Regulatory gateways for cell-specific gene expression in C_4 leaves with Kranz anatomy

Gregory Reeves, Mathieu J. Grangé-Guermente, and Julian M. Hibberd

Department of Plant Sciences, Downing Street, University of Cambridge, Cambridge CB2 3EA, United Kingdom.

*Correspondence to jmh65@cam.ac.uk

Key Words: C_4 photosynthesis, cell specificity, mesophyll, bundle sheath, gene regulation

Running title: C_4 gene expression

Highlight statement: Analysis of photosynthesis gene expression in multiple C_4 lineages indicates that individual genes are regulated at multiple levels and that the mechanisms operate in C_3 ancestors.

ABSTRACT

C_4 photosynthesis is a carbon concentrating mechanism that increases delivery of carbon dioxide to RuBisCO and as a consequence reduces photorespiration. The C_4 pathway is therefore beneficial in environments that promote high photorespiration. This pathway has evolved many times, and involves restricting gene expression to either mesophyll or bundle sheath cells. Here we review the regulatory mechanisms that control cell preferential expression of genes in the C_4 cycle. From this analysis, it is clear that the C_4 pathway has a complex regulatory framework, with control operating at epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels. Some genes of the C_4 pathway are regulated at multiple levels, and we propose that this ensures robust expression in each cell type. Accumulating evidence suggests that multiple genes of the C_4 pathway may share the same regulatory mechanism. The control systems for C_4 photosynthesis gene expression appear to operate in C_3 plants, and so it appears that pre-existing mechanisms form the basis of C_4 photosynthesis gene expression.

INTRODUCTION

Ribulose 1,5-Bisphosphate Carboxylase Oxygenase (RuBisCO) is the primary carboxylation enzyme in photoautotrophs - fixing inorganic atmospheric carbon dioxide (CO_2) into an organic form for carbohydrate anabolism (Calvin and Benson, 1948; Nelson and Cox, 2008). It has been proposed that every organic carbon molecule has passed through the RuBisCO active site at some point in time (Mauseth, 2012). When RuBisCO and photosynthesis first appeared, the earth's atmosphere was dominated by CO_2. Over a prolonged period of time, oxygenic photosynthesis transformed the atmosphere and oceans, allowing aerobic organisms to survive (Luo *et al*., 2016; Holland, 2006).

Despite its indispensable role in carbon assimilation, RuBisCO is a surprisingly inefficient enzyme. Its rate of catalysis is slow, and it has a low affinity for CO_2. Furthermore, it carries out a deleterious side-reaction that fixes molecular oxygen (O_2) rather than CO_2 (Portis and Parry, 2007). In a high CO_2 world, this would not likely affect growth, however, with approximately 21% O_2 and 0.04% CO_2 in the current atmosphere, catalysis of O_2 is common and generates a toxic two carbon compound that must be recycled via photorespiration (Bowes *et al*., 1971; Sharkey, 1988). Not only is energy expended in photorespiration, but it also leads to loss of carbon, leading to reductions in photosynthetic efficiency. To date, attempts to manipulate RuBisCO to impair or remove its oxygenase activity while maintaining carboxylase functionality have not been successful (Whitney *et al*., 2011; Spreitzer and Salvucci, 2002; Peterhansel and Offerman, 2012). However, on an evolutionary timescale, multiple independent lineages of plants have developed mechanisms to reduce oxygenation by RuBisCO in an oxygen rich world (Sage *et al*., 2012). Rather than evolving an improved CO_2-O_2 discrimination mechanism, a seemingly simple modification that would eradicate energy wasted in photorespiration, these lineages instead developed carbon-concentrating mechanisms (CCMs) that boost carboxylation by RuBisCO. Of these CCMs, C_4 photosynthesis is the most prevalent in terrestrial plants – being found in around 8100 species distributed across more than sixty lineages of plants (Sage, 2016).

C_4 photosynthesis operates as a molecular pump that generates high concentrations of CO_2 around RuBisCO (Figure 1A). A unique form of plant morphology termed Kranz anatomy has evolved in conjunction with this molecular pump to facilitate C_4 photosynthesis (Haberlandt, 1904;

El-Sharkawy and Hesketh, 1965; Downton and Tregunna, 1968; Hatch, 1987). Kranz anatomy typically consists of narrowly spaced veins with large bundle sheath (BS) cells surrounded by outer and concentric layers of mesophyll (M) cells (Figure 1B). In all plants, CO_2 enters the plant through stomata and diffuses into M cells. However, in C_4 species, rather than the initial fixation by RuBisCO CO_2 is converted into bicarbonate (HCO_3^-) by CARBONIC ANHYDRASE (CA). PHOSPHO*ENOL*PYRUVATE CARBOXYLASE (PEPC) then combines HCO_3^- with the three-carbon acceptor molecule phospho*enol*pyruvate (PEP) to produce the four-carbon acid oxaloacetate in M cells. The initial formation of the C_4 acid led to the name C_4 photosynthesis. Oxaloacetate is then commonly reduced to malate, which diffuses into BS cells where it is decarboxylated to yield pyruvate, a three-carbon compound, and a molecule of CO_2, which is secondarily refixed by RuBisCO. Key steps in the C_4 carbon assimilation cycle are therefore: initial CO_2 fixation by PEPC to form a C_4 acid in M cells, decarboxylation of the C_4 acid in BS cells to release high concentrations of CO_2 near RuBisCO for re-fixation in the Calvin-Benson-Bassham (CBB) cycle, and lastly, regeneration of the C_3 substrate PEP in M cells (Hatch, 1987; Kagawa and Hatch, 1974).

The compartmentation of C_4 photosynthesis between M and BS cells depends on the regulation of the genes encoding this complex metabolic network (Hibberd and Covshoff, 2010). Fifty years after the discovery of C_4 photosynthesis (Hatch and Slack, 1966; Furbank, 2016) there is considerable interest in understanding its regulation between M and BS cells. A better understanding of how C_4 gene expression is controlled could impact on efforts to incorporate the C_4 pathway into C_3 crops to improve their photosynthetic efficiency (Sage and Zhu, 2011; Covshoff and Hibberd, 2012; Matsuoka *et al*., 2001; von Caemmerer *et al*., 2012; Raines, 2011; Hibberd *et al*., 2008; Osborne and Beerling, 2006).

Here we aim to highlight recent advances in understanding the gene regulation associated with the C_4 pathway. The regulation of C_4 photosynthesis gene expression is stringent. Once established, environmental stimuli such as heat, cold, light or dark, and even hormonal manipulation, are not known to perturb its patterns of cell specific gene expression (Bräutigam and Weber, 2011). It is not clear if the complex C_4 phenotype is based on the evolution of conserved regulatory mechanisms that have repeatedly been recruited into C_4 photosynthesis in independent

C_4 linages. However, it is apparent that the pathway is regulated at many checkpoints including epigenetic, transcriptional, post-transcriptional, translational and post-translational processes. Analysis of the available literature supports the notion that to ensure the correct patterns of gene expression, individual genes are subject to an interconnected mosaic of gene regulation operating at many levels.

