. Development **127,** 3161-3172.

von Caemmerer, S., Quick, W.P., and Furbank, R.T. (2012). The development of C_4 rice: current progress and future challenges. Science **336,** 1671-1672.

Voznesenskaya, E., Franceschi, V., Kiirats, O., Freitag, H., and Edwards, G. (2001). Kranz anatomy is not essential for terrestrial C_4 plant photosynthesis. Nature **414,** 543-546.

Voznesenskaya, E., Franceschi, V., Kiirats, O., Artyusheva, E., Freitag, H., and Edwards, G. (2002). Proof of C_4 photosynthesis without Kranz anatomy in *Bienertia cycloptera* (Chenopodiaceae). Plant Journal **31,** 649-662.

Voznesenskaya, E., Koteyeva, N., Chuong, S., Akhani, H., Edwards, G., and Francesch, V. (2005a). Differentiation of cellular and biochemical features of the single-cell C_4 syndrome during leaf development in *Bienertia cycloptera* (Chenopodiaceae). American Journal of Botany **92,** 1784-1795.

Voznesenskaya, E.V., Chuong, S.D.X., Koteyeva, N.K., Edwards, G.E., and Franceschi, V.R. (2005b). Functional compartmentation of C_4 photosynthesis in the triple-layered chlorenchyma of *Aristida* (Poaceae). Funct Plant Biol **32,** 67-77.

Voznesenskaya, E.V., Koteyeva, N.K., Chuong, S.D.X., Ivanova, A.N., Barroca, J., Craven, L.A., and Edwards, G.E. (2007). Physiological,

anatomical and biochemical characterisation of photosynthetic types in genus *Cleome* (Cleomaceae). Funct Plant Biol **34,** 247-267.

Wakayama, M., Ueno, O., and Ohnishi, J. (2002). Cellular accumulation of photosynthetic enzymes during leaf development of *Arundinella hirta*, a C_4 grass with unusual Kranz cells without contact with vascular tissues. Plant Cell Physiol **43,** S173-S173.

Wakayama, M., Ueno, O., and Ohnishi, J. (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,** 1330-1340.

Walsh, J., Waters, C.A., and Freeling, M. (1998). The maize gene *liguleless2* encodes a basic leucine zipper protein involved in the establishment of the leaf blade-sheath boundary. Genes Dev **12,** 208-218.

Wang, B.B., and Brendel, V. (2006). Molecular characterization and phylogeny of U2AF(35) homologs in plants. Plant Physiology **140,** 624-636.

Wang, J.L., Turgeon, R., Carr, J.P., and Berry, J.O. (1993). Carbon sink-to-source transition is coordinated with establishment of cell-specific gene-expression in a C_4 plant. Plant Cell **5,** 289-296.

Wang, M.B., Li, Z.Y., Matthews, P.R., and Upadhyaya, N.M. (1998). Improved vectors for *Agrobacterium tumefaciens*-mediated transformation of monocot plants. Acta Horticulturae **461,** 401-407.

Wang, P., Kelly, S., Fouracre, J.P., and Langdale, J.A. (2013a). Genome-wide transcript analysis of early maize leaf development reveals gene

cohorts associated with the differentiation of C_4 Kranz anatomy. Plant Journal **75,** 656-670.

Wang, P., Fouracre, J., Kelly, S., Karki, S., Gowik, U., Aubry, S., Shaw, M.K., Westhoff, P., Slamet-Loedin, I.H., Quick, W.P., Hibberd, J.M., and Langdale, J.A. (2013b). Evolution of *GOLDEN2-LIKE* gene function in C_3 and C_4 plants. Planta **237,** 481-495.

Wang, S., Chang, Y., Guo, J., Zeng, Q., Ellis, B.E., and Chen, J.G. (2011). *Arabidopsis* ovate family proteins, a novel transcriptional repressor family, control multiple aspects of plant growth and development. PLoS ONE **6,** e23896.

Wang, X., Gowik, U., Tang, H., Bowers, J., Westhoff, P., and Paterson, A. (2009). Comparative genomic analysis of C_4 photosynthetic pathway evolution in grasses. Genome Biology **10**.

Wang, X.R., Wu, R., Lin, X.Y., Bai, Y., Song, C.D., Yu, X.M., Xu, C.M., Zhao, N., Dong, Y.Z., and Liu, B. (2013c). Tissue culture-induced genetic and epigenetic alterations in rice pure-lines, F1 hybrids and polyploids. Bmc Plant Biol **13**.

Wang, Y., Deng, D., Bian, Y., Lv, Y., and Xie, Q. (2010). Genome-wide analysis of primary auxin-responsive Aux/IAA gene family in maize (*Zea mays*. L.). Mol Biol Rep **37,** 3991-4001.

Wang, Y., Deng, D., Zhang, R., Wang, S., Bian, Y., and Yin, Z. (2012a). Systematic analysis of plant-specific B3 domain-containing proteins based on the genome resources of 11 sequenced species. Mol Biol Rep **39,** 6267-6282.

Wang, Y.H., Wan, L.Y., Zhang, L.X., Zhang, Z.J., Zhang, H.W., Quan, R.D., Zhou, S.R., and Huang, R.F. (2012b). An ethylene response factor *OsWR1* responsive to drought stress transcriptionally activates wax synthesis related genes and increases wax production in rice. Plant Mol Biol **78,** 275-288.

Ward, J.M., Cufr, C.A., Denzel, M.A., and Neff, M.M. (2005). The Dof transcription factor *OBP3* modulates phytochrome and cryptochrome signaling in *Arabidopsis*. Plant Cell **17,** 475-485.

Waters, M., and Langdale, J. (2009). The making of a chloroplast. EMBO Journal **28,** 2861-2873.

Waters, M., Moylan, E., and Langdale, J. (2008). *GLK* transcription factors regulate chloroplast development in a cell-autonomous manner. Plant Journal **56,** 432-444.

Waters, M., Wang, P., Korkaric, M., Capper, R., Saunders, N., and Langdale, J. (2009). *GLK* transcription factors coordinate expression of the photosynthetic apparatus in *Arabidopsis*. Plant Cell **21,** 1109-1128.

Weber, A.P., Weber, K.L., Carr, K., Wilkerson, C., and Ohlrogge, J.B. (2007). Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. Plant Physiology **144,** 32-42.

Weber, A.P.M., and von Caemmerer, S. (2010). Plastid transport and metabolism of C_3 and C_4 plants - comparative analysis and possible biotechnological exploitation. Curr Opin Plant Biol **13,** 257-265.

Wei, K.F., Chen, J., Chen, Y.F., Wu, L.J., and Xie, D.X. (2012). Molecular phylogenetic and expression analysis of the complete WRKY transcription factor family in maize. DNA Res **19,** 153-164.

Welch, D., Hassan, H., Blilou, I., Immink, R., Heidstra, R., and Scheres, B. (2007). *Arabidopsis* JACKDAW and MAGPIE zinc finger proteins delimit asymmetric cell division and stabilize tissue boundaries by restricting SHORT-ROOT action. Gene Dev **21,** 2196-2204.

Westhoff, P., and Gowik, U. (2010). Evolution of C_4 photosynthesis - looking for the master switch. Plant Physiology **154,** 598-601.

Westhoff, P., Offermannsteinhard, K., Hofer, M., Eskins, K., Oswald, A., and Streubel, M. (1991). Differential accumulation of plastid transcripts encoding Photosystem-II components in the mesophyll and bundle-sheath cells of monocotyledonous NADP-malic enzyme-type C_4 plants. Planta **184,** 377-388.