C_4 GENES ARE SUBJECT TO EPIGENETIC REGULATION

Epigenetic regulatory mechanisms, such as covalent modifications to DNA or alternations in chromatin structure, can impact on gene expression, and are not dependent on the underlying DNA sequence (Feng *et al*., 2010). As a passive barrier to gene expression, chromatin compacted by nucleosomes, is inaccessible to regulatory proteins (Loidl, 2004). Covalent modifications to DNA such as methylation at cytosine nucleotides, or histone modifications, such as lysine acetylation or methylation, are commonly associated with epigenetic regulation (Vanyushin and Ashapkin, 2011).

Based on a search of Histone H3 Lysine 4 tri-methylation (H3K4me3) and H3 lysine 9 acetylation (H3K9ac) within C_4 pathway gene promoter regions, which represent histone marks for actively transcribed genes (Wang *et al*., 2009; Dion *et al*., 2005), six C_4 cycle genes were found to share a similar histone code in the C_4 grasses *Zea mays* (maize), *Setaria italica* and *Sorghum bicolor* (Heimann *et al*., 2013). The M preferential C_4 genes, *PEPC*, *CA* and *PYRUVATE, ORTHOPHOSPHATE DIKINASE (PPDK)* showed an enrichment of H3K4me3 in M cells compared with the BS. In contrast, C_4 genes such as *NADP-MALIC ENZYME* (*NADP-ME*) and *PHOSPHO*ENOL*PYRUVATE CARBOXYKINASE* (*PEPCK*) were enriched with H3K4me3 marks in the BS compared with M cells. Interestingly, the small subunit of RuBisCO (*RbcS*) showed nearly the same degree of H3K4me3 in both M and BS cells (Heimann *et al*., 2013). All of these genes showed an enrichment of H3K9ac marks upon illumination (Heimann *et al*., 2013), which may indicate that histone modification helps regulate both cell specificity and the induction of gene expression by light. It therefore appears that H3K9ac primes genes for expression from light stimuli, and H3K4me3 further initiates preferential expression of C_4 genes in either M or BS cells but is independent of light (Offerman *et al*., 2008). A genome-wide search of maize revealed that many putative regulators of C_4 photosynthesis exhibited similar H3K9ac histone activation marks (Perduns *et al*., 2015). Among 294 genes found to be M or BS specific from maize and C_4 *Gynandropsis gynandra* (Aubry *et al*., 2014), 68 had up-regulated H3K9ac marks (Perduns *et al*., 2015). Examples of putative regulators that also show cell specific histone marks include DNA-binding with One Finger transcription factor (Yanagisawa, 2000), and an ethylene response element binding factor (Pick *et al*., 2011). Thus, histone modifications may have a broad role to

control the expression of enzymes and putative regulators involved in the induction of C_4 photosynthesis in response to light and may also prime their cell specificity.

It is also possible that DNA methylation serves as a regulator of C_4 gene expression. Methylation sensitive restriction endonucleases (Langdale *et al*., 1991) and bisulphite sequencing of DNA (Tolley *et al*., 2011) indicated that the maize *PEPC* promoter region contains several cytosine methylation sites. These cytosine residues are de-methylated in response to light specifically in M cells. However, these methylated cytosines are upstream of the 600 base pair (bp) region of the maize *PEPC* promoter that is sufficient to generate preferential expression in M cells (Matsuoka *et al*., 1994). Therefore, it is unclear whether these methylated sites play a role in M cell specific expression of maize *PEPC*. Taken together, these results indicate that epigenetic modifications likely contribute to preferential expression of C_4 genes in M or BS cells and may be an initial mechanism that primes C_4 genes for preferential expression in these cells (Wang *et al*., 2011; Heimann *et al*., 2013). In subsequent sections, we present evidence that argues for the spatial regulation of C_4 genes being subject to additional levels of regulation.

CIS-ELEMENTS AND *TRANS*-FACTORS AFFECTING TRANSCRIPTION

Transcription represents the first stage of gene expression and allows a DNA sequence to specify an RNA copy by the action of RNA polymerase. RNA polymerase requires the assistance of transcription factors that bind euchromatic DNA elements to facilitate pre-initiation of transcription. Transcription factors and the DNA elements they recognise are often referred to as *trans*- and *cis*-regulators respectively. There has been significant interest in understanding the *cis*-elements and *trans*-factors that facilitate M or BS expression of C_4 genes (Hibberd and Covshoff, 2010).

One of the best-characterised examples of transcriptional control in C_4 photosynthesis is the promoter region of the *PEPC* gene from maize. As previously mentioned, a 600bp region of the maize *PEPC* promoter is sufficient to generate strong and light-activated expression in M cells (Taniguchi *et al*., 2000; Kausch *et al*., 2001). Interestingly, these 600bp can drive M expression in C_3 rice (*Oryza sativa*) (Matsuoka *et al*., 1994; Ku *et al*., 1999). In the C_4 dicotyledon *Flaveria trinervia*, the *PEPC* promoter region is also sufficient for preferential expression in M cells

(Stockhaus *et al*., 1997). Detailed dissection of this *F. trinervia PEPC* promoter region identified a 41bp *cis*-element responsible for M expression, termed the Mesophyll Enhancing Module 1 (MEM1) (Gowik *et al*., 2004). MEM1 is also capable of conferring expression in the palisade M of C_3 *Nicotiana tabacum* (Akyildiz *et al*., 2007). Interestingly, the *PEPCK* gene from C_4 *Zoysia japonica* (a warm season turf-grass) that encodes one of the C_4 acid decarboxylases that supplies CO_2 to RuBisCO in BS cells is also able to generate expression in the BS of rice (Nomura *et al*., 2005). The most parsimonous explanation for these data is that in multiple C_4 lineages, *PEPC* and *PEPCK* have evolved novel cell specific *cis*-elements that utilised an ancestral regulatory mechanism that was based on *trans*-factors present in C_3 species.

It is becoming increasingly apparent that the spatial patterning of C_4 genes can be mediated by *cis*-elements found inside genes. In both genes that encode the heterodimeric NAD-dependent Malic Enzyme (NAD-ME), a 240bp region within the coding sequence confers BS expression in *G. gynandra* (Brown *et al*., 2011). This 240bp sequence from *G. gynandra* does not produce BS specificity in C_3 *Arabidopsis thaliana.* However, the orthologous sequence from *Arabidopsis* is able to generate BS expression in *G. gynandra*. Thus, the *cis*-elements used to generate cell specificity of *G. gynandra NAD-ME* are present in the ancestral C_3 state but the *trans*-regulator appears to be lacking.

A number of *trans*-factors have been identified *in vitro* that bind to *PEPC* promoter regions from different C_4 species. For example, DNA Binding with One Finger 1 (DOF1), DOF2, *F. trinervia* ZF-HD homeobox protein 1 (FtHB1), Maize Nuclear Factors (MNFs), and PEP-I all interact with *PEPC* promoters (Windhövel *et al*. 2001; Westoff and Gowik, 2004; Kano-Murakami *et al*., 1991). DOF transcription factors are specific to the plant kingdom and both up- and down-regulate light-responsive genes (Park *et al*., 2003; Ward *et al*., 2005). DOF1 appears to promote *PEPC* expression throughout the plant (Yanagisawa and Sheen, 1998) and it is proposed that expression is then restricted to M cells by the repressive function of DOF2 which is itself BS specific (Yanagisawa, 2000). When *DOF1* expression was repressed to very low levels in maize mutants *PEPC* expression was unaltered (Cavalar *et al*., 2007). Thus, low activity of *DOF1* is able to regulate *PEPC* expression. It has been speculated that DOF1 may regulate other C_4 genes such as *PPDK* (Yanagisawa, 2000), and it also appears to regulate *PPDK* in tissues of rice (Zhang *et al*.,

2015). Despite indications that the homeodomain proteins, FtHB1, FtHB3 and FtHB4 bind *cis*-elements present in the 5'-untranslated region (UTR) of *PEPC* from *F. trinervia* (Windhövel *et al*., 2001), there was no disruption in expression of *PEPC* when the putative binding site was deleted (Engelmann *et al*., 2008). To date, none of these *trans*-factors have been verified *in vivo* as being necessary or sufficient for expression in M cells of C_4 leaves, thus the role these proteins play in *PEPC* regulation is not yet elucidated.

Comparative transcriptomics between C_3 and C_4 species and between M and BS cells of the same C_4 species has led to the *in silico* identification of many transcription factors proposed to impact C_4 gene expression (Bräutigam *et al*., 2011; Gowik *et al*., 2011, Ding *et al*., 2015; Li *et al*., 2010; Aubry *et al*., 2014; Wang *et al*., 2014; Yu *et al*., 2015; Pick *et al*., 2011; Rao *et al*., 2016; John *et al*., 2014; Chang *et al*., 2012; Tausta *et al*., 2014). To our knowledge, none of these candidates have been validated *in vivo*. The reasons for the lack of progress in identifying the transcriptional regulators of cell specificity are not clear, but may be associated with redundancy being caused by multiple transcription factors binding relevant *cis*-elements, and/or significant amounts of post-transcriptional control.

The only transcription factor that has a fully validated role in C_4 photosynthesis is Golden Like 2 (GLK2) (Wang *et al*., 2013a). Rather than regulating genes of the core C_4 cycle, GLK2 controls genes critical for chloroplast development and the photosynthetic apparatus (Waters *et al*., 2009; Langdale and Kidner, 1994). In most C_4 species, there are two *GLK* genes that are preferentially expressed in either M and BS cells (Wang *et al*., 2013a), and differential expression of these *GLK* genes is thought to give rise to dimorphic chloroplasts in maize M and BS cells (Rossini *et al*., 2001; Wang *et al*., 2013a). Evidence is accumulating for the SCARECROW (SCR) and SHORTROOT (SHR) transcription factors being involved in specification of Kranz anatomy of C_4 leaves. In *Arabidopsis*, SCR and SHR have been shown to determine the fate of BS cells (Cui *et al*., 2014). *SCR* and *SHR* transcripts accumulate preferentially in the BS of several C_4 species, such as maize (Wang *et al*., 2013b; Tausta *et al*., 2014) and *Echinochloa glabrescens* (Covshoff *et al*., 2016). Mutations in maize *SCR* led to more than one BS layer surrounding veins, aberrant BS chloroplast differentiation and loss of minor veins (Slewinski *et al*., 2012). To our knowledge

however, these transcription factors have not yet been implicated in controlling genes of the C_4 cycle.

STABILITY OF C_4 TRANSCRIPTS

Post-transcriptional regulation is typically used to define mechanisms that impact RNA abundance prior to translation. The stability of mRNA is affected in multiple ways including pre-mRNA processing within the nucleus, or after nuclear export to the cytosol by the action of small RNAs and RNA-binding proteins. Sequences present in the mRNAs of several C_4 cycle genes contribute to cell-specificity. In *G. gynandra* the 5' UTR of *CA* confers preferential expression in M cells (Kajala *et al*., 2012). The *cis*-element responsible for this M expression is also found in the UTRs from *PPDK*, and it appears to act at the level of translational efficiency (Williams, Burgess *et al*., 2016). Although this *cis*-element is present in orthologous genes from *A. thaliana*, it does not generate M preferential expression in *G. gynandra* (Williams, Burgess et al 2016). These data imply that multiple genes (*CA* and *PPDK*) have used the same *cis*-regulatory code to generate preferential expression in M cells of the C_4 leaf. Although these *cis*-elements are present in C_3 orthologues, it is not until they are integrated into the C_4 leaf that they specify the spatial patterns of expression required for the C_4 pathway.

In C_4 amaranth, mRNAs encoding both the large subunit (LSU) and small subunit (SSU) of RuBisCO accumulate in M and BS cells early in leaf development but the polypeptides only accumulate in BS cells (Patel and Berry, 2008; Boinski *et al*., 1993). This strongly implicates post-transcriptional regulation of gene expression. In the case of *RbcS*, the 5' and 3' UTRs from *Flaveria bidentis* and *Amaranthus hypochondriacus* drive preferential expression of β-glucoronidase (GUS) in the BS, possibly indicating post-transcriptional regulation (Patel *et al*., 2004; Patel *et al*., 2006). In maize, it was also found that the promoter region including 66bp of the 5' UTR (Nomura *et al*., 2000) as well as the 3' UTR (Viret *et al*., 1994) are responsible for BS accumulation of *RbcS*. Together the data suggest transcriptional regulation and RNA stability of *RbcS* are important for expression. In maize, *rbcL* transcripts accumulate in both BS and M cells in darkness, but upon illumination transcripts are restricted to BS cells. It has been proposed that this process is likely controlled by RNA stability (Sheen and Bogorad, 1985; Kubicki *et al*., 1994).

Furthermore, in amaranth, *rbcL* transcripts appear in M and BS chloroplasts, however, the LSU of RuBisCO protein only accumulates in BS cells (Boinski *et al*., 1993). Overexpression of *RbcS* and *rbcL* by the ubiquitin promoter in maize did not lead to accumulation of RuBisCO in M cells, but was still expressed in BS cells (Wostrikoff *et al*., 2012). This is consistent with a mechanism subsequent to transcription, affecting mRNA stability of the RuBisCO holoenzyme to facilitate BS localisation. In fact, a nuclear encoded mRNA binding protein, *rbcL* RNA S1-Binding Domain (RLSB) has been shown to co-localise and bind to *rbcL* mRNA in chloroplasts (Bowman *et al*., 2013) and lower *RLSB* expression reduces expression of *rbcL* (Yerramsetty *et al*., 2016). RLSB mRNAs accumulate preferentially in BS cells in a number of C_4 species, yet accumulate throughout the leaf in *A. thaliana* (Bowman *et al*., 2013; Yerramsetty *et al*., 2016). Thus, in C_4 leaves restriction of RLSB to the BS may contribute to the accumulation of LSU in this cell-type.