Whipple, C.J., Kebrom, T.H., Weber, A.L., Yang, F., Hall, D., Meeley, R., Schmidt, R., Doebley, J., Brutnell, T.P., and Jackson, D.P. (2011). *grassy tillers1* promotes apical dominance in maize and responds to shade signals in the grasses. Proceedings of the National Academy of Sciences of the United States of America **108,** 506-512.

Williams, B.P., Aubry, S., and Hibberd, J.M. (2012). Molecular evolution of genes recruited into C_4 photosynthesis. Trends Plant Sci **17,** 213-220.

Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Pruss, M., Reuter, I., and Schacherer, F. (2000). TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Research **28,** 316-319.

Wolfe, K.H. (2001). Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet **2,** 333-341.

Wu, G., and Poethig, R.S. (2006). Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target *SPL3*. Development **133,** 3539-3547.

Wysocka-Diller, J.W., Helariutta, Y., Fukaki, H., Malamy, J.E., and Benfey, P.N. (2000). Molecular analysis of *SCARECROW* function reveals a radial patterning mechanism common to root and shoot. Development **127,** 595-603.

Xing, H., Pudake, R.N., Guo, G., Xing, G., Hu, Z., Zhang, Y., Sun, Q., and Ni, Z. (2011). Genome-wide identification and expression profiling of auxin response factor (ARF) gene family in maize. BMC Genomics **12,** 178.

Xing, Q., Creff, A., Waters, A., Tanaka, H., Goodrich, J., and Ingram, G.C. (2013). *ZHOUPI* controls embryonic cuticle formation via a signalling pathway involving the subtilisin protease *ABNORMAL LEAF-SHAPE1* and the receptor kinases *GASSHO1* and *GASSHO2*. Development **140,** 770-779.

Xu, J., Li, Y., Ma, X., Ding, J., Wang, K., Wang, S., Tian, Y., Zhang, H., and Zhu, X.G. (2013). Whole transcriptome analysis using next-generation sequencing of model species *Setaria viridis* to support C_4 photosynthesis research. Plant Mol Biol **83,** 77-87.

Yamaguchi, T., Nagasawa, N., Kawasaki, S., Matsuoka, M., Nagato, Y., and Hirano, H.Y. (2004). The YABBY gene *DROOPING LEAF*

regulates carpel specification and midrib development in *Oryza sativa*. Plant Cell **16,** 500-509.

Yanagisawa, S. (2000). Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. Plant Journal **21,** 281-288.

Yanagisawa, S. (2002). The Dof family of plant transcription factors. Trends Plant Sci **7,** 555-560.

Yanagisawa, S., and Sheen, J. (1998). Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. Plant Cell **10,** 75-89.

Yasumura, Y., Moylan, E., and Langdale, J. (2005). A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. Plant Cell **17,** 1894-1907.

Yonekura-Sakakibara, K., Kojima, M., Yamaya, T., and Sakakibara, H. (2004). Molecular characterization of cytokinin-responsive histidine kinases in maize. Differential ligand preferences and response to cis-zeatin. Plant Physiology **134,** 1654-1661.

Zhang, G.H., Xu, Q., Zhu, X.D., Qian, Q., and Xue, H.W. (2009). *SHALLOT-LIKE1* is a KANADI transcription factor that modulates rice leaf rolling by regulating leaf abaxial cell development. Plant Cell **21,** 719-735.

Zhao, C.S., Avci, U., Grant, E.H., Haigler, C.H., and Beers, E.P. (2008). *XND1*, a member of the NAC domain family in *Arabidopsis thaliana*, negatively regulates lignocellulose synthesis and programmed cell death in xylem. Plant Journal **53,** 425-436.

Zhu, T.T., Nevo, E., Sun, D.F., and Peng, J.H. (2012). Phylogenetic analyses unravel the evolutionary history of NAC proteins in plants. Evolution **66,** 1833-1848.

Zhu, X.G., Long, S.P., and Ort, D.R. (2008). What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? Curr Opin Biotech **19,** 153-159.

Zimmermann, R., and Werr, W. (2005). Pattern formation in the monocot embryo as revealed by *NAM* and *CUC3* orthologues from *Zea mays* L. Plant Mol Biol **58,** 669-685.

Appendix PSCC: Compiled Plant Science Cricket Club Correspondence 2011-2013

PSCC vs OUP, 05/05/11

Mendel, Darwin, Hooker, Borlaug, Leaver: Plant scientists are winners. After only winning a single game last year, however, success was in short supply for PSCC. After a couple of net sessions and a recent decline in the fortune of printed literature, confidence was high ahead of yesterday's friendly match against Oxford University Press. On turning up late after the departmental seminar, many of the team's chief worry was not the overcast conditions or dodgy looking pitch, but just how pervasive transcription could be in plant genomes. The paradigm shift was clearly affecting Ricki Smith and Matt Hodges who fell early but David Boshier (works on forests) and Jim Fouracre (doesn't understand paradigms) looked untroubled and steadied the ship, setting PSCC up for a strong finish. Boshier went on to top score with an impressive 31 not out and was ably supported by Ed Ronan (on loan from the Medwar - how long til he realises that you can immunise the world all you want but they still need feeding), debutant Jon Lamb and Sourav Datta. PSCC finished on 121-9 off 20 overs.

Datta opened the bowling with pace and bounce and formed a powerful opening attack with Hodges at the other end. Hodges's bowling was as accurate as it was deceptive and featured the direction his DPhil so badly needs. He was pick of the bowlers with 3 wickets. Ronan was a close second

with Lamb and Jamie Moore also keeping things tight. The game ebbed and flowed but flowed a little too quickly as Jasper Johnson came on to bowl to leave a very tense finish in very low light. PSCC clung on to win by 5 runs though in, if last season is anything to go by, probably our only victory of the year. Let's savour it. Special mention to cricket debutants Luis Santos and Liz Birch who both fielded outstandingly in what was a fine all round performance.

Men of the match: D. Boshier, OUP wicketkeeper.

PSCC vs Osler/Greent Templeton, 14/05/11

After victory against the largest university press in the world (and hence likely strongest university press cricket team that exists) PSCC approached their game with Osler (medical society)/Green Templeton college (minor college) with a sense of optimism not seen in the department since the advent of Sanger sequencing. After winning the toss and electing to bat any residual doubts vanished as Ricki Smith and David Boshier plundered 14 off the first over, setting the tone for the Plants innings. This aggressive start was continued by Jim Fouracre and Joe Forrester (tenuous links to plant sciences 1) surname, 2) occasional tending of vegetable patch) who went to town on the Osler bowling among some atrocious medic sledging. As if the NHS doesn't have enough problems. Fouracre retired on a maximum 35 not out and Forrester, on his debut in MPLS division colours, cruised to 30 retired. The little master Shyam Masakapalli helped himself to a quick fire 20 as the scurrying doctors reassessed their assumptions on human weight:power

ratios. The run rate slowed a little towards the end of the innings but after a chaotic last over (boundaries, wickets, a bemused looking Luis Santos) Plants finished on 154 all out.

Plants have never scored so many runs before and took to the field, in front of a balcony packed by a stag do and the ecology group, full of confidence. There were excited mutterings of a first competitive victory, tempered by looks of supreme determination. This was going to be the start of a glorious cup run. And then we got mauled. Mauled almost without remorse. If these doctors were at work they would have called in social services. 3 of the top 4 Osler batsmen were bigger and hairier than bears and the other was unfortunately really good at batting. Despite valiant attempts by the Plants bowlers most of the innings was spent retrieving balls from bushes, which in fairness we became quite adept at. The onslaught decreased late in the innings as the triple spin attack of John Russell, Mark Wheeler and Smith finally brought some wickets with their flight and guile, the last of which was the product of a fine slip catch by Santos. Now Boshier, who has been playing Jack Cox cricket since the department was still the Department of Agriculture, had only just commented on how he had never seen a slip catch in Jack Cox so this was a special moment for all. Although not perhaps as special as, in response to his teammates jubilation, Luis's question of "have we won the match?" merited. Osler eventually won by 7 wickets, with a handful of overs to spare. Despite the nature of the loss it was a highly enjoyable game played in very good spirit by both teams.