Control of RNA stability is also important for the glycine decarboxylase complex (GDC), a component of photorespiration. In C_4 species, GDC operates in the mitochondria of BS cells, whereas in C_3 species, it is expressed throughout all photosynthetic tissues (Bauwe *et al*., 2010). Reporter promoter fusions of the P-subunit of the GDC (*GLDPA*) confer preferential expression in the BS of C_4 *F. trinervia* and C_3 *A. thaliana* (Engelmann *et al*., 2008). Although this is consistent with transcriptional regulation, the situation is more complex because this upstream sequence contains two promoters, which together ensure BS specific and high levels of expression. With respect to the transcription start site, a proximal sub-promoter directs strong BS specificity while a distal promoter confers constitutive expression in leaf tissue. The distal promoter is strong and generates abundant transcripts in M and BS cells. However, due to the presence of a cryptic intron in the 5' UTR of transcripts derived from this promoter, these transcripts are improperly spliced during post-translational processing. Incorrectly spliced *GLDPA* transcript variants that lack the BS specific elements are suppressed via RNA decay (Wiludda *et al*., 2012). The regulation of *GLDPA* therefore epitomizes how a single gene is regulated at multiple levels, in this both transcription and RNA processing, to ensure cell specific expression in the C_4 leaf.

TRANSLATIONAL AND POST-TRANSLATIONAL MECHANISMS

Translational and post-translational regulation of C_3 photosynthesis gene expression are particularly common for plastid-encoded genes (Cohen and Mayfield, 1997; Chi *et al*., 2012; Jensen *et al*., 2007; Järvi *et al*., 2015; Schöttler *et al*., 2015). Evidence for translational and post-translational regulation in C_4 leaves is provided through analysis of genes encoding proteins of the CBB cycle and Photosystem II (PSII). In maize most of the CBB cycle operates in BS cells, whereas PSII preferentially accumulates in M cells (Friso *et al*., 2010; Li *et al*., 2010; Kawaga and Hatch, 1974; Schuster, 1985). As photosynthesis in eukaryotes depends on co-ordinate expression from both the nuclear and plastid genomes, this integration of genomes also impacts on the C_4 pathway.

Mutants for the M specific maize *high chlorophyll fluorescence 136* (*hcf136*) gene completely lack PSII in M cells (Covshoff *et al*., 2008). Interestingly, both cytosolic and plastidic transcripts encoding proteins of the PSII core accumulate to normal levels in M cells of *hcf136* mutants, yet the corresponding proteins are undetectable, strongly implying translational regulation. It has been suggested that the *hcf136* mutant is not able to process the plastid-encoded *psbB-psbT-psbH-petB-petD* polycistron, a co-transcriptional unit that encodes components of PSII and the cytochrome *b6f* complex, in M cells (Chi *et al*., 2012; Nakamura *et al*., 2003; Meierhoff *et al*., 2003). Furthermore in *hcf136* mutants, other photosynthetic genes show ectopic expression. For instance, PSI and ATP synthase complex transcripts were up-regulated in M cells of *hcf136* mutants, and many BS specific genes were expressed in M cells (Covshoff *et al*., 2008). The mechanism by which *hcf136* impacts on differential gene expression of these other photosynthesis genes between M and BS cells remains unclear.

The maize mutant, *bundle sheath defective 2* (*bsd2*) lacks RuBisCO in both M and BS cells (Roth *et al*., 1996; Langdale and Kidner, 1994). Although *bsd2* plants lack the SSU and LSU of RuBisCO, mature transcripts accumulate in each cell-type. It is proposed that BSD2 acts as a co-translational chaperone for LSU (Brutnell *et al*., 1999). In addition to translational regulation by BSD2, RuBisCO is regulated post-translationally. The RuBisCO accumulation factor 1 (Raf1) protein is specific to BS chloroplasts in maize and is required for RuBisCO expression (Feiz *et al*., 2012; Friso *et al*., 2010). Mutants for *Raf1* transcribe *RbcS* and *rbcL* mRNAs and translate both subunits normally, but do not accumulate the RuBisCO holoenzyme and therefore die as seedlings.

It is thought that Raf1 acts by direct interaction with LSU to either free or insulate the LSU from chaperones during the early assembly process, and lack of Raf1 leads to a degradation of both LSU and SSU (Feiz *et al*., 2012).

Lastly, some C_4 cycle enzymes themselves are post-translationally chemically modified. PPDK is phosphorylated by pyruvate,orthophosphate dikinase regulatory protein (PDRP) in a dynamic light-responsive manner (Chen *et al*., 2014; Chastain *et al*., 2011). PDRP is capable of both phosphorylating and dephosphorylating PPDK at serine and threonine residues in the PPDK active site, which deactivate and activate its function respectively (Astley *et al*., 2011). PEPC is phosphorylated in several C_4 and C_3 species (Aldous *et al*., 2014; Nimmo *et al*., 2001). Additionally, the phosphorylation states of C_4 *Panicum maximum* PEPC and PEPCK are tightly coordinated with each other, despite being localised in M and BS cells respectively (Bailey *et al*., 2007). However to our knowledge, these modifications have not yet been associated with cell specificity, but are thought to adjust the C_4 cycle for different conditions within each cell type.

INDIVIDUAL C_4 GENES ARE CONTROLLED BY MULTIPLE REGULATORY MECHANISMS

The above analysis indicates that epigenetic, transcriptional, post-transcriptional, translational and post-translational regulatory mechanisms mediate preferential gene expression in either M or BS cells of C_4 leaves (Figure 2). However, these analyses of C_4 gene regulation are scattered across C_4 plants from a broad range of taxa. This makes it difficult to model the levels of regulation that are controlling each enzyme. Nevertheless, it is clear that C_4 genes within the same species are controlled at multiple levels. In addition to *GLDPA* (see above), evidence is provided by BS specific RuBisCO and M specific PEPC in maize (Table 1), two enzymes known to operate in all C_4 subtypes.

For example, accumulation of the RuBisCO holoenzyme in BS cells involves transcriptional, post-transcriptional, translational and post-translational mechanisms in maize (Berry *et al*., 2016). The small subunit is transcriptionally and post-transcriptionally regulated by *cis*-elements in its promoter and UTRs (Nomura *et al*., 2000; Viret *et al*., 1994). To ensure strong BS specificity, the stability of the large subunit transcript is regulated by the RNA binding protein RLSB (Bowman *et al*., 2013), translationally by BSD2 (Brutnell *et al*., 1999) and post-translationally by Raf1 (Feiz *et*

al., 2012). Additionally, the RuBisCO holoenzyme has a complex post-translational assembly process, mediated by chaperones, many of which are themselves BS specific, and influenced by the amount of both SSU and LSU (Hauser *et al*., 2015; Berry *et al*., 2016).