Man of the Match: Luis Santos. Intrepid running, fine catching, impressive building of hay piles in the outfield.

PSCC vs Chemistry, 23/05/11

The Oxford University Chemistry Department occupies 3 major laboratory buildings (with over £4 million committed to a fourth) and employs over 80 academic members of staff. It is difficult to quantify exactly how much of the £80 million that they have raised in spin out money has been dedicated to training their cricket team, but there have been rumours within MPLS of winter breaks to the subcontinent. In comparison, although bigger in reputation, impact and international standing, the Oxford University Plant Sciences Department is more of a research cottage industry in terms of size and number. And so Plants took to the field against Chemistry on Friday with hopes of what would be one of the biggest upsets in departmental cricket since David Boshier's last team barbeque. Chemistry won the toss and elected to bat with the arrogance that comes naturally to those who work with metal nanoclusters. Two overs of raw pace and two wickets from Sourav Datta later however and the chemists were rushing for their helmets. The upset was on. Briefly. Unfortunately, it was soon to be Plants in need of protection as we started heading for those all too familiar bushes. And the roads beyond the bushes. Joe Forrester always asks questions with his bowling and finally found out whether the windows in the Medwar building can stand a cricket ball (yes) after several near misses. Jamie Moore let it rain on the squash courts and only the permanently capped little master Shyam

Masakapalli managed to hold them down for a while. After the smoke cleared Chemistry had finished on 217 off their 20 overs. For those that understand cricket – indeed. For those that don't – Plants were magnificent.

With such a mammoth target to chase Plants set about playing their natural game: Ricki Smith a wall of textbook poses, Matt Hodges a sporting reincarnation of that bloke from the BT adverts/My Family. The trip to watch Eurovision the previous weekend had clearly paid dividends for Hodges who looked completely at ease, retiring on 30 not out. Forrester looked similarly at ease until Jim Fouracre ran him, and then himself, out to leave Boshier and John Russell to guide the team home, gallivanting to 120-3, with Russell particularly impressive. We hit our target of losing by less than 100 runs emphatically, Chemistry looking downtrodden in victory. While conceding more runs and losing by the biggest margin of the tournament so far may have crushed other teams, Plants left the field in strong spirits, knowing that we'd really achieved something special. Team spirits further escalated following sympathy beers from the chemists who, despite being chemists, exhibited surprisingly well developed social skills.

Man of the match: Luis Santos, flying the PSCC flag high in Barcelona.

<u>PSCC vs Earth Sciences/Pharmacology, 14/06/11</u>

To put this game in a sporting context, both PSCC and Earth Sciences/Pharmacology (ES/P) had played two lost two coming into this, the

last fixture of the Jack Cox group stages, with the outcome to count towards the 'plate' competition. Both teams therefore were desperately in need of a rare win. To put this game in a personal context, prior to the season the ES/P captain, an acquaintance of mine (i wouldn't go as far as to say friend) contacted me asking if PS would like to join forces with ES to form some kind of Gaia superteam. I politely replied saying yes if he could guarantee us a minimum number of places. He never replied. I don't care who you are, or how easily you beat me at squash, but nobody dances with a plant scientist and then jumps into a pharmacologist's bed. To put this game in a meteorological context, the weather was fair with a light easterly.

PSCC won the toss and opted to bat, with veterans David Boshier and Ricki Smith opening the innings like they have so many times before: calmly, gracefully and in no great hurry. They built a foundation any earth scientist would be proud of, with Boshier scoring a comfortable 15 and Smith retiring on 30, finishing with a sublime pair of boundaries. The first 10 overs had brought 53 runs, the next 10 were to bring 106. PSCC rampaged through the ES/P change bowlers, with no one rampaging harder than Matt 'the engine room' Hodges. He pilfered 30 (not out) in 4 overs, despite his engine burning out in the 3rd. Jim Fouracre ground his way to 32 retired and with able and athletic support down the order PSCC finished on a season's (and possibly departmental) best of 159-3 off 20 overs.

Sourav Datta opened the bowling for PSCC with characteristic pace and aggression and thought he had their captain first ball, well caught by Jamie

Moore, only to be denied by a late call of no ball from a shifty looking pharmacologist. This is exactly the kind of thing that happens if you go running off with just any old department. This only inspired Datta to run in harder, with Hodges running in a bit less hard from the other end. Despite tight bowling, ES/P were starting to accelerate worryingly, until Hodges, clearly in need of a sit down, took an outstanding diving catch off Jon Lamb's bowling. Lamb bowled excellently in combination with the little master Shyam Masakapalli. ES/P recovered though and started to up the pace, exactly matching the required run rate. Another hard fought loss looked on the cards. Once the two major ES/P threats had bludgeoned their way to retirement however the game swung very much back in PSCC's favour. Joe Forrester came into the attack bowling the best he has all season, sending down nothing but questions, rising up off a length, and ES/P started to run out of answers. Fault lines started to appear. If the ES/P innings were a drug trial it would have been halted at phase two. With ES/P flagging the twin spin wizards of Smith and John Russell came on to tighten the screw, taking two wickets each. ES/P finished 22 runs short to much celebration by PSCC, having secured a first competitive victory in well over 2 years. Bring on the plate.

Men of the match: Chris Dixon - complete professionalism; Luis Santos - the only player the crowd wants to watch (playing or otherwise); Ricki Smith - foundation building/pastoral care.

PSCC vs Pathology, 22/07/11

Not since Arabidopsis research was just a glint in Elliot Meyerowitz eye has PSCC made it to the Jack Cox memorial trophy plate competition. Victory in Friday's match against Pathology would guarantee our place in this most heralded of semis. Despite the presence of 'Big Clive' in the Pathology line up, renowned the length of South Parks Road as being that bloke who hit the ball over pharmacology, they had not won a game all season and so PSCC entered the contest with high hopes and unfamiliar confidence. Unfortunately Big Clive's fingers are as nimble as his hunger is vast and Pathology won an important toss, electing to field.

With the twin pillars of Smith and Boshier away it fell to Matt Hodges and Jim Fouracre to open the batting. Fouracre hit a quick fire 25 before being caught on the boundary while part time pathologist Hodges batted with the lethargy familiar to those that have watched him go about writing up but, despite only firing on one cylinder, the engine room made it to 30 retired and collapsed on the boundary. Perhaps it was because it was his birthday, perhaps it was because Fouracre was no longer in to run him out, but Joe Forrester batted with supreme ease and zest, cruising to 30 retired. PSCC were in a rampant mood and all over Pathology like a trypanosome outbreak, with the CSI wannabes running out of bowling options. With Big Clive fretting, albeit fretting very slowly, Ed Ronan belied his lack of botanical knowledge and proceeded to make hay before being trapped LBW on 25. Following late cameos by Jack Wright and Shyam Masakapalli PSCC finished on 152-2 - a pretty worrying diagnosis for the pathologists. It just goes to show that, aside from attracting

world leading researchers and providing a congenial working environment, it doesn't matter how many new buildings with fancy cladding and lovely outdoor seating areas you build if you haven't got any bowlers to sit there.

Aside from Big Clive's forearms, the only concern for PSCC was the weather, with dark clouds looming overhead. Both concerns turned out to be real, as Big Clive lived up to his reputation in the early overs and drizzle started to come down. Both sides played on valiantly in worsening conditions but when the drizzle became torrential the game had to be abandoned after only 5 overs when neatly poised. With the strongest bowling line up we've had all season however there is no doubt it would have been poised in PSCC's favour by the end. While we may have been cruelly denied a result here the semi final dream still flickers with one more fixture against Engineering still to play in the group stage.

Men of the match: Luis Santos - really nice haircut; Ian Moore - strong support; Joe Forrester – birthday

PSCC vs Physics, 17/08/11

On Wednesday PSCC took on the Physics department for a place in the Jack Cox inter-departmental cricket competition final (plate version), a place that no botanist has ever been. But, despite our miraculous topping of the qualifying group, this would be no easy feat. Physics, with their superior numbers and understanding of the fundamental laws of motion, are giants of the Jack Cox

trophy. However this was not the Physics team of old, gone are the glory days of Hubble and Ryle. With the department out of the trophy competition and at war with itself over the architectural style of a proposed new building, star gazing was out, and navel gazing was in.

PSCC won the toss and elected to bat in conditions that even the best climate models wouldn't predict. Ricki Smith, refreshed by a holiday, and Matt Hodges, permanently refreshed, opened the batting against some accurate Physics bowling. Any doubts over Smith's mindset following his recent acquisition of a new puppy were quickly laid to rest as he and Hodges started emphatically. Whether or not Hodges's thoughts turned to puppies or not is unclear, but they turned to something as he was bowled on 17 attempting a huge slog over mid wicket having looked very comfortable. Smith kept comfortable and was unlucky to be caught behind on 24 having started to open up. Despite the entry of an octogenarian into the bowling attack Jim Fouracre and Joe Forrester were unperturbed and built on the good start batting freely, almost expansively, until Fouracre was caught on 29 having been distracted by an overexcited Jon Lamb on the boundary and Forrester bowled on 14 attempting to up the run rate. This aggression was maintained in glorious fashion by Sourav Datta, Shyam Masakapalli and John Russell. Datta was outstanding, hitting his first 30 retired of the season, with characteristic speed between the wickets. Nobody was more up for victory than Masakapalli and he batted with even more intent than usual, hitting a rapacious 12, before Russell made the cameo of the season hitting 12 of his

first 4 balls. PSCC finished on 161-5 in what was the best team batting performance of the season.

162 was a competitive target in difficult conditions and PSCC took to the field earnestly, knowing we carried the hopes and dreams of a department that has tolerated drought for too long. Forrester opened the bowling in emphatic fashion, taking a wicket with his 4th ball. Unbelievable scenes. If there were a physicist feeling any greater pressure than the incoming batsman they might actually find the Higgs boson. Datta kept up the ante from the other end, playing out some sweet chin music to the rattled physicists. After such a strong start however Physics found a footing and started to gain momentum, in part due to a Hodges over that pushed the boundaries of the space-time continuum in every direction. PSCC stayed focussed though and secured a double breakthrough thanks to Lamb who was swinging the ball like a sinusoidal wave. A wicket and a sharp run out kept the game neatly poised, it was too close to call. But Physics recovered and, despite determined bowling from Jamie Moore, Smith and on loan zoologist Nathan Kenny, started to edge in front. And then stride. Even Masakapalli's rousing cries from the deep couldn't stem the flow of runs. Physics ran out eventual deserved winners by 7 wickets. PSCC left the field disheartened but united. To have come so far this season only to fall at the last hurdle is tough, crushing even, and we will have to make it through the months of intense introspection and questioning that will follow, but to have come anywhere at all this season is an achievement. To celebrate we will be watching the plate final at the Uni Club on the 30th August before heading out for a meal. All are welcome to join.

Men of the match: Sourav Datta - runs and passion; Mark Wheeler - well informed sledging; Luis Santos - what a season we've had.

PSCC vs OUCS, 30/05/12

After a long winter of gym work and sports psychiatry, nutritional plans and publicity work, the Plant Sciences Cricket Club (PSCC) kicked off the new cricket season with excitement and inappropriate sporting metaphors against giants of the Jack Cox trophy, Oxford University Computer Services, yesterday evening. But with physical growth had come emotional loss. Gone to the world of non-photosynthetic labour were seasoned veterans Johnson (commercial consultancy), Moore (social statistics) and Hodges (full time carer of new family kittens Bella and Sienna). In their honour, a traditional pre-game minute's silence was replaced by practising running each other out and the fetching of balls from bushes.

PSCC won the toss and elected to bat, sending the computer servicemen into a still damp outfield without a hard disk fan in sight. The evergreen Ricki Smith opened up with Jack Cox veteran, but new plant science recruit, Tom Wrobel from our neighbours in the Medieval Latin Dictionary project. Salve Tom. Both showed their experience in dealing with some rapid opening bowling from an aggressive OUCS attack reaching upwards of 10mbps. Smith was unfortunate to fall LBW just as he looked settled bringing Jim Fouracre to the crease, who didn't look settled at all. Wrobel however started to open up

and made the most of some over zealous backing up by the OUCS fielders to cruise to a debut 30 retired. He was well supported by another debutant Sam Satish (Life Technologies), displaying the power of industry by breaking a bat, before the twin thunder of the PSCC Ultras and the little master Shyam Masakapalli arrived. With vuvuzelas booming from the sidelines and Masakapalli booming in the middle PSCC raced onwards, leaving OUCS in meltdown. Masakapalli retired on 30 but late cameos from Ohhhh Johnny Russell, Andries Van Tonder (Zoology) and Jon Lamb helped PSCC to a competitive 132 from our 20 overs.

Lamb managed to save enough energy to open the bowling with zest having gained a yard of pace and a lot of little brothers and sisters in the springtime. He was ably supported by Satish from the other end and initially OUCS were kept at bay. Things looked started to look ominous though as the OUCS opening pair Big Joe and the even Bigger Chris started to find their feet, with Big Joe soon retiring. Just as the game looked to be running away from us however, and despite having spent the last 3 days fighting off allegations of fraud against his employer, Mark 'the (legal) dealer' Wheeler produced a moment of brilliance in the field to run out the OUCS number 3. The OUCS mainframe started shaking even further with Fouracre taking a rare wicket and the Ultras in full voice regardless of whatever was actually going on in the game. Exceptional fielding from debutants Brendan O'Leary and cricketing sage Niloufer Irani (quickest PSCC debut at 8 working days), tight bowling from Van Tonder, Russell and Smith, a wicket first ball from Wheeler and

another run out from Fouracre kept PSCC in the game but in the end OUCS proved too strong, winning with a couple of overs to spare.

A really enjoyable start to the season, thanks particularly to all the fans that came down (covering all 3 departmental research programmes).

MOM: Masakapalli (outstanding innings), Wrobel (batting and wicketkeeping), Luis Santos (out of sight, never out of mind).

***PSCC are holding a cake sale at 10.30 in the coffee room on Wednesday 6th June to raise money for, among other things, a new bat (although i think Life Technologies can probably pay for this) and some abdominal guards. Do your bit to protect the next generation of botanists. ***

PSCC vs Stats/Medawar, 13/06/12

PSCC took on the conjoined might of Statistics and Medawar (S&M), semi-finalists last year, yesterday afternoon in front of a Club packed with fans of Northwest European football if not amateur cricket. But who better to take on while proudly sporting the new studded leather jockstraps that the cake sales over the last couple of weeks have allowed us to buy. Buoyed with the confidence of superior abdominal protection and the knowledge that brothers and sisters in the department had laid down an entire Banbury Coop's worth of of flour and eggs for the team, PSCC elected to bat in refreshingly dry conditions.