Likewise, maize *PEPC* appears to be regulated at multiple levels. It is epigenetically predisposed to cell specific expression through an enrichment of histone marks for actively transcribed genes, and through demethylation of DNA sites in its promoter region that are all enriched in M cells and responsive to light (Heimann *et al*., 2013; Tolley *et al*., 2011). Cell specificity is further ensured by transcriptional control mediated by *cis*-elements in the maize *PEPC* promoter that confer M specificity (Kausch *et al*., 2001; Taniguchi *et al*., 2000). These examples of PEPC and RuBisCO indicate that their cell specific expression patterns are underpinned by regulatory redundancy operating at multiple levels. This may well be the case for many more C_4 genes in multiple C_4 lineages.

Despite being a seemingly complex system, C_4 photosynthesis has also likely co-opted multiple ancestral regulatory mechanisms (Sage, 2004; Williams *et al*., 2013). This is evidenced by C_4 cycle genes from C_3 species already being regulated by light and chloroplast signalling networks (Burgess *et al.,* 2016), but also C_4 genes from independent C_4 lineages sharing regulatory mechanisms, including a shared epigenetic histone code for *NADP-ME, PEPCK, PEPC, CA*, and *PPDK* (Heimann *et al*., 2013) and shared post-transcriptional *cis*-elements for *PPDK* and *CA* (Williams, Burgess *et al*., 2016; Hibberd and Covshoff, 2010). Furthermore, certain C_4 cycle genes may have gained cell specificity from pre-existing regulatory mechanisms, as in some cases, orthologous genes in C_3 plants contain *cis*-elements that confer cell-specificity in C_4 leaves (Williams *et al*., 2012; Kajala *et al*., 2012; Brown *et al*., 2011), whereas in others, they are regulated correctly when expressed in C_3 congeners (Nomura *et al*., 2005; Matsuoka *et al*., 1994; Ku *et al*., 1999). This is an indication that *cis*-elements, *trans*-factors and other regulators for C_4 photosynthesis are present in C_3 species.

CONCLUSION

The spatial patterning of gene expression required in the C_4 leaf is determined by a variety of mechanisms. Although there are few examples where an individual gene has been catalogued as being controlled at multiple levels of regulation, the above summary indicates that this may well be true for many genes. It is not yet clear the extent to which these regulatory mechanisms operate in C_3 plants, nor how many changes in *cis*- and *trans*-regulation are responsible for C_4 enzymes to be partitioned between mesophyll and bundle sheath cells.

Table I: C_4 genes in maize for which there is evidence that cell-specificity is generated from multiple levels of regulation.

Cell type	C_4 gene	Type of regulation		References
Mesophyll	Phospho*enol*pyruvate carboxylase (*PEPC*)	Epigenetic regulation	Histone acetylation and methylation, as well as DNA methylation of promoter region	Heimann *et al*., 2013; Perduns *et al*., 2015; Tolley *et al*., 2011
		Transcriptional regulation	*Cis*-elements in promoter region	Kausch *et al*., 2001; Taniguchi *et al*., 2000
Bundle Sheath	Small subunit of RuBisCO (*RbcS*)	Transcriptional and post-transcriptional regulation	*Cis*-elements in promoter region and UTRs	Nomura *et al*., 2000; Viret *et al*., 1994
	Large subunit of RuBisCO (*rbcL*)	Post-transcriptional regulation	RNA stability mediated by RNA binding protein RLSB	Bowman *et al*., 2013
		Translational regulation	Interaction with BSD2	Brutnell *et al*., 1999
		Post-translational regulation	Interaction with Raf1	Feiz *et al*., 2012

FIGURE LEGENDS

Figure 1. Biochemical and anatomical hallmarks of C_4 photosynthesis. (A) A unified cycle for the three decarboxylation subtypes of the C_4 carbon assimilation cycle. Enzyme abbreviations: *CA*, *carbonic anhydrase*; *PEPC*, *PHOSPHO*ENOL*PYRUVATE CARBOXYLASE*; *ASPAT*, *ASPARTATE AMINOTRANSFERASE*; *ALAAT*, *ALANINE AMINOTRANSFERASE*; *PEPCK*, *PHOSPHO*ENOL*PYRUVATE CARBOXYKINASE*; *NAD-MDH*, *NAD-MALATE DEHYDROGENASE*; *NADP-MDH*, *NADP-MALATE DEHYDROGENASE*; *NAD-ME*, *NAD-MALIC ENZYME*, *NADP-ME*, *NADP-MALIC ENZYME*; *PPDK*, *PYRUVATE ORTHOPHOSPHATE DIKINASE*; *RUBISCO*, *RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE OXYGENASE*. Metabolite abbreviations: OAA, oxaloacetate; Asp, aspartate; Ala, alanine; Pyr, pyruvate; M, malate; PGA, 3-phosphoglyceric acid; PEP, phospho*enol*pyruvate. Coloured dots represent the number of carbon atoms present in each metabolite. Red dots represent the path of carbon to RuBisCO in one round of the C_4 cycle. (B) Anatomical differences between C_4 and C_3 leaves. Kranz anatomy in C_4 *Gynandropsis gynandra* (left) and typical non-Kranz anatomy in C_3 *Tarenaya hassleriana* (right) are shown. Leaves were fixed in resin, sectioned, stained in Toluidine-Blue and imaged under light-microscopy. Scale bar = 100µm. Cell type abbreviation: BS, bundle sheath; M, mesophyll.

Figure 2. Multiple regulatory mechanisms control cell specific expression of C_4 genes. Genes listed under each regulatory checkpoint are known to be controlled at that step, as discovered in specific C_4 species indicated by superscripted numbers. C_4 gene abbreviations: *CA*, *CARBONIC ANHYDRASE; PEPC, PHOSPHO*ENOL*PYRUVATE CARBOXYLASE; PEPCK, PHOSPHO*ENOL*PYRUVATE CARBOXYKINASE; PPDK, PYRUVATE ORTHOPHOSPHATE DIKINASE; NAD-ME, NAD-MALIC ENZYME, NADP-ME, NADP-MALIC ENZYME*; RBCS, SMALL SUBUNIT OF RUBISCO; RBCL, LARGE SUBUNIT OF RUBISCO; GLDPA, P-SUBUNIT OF THE GLYCINE DECARBOXYLATION COMPLEX; *psbB-psbT-psbH-petB-petD*, polycistron encoding components of Photosystem II.

REFERENCES

Aldous SH, Weise SE, Sharkey TD, Waldera-Lupa DM, Stuhler K, Mallmann J, Groth G, Gowik U, Westhoff P, Arsova B. 2014. Evolution of the phospho*enol*pyruvate carboxylase protein kinase family in C_3 and C_4 *Flaveria* spp. *Plant Physiology* **165,** 1076–1091.

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

Astley HM, Parsley K, Aubry S, Chastain CJ, Burnell JN, Webb ME, Hibberd JM. 2011. The pyruvate, orthophosphate dikinase regulatory proteins of *Arabidopsis* are both bifunctional and interact with the catalytic and nucleotide-binding domains of pyruvate, orthophosphate dikinase. *The Plant Journal* **68**, 1070–1080.