Thanks to a suggestion by tactical guru Mark Wheeler, Ohhh Johny Russell was promoted to open the batting with old master and still current head of pastoral care Ricki Smith, who arrived at the crease fresh from a year 9 enrichment session. Smith batted with the ease of a man who knows he's got a 2 month personal enrichment session coming up with term nearly over and he and Russell comfortably saw off the opening overs. Just as Smith was starting to open up he fell to a catch in the covers, shortly followed by Jim Fouracre, frustrated by some generally linear bowling. This brought Joe Forrester and his 2nd holiday beard of the summer to the crease. Unfortunately for S&M, Forrester is a man with a masters level qualification in statistics and he quickly found holes in the normally distributed field. He and the elegant Russell kept the scoreboard ticking over nicely before both retired on 33 and 30 respectively. Following their departure however runs became hard to come by as the bottom tail struggled with some missing parameters. Brave running and a splendid cameo from first time batsman Brendan O'Leary, despite attempts to throw his bat and slide into the crease, left the Plants total at 126 off our allotted 20 overs.

Considering the average (mean) age of the S&M attack 126 felt slightly light but with a slow outfield and Wheeler's considerable research of stat based sledging PSCC took to the field with vigour. And no one more so than debutant Tom Broughton, who is the fastest bowler PSCC have had since the forestry masters shut down, indeed sending our own Forrester a few yards back at wicketkeeper. Broughton and fellow pace merchant Sourav Datta

bowled an incredibly tight first 6 overs, with Broughton only going for 8 runs off his 3 overs. This trend was continued by Jon Lamb who, with a subtle Bonferroni correction to his action, bowled his best 3 overs for PSCC, taking 1 for 11. S&M started to look frightened, but at the same time just a little bit excited, as the runs continued to dry up. Despite their mammoth innings Ohhh Johnny Russell kept one end on absolute lock down while Forrester probed delicately from the other. The situation could still have gone either way, with S&M looking to hit out, until O'Leary, running in from the boundary and half blinded by the sun took a stunning catch in the deep. Not to be outdone, debutant Michiel 'the windmill' Kwantes, who had spent most of the innings trying to work out what the Dutch score was from the cheers inside the Club, took an equally stunning catch the over after. S&M finally submitted and were finished off in the final overs by Smith and Shyam Masakapalli, PSCC winning against all the odds by 20 runs. It was an outstanding bowling and fielding performance, celebrated afterwards with the entire German population of Oxford. Thanks very much to the Ultras the came down and also to everyone who bought and baked cakes, we raised over £200 to put towards equipment. Unbelievable.

MOM: Forrester and Russell (fine all round performances), PSCC bakers (such moist sponges), O'Leary and Kwantes (inspired fielding), Luis Santos (the only icing missing from the cake)

<u>PSCC vs Pathology, 26/06/12</u>

Having snaffled Dr. Steve Kelly from under their noses and following a recent saga involving electron microscopy that turned even the normally self-disciplined Dr. Peng Wang to drink, relations between the Pathology and Plant Sciences departments have never been so bad. They might have a councillor for Jericho and Osney on their payroll but we've got the guardians of the natural world over here. With last season's fixture abandoned due to rain with PSCC in a dominant position, there were scores to settle on and off the pitch, and no man covers ground off an on a pitch combined like the the Pathology captain Big Clive. A man, a myth, a bedtime story saved only for the worst behaved of MPLS division offspring, he won an important toss for the pathologists, his glistening paws held aloft triumphant.

PSCC were thus sent into the field, but with an almost unchanged bowling attack to that which had so ruthlessly contained Stats two weeks ago. Tom Broughton and Sourav Datta thundered in despite the slippery conditions, bowling tightly and menacingly and restricting the wannabee CSI agents in the early overs. There cannot be many things more professionally demeaning than choosing a field of research solely so you can better interpret a low rent drama shown on Channel 5. But the Pathology openers weathered the early storm and started to expand. Datta, asked if he wanted his final over immediately or later, chose the former so he could finish off some work on the confocal. His academic endeavour was to pay dividends as the very next ball he took PSCC's first wicket, Jim Fouracre taking what was described by onlookers as possibly the greatest catch ever taken in the Jack Cox group stages. 3 balls later the pair were to combine again to remove the other

opener, Fouracre taking another catch off fine Datta bowling. However Pathology, left reeling by two quick wickets, were quickly stabilised by the arrival of some extra ballast in the shape of Big Clive. The Pathology captain played with power and surprising poise to steady the ship in the face of some demanding bowling by Jon Lamb (of the batsmen) and Joe Forrester (of the returning David Boshier behind the stumps). With BC soon hitting out and the game moving away from us, Andries Van Tonder bowled extremely tightly to lock it down from one end taking a deserved wicket, caught athletically by fielding find of the season Brendan O'Leary. The Saskatchewan softball league's loss is PSCC's gain. Mark Wheeler came on to turn and burn at the other end and took a wicket with his 2nd ball, stumped dynamically by recently nationally certified athletics coach Boshier. Wheeler's wicket prompted a display of defiance, gleefully withdrawing a warm and crumpled piece of paper from his right pocket annotated with a concise patois critique of the captain for giving him so few bowling opportunities this season. Wheeler did not bowl another over. These two wickets and tight closing overs from Ohhhh Johnny Russell, Forrester and Broughton left Pathology on 118 off their 20 overs.

Last match's successful opening partnership of Ricki Smith and O.J.R. opened up again for Plants in the face of some fierce Pathology bowling. A barrage of short pitched deliveries from both ends led to even the experienced Smith calling for his helmet as PSCC struggled for runs in the opening overs. Russell was unlucky to fall to one aimed at his ribs while Smith demonstrated exactly why he is such a respected Head of Pastoral Care throughout

Oxfordshire, and even parts of Bucks, walking after being caught behind despite being given not out. Fouracre quickly came and went, leaving Forrester and Broughton to threaten the Pathology field like a poorly controlled trypanosome outbreak. The pair ticked along nicely, just keeping on top off the run rate, until a couple of slow overs led to PSCC looking like falling when the finish line was so close in sight. Forrester retired for a 2nd consecutive 30 leaving PSCC needing 13 off the last 12 bowls to win. With O'Leary so worried he had to come down from the bar things looked bleak. And then kappow. Broughton, the Banbury housewives favourite, creamed a couple of successive boundaries to retire, leaving Boshier and Datta to score the winning runs. A nervy finish to give PSCC 2 wins in a row, something even Boshier can't remember happening, thanks in part to our new cake-funded kit bag which is on display by the North building entrance. Our new bat, and some heavy linseed oil aromas, are currently filling N114 for those that enjoy English willow and organically induced headaches.

MOM: Broughton (all round performance), Van Tonder (tight bowling), O'Leary (the new Santos)

PSCC vs Bodleian and Engineering

Against all the odds PSCC won 2 games in 3 days last week, overcoming the mighty and well indexed Engineering and Bodleian teams respectively. With several regulars away new caps were handed out to Cust (economics), O'Toole (earth sciences), Meddler (logistics provision) and Tapper (pufferfish

studies) who all filled in valiantly. O'Toole's first contribution may have been to hurl a ball into wicketkeeper Forrester's face against Engineering, removing him from the rest of the innings and requiring subsequent trips to a dentist and radiology, but he later made up for it with 2 wickets and a 30. Forrester returned in the 2nd innings to score a match winning 30 despite the use of only 1 eye. Engineering also saw the final game in a glorious PSCC career for Shyam Masakapalli, scoring a 4 with the first ball faced by the new cake funded bat, and bowling with characteristic enthusiasm (for the romantics a photo of Shyam's last game is attached). Despite only having 10 men against the Bodleian we won with an over to spare, thanks to 2 wickets apiece from Broughton, Forrester, Wheeler (such guile) and returning comrade Duncan, whose bowling was as tight as his 5 year old University basketball vest, alongside 30s from Forrester and Fouracre.

Our next fixture is against University Offices on Tuesday, please could all those available to play let me know. Incredibly we are already guaranteed a Plate semi-final spot but victory will take us into the Jack Cox Memorial Cup semi finals for the first time in a generation. Needless to say calls to rope in some Chinese badminton players and go for the Plate have been dismissed.

PSCC vs Biochemistry, 21/08/12

The Oxford University Biochemistry department was one of only two departments to finish higher than the Oxford University Plant Sciences department in the 2008 RAE, biological sciences category. How much

emphasis the RAE placed on impressively designed glass fronted buildings and canteen food mark ups, or whether mucking about with derived cell lines really is 4*, is unclear. One thing is for certain though, it all counts for nothing when there's a place in the Jack Cox Plate Competition Final (JCPCF) up for grabs.

Having lost our previous fixture in the pouring rain against a well organised University Offices side, PSCC took on Jack Cox giants Biochemistry hoping for blue skies and a return to winning ways. With only 9 1/2 players and thick cloud things looked ominous. Things looked even worse as, after being put into bat, PSCC lost their 3 leading run scorers in O'Toole, Fouracre and Forrester cheaply in the first few overs against a sharp Biochemistry attack. This left Ricki Smith and Sourav Datta to steady a ship in worsening conditions. Despite unsubstantiated claims to have hit a former Test bowler for 4 earlier in the season, Smith hasn't made 30 all summer and Datta was playing almost one handed, having damaged his wrist in a previous game. But while form is temporary, class and cartilage in at least one hand is permanent. Smith played commandingly to retire on 30, hitting 3 fours in a row at one point, while Datta played with grace and panache to retire shortly after. With Kalde and Schuler in full voice on the balcony PSCC looked to press on but, as the rain poured, batting became increasingly difficult. Van Tonder, Canon O'Leary, Tapper and Wheeler all contributed vital runs but wickets continued to fall before The Human Windmill, Michiel Kwantes, tore into the flailing Biochemists and, with Smith returning to bat, left Plants on a very respectable 143 off 20 overs.

Respectable, but would it be enough? It's a question Mark Wheeler asks himself every morning in front of the mirror, only this time he might be strolling to the JCPCF, rather than the Coop round the corner. With Broughton away and Datta injured it was left to O'Toole and Van Tonder to open the bowling. Both answered the call and the pressure told immediately as The Canon in the deep, Brendan O'Leary, rifled in a throw to take a stunning run out. What a guy, what a right arm. O'Toole was managing to find bounce on the drenched pitch while at the other end Van Tonder gave a spin masterclass, taking two wickets. But the rain continued and so did Biochemistry, striving for a foothold in the game. Bowling through the pain Datta performed valiantly and in partnership with a miserly Forrester kept Biochemistry at bay. With bowling options scarce on the ground Fouracre came on to bowl and tide quickly turned, over his and everyone else's head. Just as the game looked to be running, or swimming, away from PSCC the clouds parted and a rainbow appeared. A bedazzled Forrester turned to senior pro and de facto team chaplain Smith at 3rd man and asked "What's that Ricki?" "That's hope my child." A hope seized on by Wheeler and Smith who bowled with enormous flight and guile to restrict the Biochemists. But it was still very close and getting very dark, fielding and batting becoming incredibly difficult. By now Biochemistry's strongest and angriest batsman was back in and looking menacing but thanks to the good work by Wheeler and Smith Biochemistry required 18 off the last over to win. With everyone else bowled out Fouracre returned and tried his hardest to let them, bowling no ball after no ball. But after intense controversy, team square ups and a couple of gentle

misunderstandings PSCC held on to win by 3 runs in an absolutely classic semi-final. Things look good for REF 2014.

PSCC take to the field in our first final in living memory on Tuesday 28th at the University Club, it would be great to have as many Ultras there as possible. It will be followed by our annual end of season dinner and awards ceremony at Chutneys Indian restaurant in central Oxford. If you would like to come please let me know.

MOM: Smith and Datta (match winning innings)

Plate final and end of season awards

It is with great pain that i say that yesterday PSCC lost in the plate final to Stats/Medawar, despite a pre-Paralympic opening ceremony dance routine from Monika Kalde and Yuki Yasumura and an on-field debut from a hedgehog. It was a very close and enjoyable game with particularly strong performances from Ricki Smith, Tom O'Toole and Michiel Kwantes, who seemed to attract the ball wherever he wandered.

When i woke up this afternoon still covered in sweat, shame and a neighbouring diner's korai sauce i could see no light. But, having opened the curtains and considered the journey we have been on, appreciated the full extent of our achievements. For a small department made up of florists a semi-final last year and final this year is something to celebrate and put in all

grant applications. Thanks to all those who played and supported the team, especially with the cake sales. Thanks also to a number of players who are leaving, including veterans Shyam and John Russell. Others are looking for new contracts, please give them one.

PSCC 2012 End of Season Award Winners

The Rembrandt Award for Dutch mastery - Michiel Kwantes

The Nicola Adams Award for best on-field impersonation of an amateur boxer - Joe Forrester

The K19 Award for rockets from the deep - Brendan O'Leary

The Tinman Award for biggest heart - Shyam Masakapalli

The Chinese Badminton Federation Award for sportsmanship - Jim Fouracre

The Lance Armstrong Award for performance enhancement - Brendan O'Leary

The Headington School for Girls Department of English Award for Coco the Clown of the season - Ricki Smith and Mark Wheeler

The Hammadou Djibo Issaka Award for speed between the wickets - Jon Lamb

Best Bowler - Tom Broughton

Best Batsman - Joe Forrester

Most improved - Michiel Kwantes

The David Boshier Lifetime Achievement Award - Ricki Smith

<u>PSCC vs OUP, 01/05/13</u>

All the great competitive spectacles have their curtain raisers. The Premier League has the Community Shield, the US Masters the opening par-3 competition and the Annual Maize Meeting the all too familiar dance routine to Agadoo led by Prof Jane Langdale. And so it is that PSCC take on publishing behemoths Oxford University Press (OUP) at the start of each cricketing season in the friendliest of friendlies. Despite the glorious sunshine and the guarantee of an enhanced social network at the end of the tie, PSCC entered into the game with a tinge of sadness however, knowing they would face the season without many comrades lost to career progression and pastures new, including seasoned veterans Shyam 'the Little Master' Masakapalli and John 'Mr Cricket' Russell.

However, just as the fall of every might oak allows a new sapling to emerge, the broad church that is PSCC was thus overjoyed to welcome no fewer than 4 debutants in this opening fixture: Tom Hughes and Elanor Jaskowska (Plants), Roger Close (Earth Sciences) and Sat Satish (St John's Care Homes). And what better way for them to learn their cricketing trade than by watching experienced craftsmen Ricki Smith and Tom O'Toole open the batting after OUP won the toss and elected to field. Without any restrictions on retiring and a winter without runs, both had their eyes on big scores on a lifeless pitch among a ring of fielders downtrodden by the decline of print media. They started effortlessly, untroubled by awkward bowling or calls for open access stroke making. After such a solid start Smith was unfortunate to fall on 8 when looking settled. Luckily O'Toole was in scintillating form and proceeded to lay waste to the publishers in a way not seen since the Luddites

got their hands on some printing presses. Support came from a scratchy Jim Fouracre (13) and a very friendly Joe Forrester (23) before a quick cameo from Satish (4 not out) closed the innings, O'Toole carrying his bat, albeit looking desperate to put it down somewhere. PSCC finished on a very respectable 145-3 off 20 overs, with O'Toole scoring 78, possibly the highest ever recorded by woodsman, well done Tom.