Aubry S, Kelly S, Kümpers BM, Smith-Unna RD, Hibberd JM. 2014. Deep evolutionary comparison of gene expression identifies parallel recruitment of *trans*-factors in two independent origins of C_4 photosynthesis. *PLoS Genetics* **10,** e1004365.

Bailey KJ, Gray JE, Walker RP, Leegood RC. 2007. Coordinate regulation of phosphoenolpyruvate carboxylase and phospho*enol*pyruvate carboxykinase by light and CO_2 during C_4 photosynthesis. *Plant Physiology* **144,** 479–486.

Bauwe H, Hagemann M, Fernie AR. 2010. Photorespiration: Players, partners and origin. *Trends in Plant Science* **15,** 330–336.

Berry JO, Mure CM, Yerramsetty P. 2016. Regulation of Rubisco gene expression in C_4 plants. *Current Opinion in Plant Biology* **31,** 23-28.

Boinski JJ, Wang JL, Xu P, Hotchkiss T, Berry JO. 1993. Post-transcriptional control of cell type-specific gene expression in bundle sheath and mesophyll chloroplasts of *Amaranthus hypochondriacus*. *Plant Molecular Biol*ogy **22,** 397-410.

Bowes G, Ogren WL, Hageman RH. 1971. Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochemical and Biophysical Research Communications* **45,** 716–722.

Bowman SM, Patel M, Yerramsetty P, Mure CM, Zielinski AM, Bruenn JA, Berry JO. 2013. A novel RNA binding protein affects *rbcL* gene expression and is specific to bundle sheath chloroplasts in C_4 plants. *BMC Plant Biology* **13,** 138.

Bräutigam A, Kajala K, Wullenweber J, Sommer M, Gagneul D, Weber KL, Carr KM, Gowik U, Maß J, Lercher MJ, Westhoff P, Hibberd JM, Weber APM. 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.

Bräutigam A, Weber AP. 2011. Do metabolite transport processes limit photosynthesis? *Plant Physiology* **155,** 43-48.

Brown NJ, Newell CA, Stanley S, Chen JE, Perrin AJ, Kajala K, Hibberd JM. 2011. Independent and parallel recruitment of preexisting mechanisms underlying C_4 photosynthesis. *Science* **331,** 1436–1439.

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

Brutnell TP, Sawers RJ, Mant A, Langdale JA. 1999. BUNDLE SHEATH DEFECTIVE2, a novel protein required for post-translational regulation of the *rbcL* gene of maize. *The Plant Cell* **11,** 849–864.

Burgess SJ, Granero-Moya I, Grangé-Guermente MJ, Boursnell C, Terry MJ, Hibberd JM. 2016. Ancestral light and chloroplast regulation form the foundations for C_4 gene expression. *Nature Plants* doi: 10.1038/NPLANTS.2016.161

Calvin M, Benson AA. 1948. The path of carbon in photosynthesis. *Science* **107,** 476-480.

Cavalar M, Phlippen Y, Kreuzaler F, Peterhansel C. 2007. A drastic reduction in DOF1 transcript levels does not affect C_4-specific gene expression in maize. *Journal of Plant Physiology* **164,** 1665–1674.

Chang YM, Liu WY, Shih ACC, Shen MN, Lu CH, Lu MYJ, Yang HW, Wang TY, Chen SCC, Chen SM, Li WH, Ku MSB. 2012. Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. *Plant Physiology* **160,** 165-177.

Chastain CJ, Failing CJ, Manandhar L, Zimmerman MA, Lakner MM, Nguyen TH. 2011. Functional evolution of C_4 pyruvate, orthophosphate dikinase. *Journal of Experimental Botany* **62,** 3083–3091.

Chen YB, Lu TC, Wang HX, Shen J, Bu TT, Chao Q, Gao ZF, Zhu XG, Wang YF, Wang BC. 2014. Posttranslational modification of maize chloroplast pyruvate orthophosphate dikinase reveals the precise regulatory mechanism of its enzymatic activity. *Plant Physiology* **165,** 534–549.

Chi W, Ma J, Zhang L. 2012. Regulatory factors for the assembly of thylakoid membrane protein complexes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **367,** 3420–3429.

Cohen A, Mayfield SP. 1997. Translational regulation of gene expression in plants. *Current Opinion in Biotechnology* **8,** 189-194.

Covshoff S, Hibberd JM. 2012. Integrating C_4 photosynthesis into C_3 crops to increase yield potential. *Current Opinion in Biotechnology* 23, 209-214.

Covshoff S, Majeran W, Liu P, Kolkman JM, van Wijk KJ, Brutnell TP. 2008. Deregulation of maize C_4 photosynthetic development in a mesophyll cell-defective mutant. *Plant Physiology* **146,** 1469–1481.

Covshoff S, Szecowka M, Hughes TE, Smith-Unna R, Kelly S, Bailey KJ, Sage TL, Pachebat JA, Leegood R, Hibberd JM. 2016. C_4 Photosynthesis in the Rice Paddy: Insights from the Noxious Weed *Echinochloa glabrescens*. *Plant Physiology* **170,** 57–73.

Cui H, Kong D, Liu X, Hao Y. 2014. SCARECROW, SCR-LIKE 23 and SHORT-ROOT control bundle sheath cell fate and function in *Arabidopsis thaliana*. *Plant Journal* **78,** 319–327.

Ding Z, Weissmann S, Wang M, Du B, Huang L, Wang L, Tu X, Zhong S, Myers C, Brutnell TP, Sun Q, Li P. 2015. Identification of Photosynthesis-Associated C_4 Candidate Genes through Comparative Leaf Gradient Transcriptome in Multiple Lineages of C_3 and C_4 Species. *PLoS One* **10,** e0140629.

Dion MF, Altschuler SJ, Wu LF, Rando OJ. 2005. Genomic characterization reveals a simple histone H4 acetylation code. *Proceedings of the National Academy of Sciences USA* **102,** 5501–5506.

Downton WJS, Tregunna EB. 1968. Carbon dioxide compensation-its relation to photosynthetic carboxylation reactions, systematics of the Gramineae, and leaf anatomy. *Canadian Journal of Botany* **46**, 207–215.

El-Sharkawy M, Hesketh J. 1965. Photosynthesis among species in relation to characteristics of leaf anatomy and CO_2 diffusion resistances. *Crop Science* **5,** 517–521.

Engelmann S, Wiludda C, Burscheidt J, Gowik U, Schlue U, Koczor M, Streubel M, Cossu R, Bauwe H, Westhoff P. 2008. The gene for the P-subunit of glycine decarboxylase from the C_4 species *Flaveria trinervia*: Analysis of transcriptional control in transgenic *Flaveria bidentis* (C_4) and Arabidopsis (C_3). *Plant Physiology* **146,** 1773–1785.

Feiz L, Williams-Carrier R, Wostrikoff K, Belcher S, Barkan A, Stern DB. 2012. Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation factor1 is required for holoenzyme assembly in maize. *The Plant Cell* **24,** 3435-46.

Friso G, Majeran W, Huang M, Sun Q, van Wijk KJ. 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.

Feng S, Jacobsen SE, Reik W. 2010. Epigenetic reprogramming in plant and animal development. *Science* **330,** 622–627.