Sourav Datta opened the bowling for Plants like a Godolphin-trained horse bolting from a stable. Despite vigour from his end and (very friendly) variety from Forrester at the other OUP started at pace. 4 overs in and OUP were cruising until a moment of magic from debutant Hughes in the field changed the game, running out a monster (physically speaking, i'm sure he's a very generous editor) of a publisher with a direct hit. As his teammates rushed in to celebrate, a bemused looking Hughes asked what he'd done. That innocence will not last long in the Jack Cox trophy where Material Scientists and Engineers lurk. Not to be outdone, 2 minutes later Mark Wheeler produced a carbon copy run out, the batsman not facing a ball. There hasn't been a look of innocence on Mark's face since 1996. With PSCC in the ascendancy OUP started to crumble. Taking 2 balls to find the pitch, Satish took a wicket with his third and bowled exquisitely, taking 3 for 10 overall, sadly bringing a very brief career at PSCC to a close. PSCC's loss is occupational health in Plymouth's gain (except for perhaps the first 2 patients). Hughes continued his fine debut with a wicket and some tight lines, O'Toole bowled as well as he batted and Close took an exceptional wicket with his very first ball before retiring gracefully with wrist spasms. With the shadows lengthening Wheeler

stepped up and completed the rout in the last over with characteristic flight, PSCC winning by 46 runs. Special mention to Brendan O'Leary (better men than Brendan have crossed the North-South building divide and lost the ability to hurl a ball but not he) and Jaskowska for real quality in the field. Roll on the real thing.

Men of the Match: Tommy O'Toole (runs), Michiel 'The Windmill' Kwantes (inspiration behind the runs)

PSCC vs Pathology, 06/06/13

For the 3rd season in a row PSCC took on Pathology on Thursday in our opening game of this season's Jack Cox tournament. Unbeaten in our last two fixtures against them Pathology were hungry for revenge, especially with big money Path-Plants transfer Dr Steve Kelly so brazenly celebrating his 26th birthday from the Uni Club balcony, leading the Ultras in sectarian scientific voice. And there is no hungrier man in the Jack Cox than old PSCC foe, Pathology captain Big Clive. We dance again mon brave. Hurrying towards the assorted botanists to conduct the toss with a gentle tide of perspiration flowing across his gleaming skull, a glint in his deep set eyes, here was a man looking for a quick victory, anxious in the knowledge that tickets for the Club Thursday curry buffet were limited in number. With Sourav Datta confined to babysitting duty for the first 45 minutes of the match, PSCC were relieved to be put into bat following another BC toss victory.

After an auspicious partnership in our friendly against OUP earlier in the summer, seasoned veterans Ricki Smith and Tommy O'Toole again opened up for PSCC. Separated only by 20 years and an extended period of military service, there is a tacit mutual respect between these two strokemakers at the crease. Datta could not have asked for better rolemodels to inspire young Isha, watching keenly from the sidelines. Playing calmly and elegantly, they clinically set about the Pathology bowlers, building a perfect platform for the rest of the innings. Both retired on 31, much to the relief of the ailing pathologists. Joe Forrester came in and quickly went, clearly worried about the overdue arrival of BC into the bowling attack, his presence creating both physical and figurative shadows over the PSCC batsmen. Jim Fouracre and Tom Broughton steadied the ship for a few overs, Fouracre scoring almost exclusively between his own legs and Broughton shuffling aggressively down the crease, before Fouracre succumbed to a straight one on 22. And then the big guy arrived, pregame rumours of an achilles problem clearly fabricated by the Pathology press department as he took only 3 balls to remove Broughton for 16. Brendan O'Leary followed 2 balls later and suddenly PSCC were in trouble. Luckily Mrs Datta finally arrived and freed Mr to play an inspired innings, blasting 22 off one over. With able support from debutant Richard O'Toole and Liam Medler, who stood fast in the face of a BC torrent, PSCC finished on a very competitive 159 off 20 overs, Datta unbeaten on 28.

Competitive perhaps in normal circumstances, it is well known that the presence of BC adds 20 to any total (or round) however. So it was with great surprise and relief that, instead of seeing BC lumber out to the middle, his bat

a matchstick between his giant paws, a hushed PSCC witnessed a little Australian pixie like creature amble daintily to the crease in his place. What trickery could this be? What sorcery does this pixie possess? Opening Toms Broughton and O'Toole were not to be enchanted and bowled quickly and tightly, conceding only occasionally, and were unfortunate not to take any wickets. With the required run rate spiraling Datta finally enticed the pixie out of its defensive grotto with a comfortable catch taken by Smith. PSCC were in a dominant position but with BC next in to bat nothing was certain. Further tight bowling from O'Toole Jnr and a resurgent Jon Lamb kept any onslaught at bay though, with Lamb taking a deserved wicket. BC could not be held off forever though and finally unleashed, despite valiant attempts by spin twins Mark Wheeler and Smith to recreate the "ball of the century" in its 20th anniversary week. His momentum eventually got the better of him however and he was run out by Fouracre, the look of disappointment in his eyes fleeting as he bounded off for an early korma bath. With BC gone Forrester was suddenly a lot keener to bowl and finished off the innings with 3 wickets for 13 in fine fashion, supported magnificently by a gazelle like O'Leary in the field (if gazelles had dishes for hands), PSCC finally winning comfortably by 28 runs.

Men of the match: Smith/O'Toole (batting), Datta (exemplary balance of childcare and sporting achievement), O'Leary (champagne fielding).

<u>PSCC vs Materials, 13/06/13</u>

Yesterday saw one of the great Jack Cox matches of recent times take place at the Uni Club: home of flavour, home of sport. It is only a shame that Jack Cox was not around to enjoy it (if he ever existed at all). PSCC, looking to avenge a bad tempered defeat in 2010 the last time we played, took on a socially reformed Materials side knowing that victory would take us through to the Cup stages of the competition for the first time since ash dieback was only a speck on a Polish sapling. With the silver Bosh away on the ash resistance front line and key all rounders Forrester, Broughton and O'Toole also absent PSCC were asked to field with only 9 men, a nervous looking Fouracre behind the stumps and belted by 40 mph gusts.

The Materials department may be an enormous and technologically advanced interdisciplinary centre but they have found nothing as hard, no lattice as rigorous nor a crystal as bright as Sourav Datta. With all the odds stacked against us he struck crucially in the first over, trapping the Materials opener (roughly) leg before wicket. There was hope after all. With the pressure rising Andries Van Tonder had a cast iron LBW decision and stumping turned down in the next as the Materials umpires threatened to revert to nefarious type. Not to be dismayed, Van Tonder continued to bowl menacingly and was rewarded 2 overs later with a wicket. At this point Pastor Ricki Smith finally arrived from a delayed school governors finance meeting. He wouldn't say what had happened but it was perhaps telling he arrived on the park and ride service rather than by car. Buoyed by some pastoral care PSCC started ripping through the Materials line up like a chainsaw through balsa. Swing kings Tom Hughes and Jon Lamb used the wind to great effect, finding big big

movement, to take 1 and 2 wickets respectively. Smith found solace in a couple of wickets and a run out (if only he'd covered his accounting tracks like he covers ground) and Roger Close bowled tidily to also take a wicket. With one retired batsman to return to the crease Datta fired himself up again to finish the innings, Materials all out for 127 off just 14 overs. Lets hope some of their substrates have greater durability.