Furbank RT. 2016. Walking the C_4 pathway: past, present, and future. *Journal of Experimental Botany* doi:10.1093/jxb/erw161.

Gao X, Wang C, Cui H. 2014. Identification of bundle sheath cell fate factors provides new tools for C_3-to-C_4 engineering. *Plant Signaling and Behavior* **9**, e29162.

Gowik U, Burscheidt J, Akyildiz M, Schlue U, Koczor M, Streubel M, 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. *The Plant Cell* **16,** 1077–1090.

Gowik U, Brautigam A, Weber KL, Weber AP, Westhoff P. 2011. Evolution of C_4 photosynthesis in the genus *Flaveria*: How many and which genes does it take to make C_4? *The Plant Cell* **23,** 2087-2105.

Haberlandt, G. 1904. *Physiologische Pflanzenanatomie* (Engelmann).

Holland HD. 2006. The oxygenation of the atmosphere and oceans. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **361,** 903–915.

Hatch MD. 1987. C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta* **895,** 81–106.

Hatch MD, Slack CR. 1966. Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. *Biochemical Journal* **101,** 103-111.

Hauser T, Popilka L, Hartl FU, Hayer-Hartl M. 2015. Role of auxiliary proteins in Rubisco biogenesis and function. *Nature Plants* doi: 10.1038/nplants.2015.65.

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

Hibberd JM, Covshoff S. 2010. The regulation of gene expression required for C_4 photosynthesis. *Annual Review of Plant Biology* **61,** 181–207.

Hibberd JM, Sheehy JE, Langdale JA. 2008. Using C_4 photosynthesis to increase the yield of rice—rationale and feasibility. *Current Opinion in Plant Biology* **11,** 228-231.

Järvi S, Suorsa M, Aro E-M. 2015. Photosystem II repair in plant chloroplasts — Regulation, assisting proteins and shared components with photosystem II biogenesis. *Biochimica et Biophysica Acta* **1849,** 900-909.

Jensen PE, Bassi R, Boekema EJ, Dekker JP, Jansson S, Leister D, Robinson C, Henrik Vibe Scheller HV. 2007. Structure, function and regulation of plant photosystem I. *Biochimica et Biophysica Acta* **1767,** 335-352.

John CR, Smith-Unna RD, Woodfield H, Covshoff S, Hibberd JM. 2014. Evolutionary convergence of cell-specific gene expression in independent lineages of C-4 grasses. *Plant Physiology* **165,** 62–75.

Kajala K, Brown NJ, Williams BP, Borrill P, Taylor LE, Hibberd JM. 2012. Multiple Arabidopsis genes primed for recruitment into C_4 photosynthesis. *Plant Journal* **69,** 47–56.

Kano-Murakami Y, Suzuki I, Sugiyama T, Matsuoka M. 1991. Sequence-specific interactions of a maize factor with a GC-rich repeat in the phosphoenolpyruvate carboxylase gene. *Molecular and General Genetics* **225,** 203-208.

Kagawa T, Hatch MD. 1974. C_4-acids as the source of carbon dioxide for Calvin cycle photosynthesis by bundle sheath cells of the C_4-pathway species *Atriplex spongiosa*. *Biochemical and Biophysical Research Communications* **59,** 1326–1332.

Kausch AP, Owen TP, Zachwieja SJ, Flynn AR, Sheen J. 2001. Mesophyll-specific, light and metabolic regulation of the C_4 *PPCZm1* promoter in transgenic maize. *Plant Molecular Biology* **45,** 1-15.

Ku MS, Agarie S, Nomura M, Fukayama H, Tsuchida H, Ono K, Hirose S, Toki S, Miyao M, Matsuoka M. 1999. High-level expression of maize phospho*enol*pyruvate carboxylase in transgenic rice plants. *Nature Biotechnology* **17,** 76-80.

Kubicki A, Steinmüller K, Westhoff P. 1994. Differential transcription of plastome-encoded genes in the mesophyll and bundle-sheath chloroplasts of the monocotyledonous NADP-malic enzyme-type C_4 plants maize and Sorghum. *Plant Molecular Biology* **25,** 669-679.

Langdale JA, Taylor WC, Nelson T. 1991. Cell-specific accumulation of maize phospho*enol*pyruvate carboxylase is correlated with demethylation at a specific site >3 kb upstream of the gene. *Molecular and General Genetics* **225,** 49-55.

Langdale JA, Kidner CA. 1994. Bundle sheath defective, a mutation that disrupts cellular differentiation in the maize leaves. *Development* **120,** 673–681.

Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, Reidel EJ, Turgeon R, Liu P, Sun Q, Nelson T, Brutnell TP. 2010. The developmental dynamics of the maize leaf transcriptome. *Nature Genetics* **42,** 1060-1067.

Loidl P. 2004. A plant dialect of the histone language. *Trends in Plant Science* **9,** 84-90.

Luo G, Ono S, Beukes NJ, Wang DT, Xie S, Summons RE. 2016. Rapid oxygenation of Earth's atmosphere 2.33 billion years ago. *Science Advances* **2,** e1600134.

Matsuoka M, Kyozuka J, Shimamoto K, Kano-Murakami Y. 1994. The promoters of two carboxylases in a C_4 plant (maize) direct cell-specific, light-regulated expression in a C_3 plant (rice). *Plant Journal* **22,** 211-221.

Matsuoka M, Furbank RT, Fukuyama H, Miyao M. 2001. Molecular engineering of C_4 photosynthesis. *Annual Review of Plant Physiology and Molecular Biology* **52,** 297-314.

Meierhoff K, Felder S, Nakamura T, Bechtold N, Schuster G. 2003. HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast *psbB-psbT-psbH-petB-petD* RNAs. *The Plant Cell* **15,** 1480–1495.

Mauseth JD. 2012. Plants and People. *Jones & Bartlett Publishers* ISBN 9780763785505.

Nakamura T, Meierhoff K, Westhoff P, Schuster G. 2003. RNA-binding properties of HCF152, an Arabidopsis PPR protein involved in the processing of chloroplast RNA. *European Journal of Biochemistry* **270,** 4070–4081.

Nelson DL, Cox MM. 2008. Principles of biochemistry. Fifth ed. *W.H. Freeman & Company*. ISBN: 1429222638.

Nimmo H, Fontaine V, Hartwell J, Jenkins G, Nimmo G, Wilkins M. 2001. PEP carboxylase kinase is a novel protein kinase controlled at the level of expression. *New Phytologist* **151,** 91-97.

Nomura M, Higuchi T, Ishida Y, Ohta S, Komari T, Imaizumi N, Miyao-Tokutomi M, Matsuoka M, Tajima S. 2005. Differential expression pattern of C_4 bundle sheath expression genes in rice, a C_3 plant. *The Plant Cell Physiology* **46,** 754-761.