With half the top order batsmen missing all PSCC had to do was bat sensibly, see off the opening bowling and edge home. 2nd over Fouracre swings at a straight one and is bowled. Fortunately Smith and Datta had more sense and played perfectly for the next 10 overs, keeping just on top of the run rate and looking very comfortable. Then disaster struck, Smith and Datta both fell in consecutive overs and all of a sudden PSCC were looking very shaky. First Cannonball O'Leary, then Van Tonder, then Mark Wheeler and Close all fell and we'd lost 6 wickets in 5 overs of carnage. With 19 runs to win required off the last 4 overs and only debutant Luke Tomlinson (Political Theory) left to bat, it was on Hughes and Lamb to bring us home. Aware that Tomlinson had spent most of the game naming passing gull species and reminiscing about music festivals spent with opposing fielders they knew it was all on them. You could have cut the atmosphere with a knife. The fielders looked nervous, Hughes and Lamb looked nervous, the windswept Club balcony looked cold and disinterested. But, despite some lofted chips and nervous running, HuLa brought us home with aplomb to rapturous applause from the onlooking botanists. Smith may not be able to play cricket in prison but at least he will be able to while away the hours in solitary knowing we made the Cup this one

time. A fantastic all round performance, particularly in the field where O'Leary and Tomlinson always looked dangerous, and a very enjoyable victory.

Men of the Match: Hughes and Lamb (bowling and match winning partnership), Tomlinson (outstanding knowledge of gull species)

PSCC vs Chemistry, 18/07/13

After narrow defeats against Jack Cox behemoth Medicine and a free scoring collection of graduate guns for hire in a friendly versus Linacre and friends, PSCC took to the field against Chemistry looking to get back to winning ways in a game we had to win to stay in the Jack Cox inter-departmental cup. It had been billed by some as The Biggest Game Of Our Lives, and with #TBGOOL trending just behind #royalbaby on Twitter the pressure was intense. The last time we played Chemistry in 2011 it resulted in a heavy loss followed by a heavy session with the Chemistry captain, so we knew we were in for a tough if well catalysed time.

Opening the bowling after a rare toss victory, Tom Broughton and Sourav Datta stormed in over a sun bleached Club turf against an opening batsman who'd hit us for 30 in 3 overs the previous week. A solar powered Broughton was too much for him this time however, snaring him in the 3rd over in a crucial breakthrough. Chemistry were reeling and the pressure rising but the first to crack was Jim Fouracre, or more specifically the webbing between his fingers, having to leave the field after botching a catch to a less than

sympathetic audience of Club staff. Solace and strapping are in scant supply when the Sky Sports rolling ticker is on the big screen. In his absence Plants were to prosper however, with the outstanding Andries Van Tonder taking a couple of key wickets, finishing with 2 for 11 off his 3 overs. But despite these early wickets, and further tight bowling from Tom Hughes and Roger Close, Chemistry were still scoring relatively freely. Nimble work in the field by the gorgeously bearded Brendan O'Leary and debutant Matt Pickles (Press Office, gorgeous prescription sunglasses) kept the damage down, before late wickets from a bewitching Mark Wheeler and the less than energetic Joe Forrester, his daal only diet is clearly taking its toll. A final double wicket nearly-maiden by the impressive Broughton and Chemistry closed on 155-9 off 20 overs, an exceptional effort by the PSCC bowlers.

Without opening stalwarts Smith and O'Toole, the PSCC batting line up was already stretched and the omens did not improve when the speakers for our much heralded introductory tunes magically ran out of electricity and Datta asked to be excused so he could pick up an ingredient for his wife's special BBQ marinade. Request refused, sorry spice fans. With Forrester unfortunate to fall early Datta channeled his anger and concerns over the minimum recommended marinating period by batting aggressively, he and Broughton taking on the Chemistry attack. With Broughton triggered when looking set on 30 Datta and Fouracre continued, Datta impassioned and Fouracre's fingers stuffed in a glove. At the halfway stage PSCC were just behind the required run rate with wickets in hand and a famous win looked possible. However just as we dared to dream the all too familiar collapse set in. Despite being so

impassioned he managed to break a bat simply through running, Datta survived to retire on 30 but Fouracre, O'Leary and Hughes all disappeared in successive overs. Evergreen athlete David Boshier marked his season's debut with some staunch resilience but, combined with late runs from Pickles and Van Tonder, it was not quite enough, PSCC falling 35 runs short.

Maybe we dreamed too big, maybe our only mistake was believing in ourselves. Having had 5 hours in hospital waiting rooms to reflect on it i'm still not sure. Perhaps these words from the sage Boshier sum it up best, words that only come with 40 years Jack Cox experience: you can't always be a big fish in a little pond, sometimes you have to be that little fish. One thing's for sure, if Forrester sticks to this daal fad he'll be a minnow soon enough. We might be out of the Jack Cox but we've still got at least one more game this season, keep believing.

Men of the match: Broughton (bowling), Datta (runs, culinary cooperation), Ultras (great support)

<u>End of season gala awards dinner</u>

Following a crushing victory for a PSCC side led by Tom Hughes against the Said Business School in our final match, this ground breaking season is unfortunately coming to a close. However there can be no closure, no farewells to fallen comrades, no placating of the cricketing gods, without the Annual End of Season Gala Awards Dinner (AEOSGAD). This will take place

on Thursday 22nd at Oxford's finest BYOB curry house Arzoo (sister restaurant to Bicester's finest BYOB curry house, Arzoo) at 8.30. For those that are able we will assemble at 7.30 in the Jericho Tavern across the road. If you would like to come and haven't already informed me please let me know. All members of the department are welcome, whether you understand the rules of cricket or otherwise. It's not stopped many playing this season. If you are unable to attend but would still like to vote in the end of season awards, please email me your votes in the categories outlined below.

I would like to remind those that are attending of the Gala status of the dinner, please dress appropriately.

Following a gala dinner rich in posturing but low in actual gala content, i am pleased to announce the winners of the PSCC end of season awards (see below). But as always at this time of year, the feelings of joy and emotional relief are tempered with the need to say goodbye to some departing heroes. There has been no more heroic interdepartmental cricketer over the past 30 (+/- 15) years than David Boshier, who's classical batting style and athletic anecdotes have kept us on the straight and narrow. We also say goodbye to Sourav Datta, who has contributed more runs, overs and utter confusion between the wickets combined than anyone over the past few seasons, while Jon Lamb is on loan in Germany. Outside of the department, cornerstones of the team Joe Forrester, Tom Broughton and Tommy O'Toole are sadly

additionally moving to outfields new. I am also off and am very pleased to announce that Tom Hughes will be taking over the captaincy next year. As a teenager he claims to have had equivalent sporting prowess to school mate Jess Ennis, lets hope he rediscovers it in time for next summer. Many thanks to all those who have supported the team over the last few years, be it through squawking through a vuvuzela at the Club, baking for cake sales or feigning interest in cricket at coffee times.

The Rev Rowan Williams award for services to spiritual leadership - Jim Fouracre

The Kraken award for monster in the deep - Brendan O'Leary

The Gareth Batty award for spirit of cricket - Jim Fouracre

The Oxford Sports Masseur Association award for limbering up - Roger Close

The Adrian Mutu award for best summer signing - Tom Hughes

The Monty Panesar award for celebrating success - Andries Van Tonder

The George Osborne award for a safe pair of hands - Matt Pickles

Best bowler - Tom Broughton

Best batsman - Tommy O'Toole

Most improved player - Brendan O'Leary

The David Boshier Lifetime Achievement award - David Boshier and Sourav Datta