Nomura M., Katayama K., Nishimura A., Ishida Y., Ohta S., Komari T., Miyao-Tokutomi M., Tajima S., Matsuoka M. 2000. The promoter of *rbcS* in a C_3 plant (rice) directs organ-specific, light-dependent expression in a C_4 plant (maize), but does not confer bundle sheath cell-specific expression. *Plant Molecular Biology* **44,** 99–106.

Offermann S, Dreesen B, Horst I, Danker T, Jaskiewicz M, Peterhansel C. 2008. Developmental and environmental signals induce distinct histone acetylation profiles on distal and proximal promoter elements of the C_4- Pepc gene in maize. *Genetics* **179,** 1891–1901.

Osborne CP, Beerline DJ. 2006. Nature's green revolution: the remarkable evolutionary rise of C_4 plants. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **361,** 173-194.

Park DH, Lim PO, Kim JS, Cho DS, Hong SH, Nam HG. 2003. The Arabidopsis COG1 gene encodes a Dof domain transcription factor and negatively regulates phytochrome signaling. *Plant Journal* **34,** 161-71.

Patel M, Berry JO. 2008. Rubisco gene expression in C_4 plants. *Journal of Experimental Botany* **59,** 1625–1634.

Patel M, Corey AC, Yin LP, Ali S, Taylor WC, Berry JO. 2004. Untranslated regions from C_4 amaranth *AhRbcS1* mRNAs confer translational enhancement and preferential bundle sheath expression in transgenic C_4 *Flaveria bidentis*. *Plant Physiology* **136,** 3550-3561.

Patel M, Siegel AJ, Berry JO. 2006. Untranslated regions *FbRbcS1* mRNA mediate bundle sheath cell-specific expression in leaves of a C_4 plant. *Journal of Biological Chemistry* **281,** 25485-25491.

Perduns R, Horst-Niessen I, Peterhansel C. 2015 Photosynthetic Genes and Genes Associated with the C_4 Trait in Maize Are Characterized by a Unique Class of Highly Regulated Histone Acetylation Peaks on Upstream Promoters. *Plant Physiology* **168,** 1378-1388.

Peterhansel C, Offerman S. 2012. Re-engineering of carbon fixation in plants—challenges for plant

biotechnology to improve yields in a high-CO_2 world. *Current Opinion in Biotechnology* **23**, 204-208.

Pick TR, Bräutigam A, Schlüter U, Denton AK, Colmsee C, Scholz U, Fahnenstich H, Pieruschka R, Rascher U, Sonnewald U, Weber APM. 2011. Systems analysis of a maize leaf developmental gradient redefines the current C_4 model and provides candidates for regulation. *The Plant Cell* **23,** 4208-4220.

Portis AR Jr, Parry MA. 2007. Discoveries in Rubisco (Ribulose 1,5-bisphosphate carboxylase / oxygenase): A historical perspective. *Photosynthesis Research* **94,** 121–143.

Raines CA. 2011. Increasing photosynthetic carbon assimilation in C_3 plants to improve crop yield: Current and future strategies. *Plant Physiology 155,* 36-42.

Rao X, Lu N, Li G, Nakashima J, Tang Y, Dixon RA. 2016. Comparative cell-specific transcriptomics reveals differentiation of C_4 photosynthesis pathways in switchgrass and other C_4 lineages. *Journal of Experimental Botany* **67,** 1649-1662.

Rossini L, Cribb L, Martin DJ, Langdale JA. 2001. The maize golden2 gene defines a novel class of transcriptional regulators in plants. *The Plant Cell* **13,** 1231–1244.

Roth R, Hall LN, Brutnell TP, Langdale JA. 1996. Bundle sheath defective2, a mutation that disrupts the coordinated development of bundle sheath and mesophyll cells in the maize leaf. *The Plant Cell* **8,** 915–927.

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

Sage R. 2016. A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *Journal of Experimental Botany* DOI: 10:1093/jxb/erw156.

Sage RF, Sage TL, Kocacinar F. 2012. Photorespiration and the evolution of C_4 photosynthesis. *Annual Review of Plant Biology* **63,** 19–47.

Sage RF, Zhu XG. 2011. Exploiting the engine of C_4 photosynthesis. *Journal of Experimental Botany* **69,** 2989-3000.

Schöttler MA, Tóth SZ, Boulouis A, Kahlau S. 2015. Photosynthetic complex stoichiometry dynamics in higher plants: biogenesis, function, and turnover of ATP synthase and the cytochrome b6f complex. *Journal of Experimental Botany* **66,** 2373-2400.

Schuster G, Ohad I, Martineau B, Taylor WC. 1985. Differentiation and development of bundle sheath and mesophyll thylakoids in maize. Thylakoid polypeptide composition, phosphorylation, and organization of photosystem II. *Journal of Biological Chemistry* **260,** 11866–11873.

Sharkey T. 1988. Estimating the rate of photorespiration in leaves. *Physiologia Plantarum* **73,** 147–152.

Sheen J-Y, 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.

Slewinski TL, Anderson AA, Zhang C, Turgeon R. 2012. Scarecrow plays a role in establishing Kranz anatomy in maize leaves. *Plant and Cell Physiology* **53,** 2030–2037.

Spreitzer RJ, Salvucci ME. 2002. Rubisco: structure, regulatory interactions and possibilities for a better enzyme. *Annual Review of Plant Biology* **53,** 449–475.

Stockhaus J, Schlue U, Koczor M, Chitty JA, Taylor WC, Westhoff P. 1997. The promoter of the gene encoding the C_4 form of phospho*enol*pyruvate carboxylase directs mesophyll specific expression in transgenic C_4 *Flaveria* spp. *The Plant Cell* **9,** 479–489.

Taniguchi M, Izawa K, Ku MSB, Lin JH, Saito H, Ishida Y, Ohta S, Komari T, Matsuoka M, Sugiyama T. 2000. Binding of cell type-specific nuclear proteins to the 5'-flanking region of maize C_4 phospho*enol*pyruvate carboxylase gene confers its differential transcription in mesophyll cells. *Plant Molecular Biology* **44,** 543-557.

Tausta SL, Li P, Si Y, Gandotra N, Liu P, Sun Q, Brutnell TP, Nelson T. 2014. Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C_4-related processes. *Journal of Experimental Botany* **65,** 3543-3555.

Tolley BJ, Woodfield H, Wanchana S, Bruskiewich R, Hibberd JM. 2011. Light-regulated and cell-specific methylation of the maize PEPC promoter. *Journal of Experimental Botany* **63,** 1381-1390.

Vanyushin BF, Ashapkin VV. 2011. DNA methylation in higher plants: past, present and future. *Biochimica et Biophysica Acta* **1809,** 360-368.

Viret J-F, Mabrouk Y, Bogorad L. 1994. Transcriptional photoregulation of cell-type-preferred expression of maize rbcSm3: 3' and 5' sequences are involved. *Proceedings of the National Academy of Sciences USA* **91**, 8577–8581.

von Caemmerer S, Quick WP, Furbank RT. 2012. The development of C_4 rice: Current progress and future challenges. *Science* **336,** 1671-1672.

Wang X, Elling AA, Li X, Li N, Peng Z, He G, Sun H, Qi Y, Liu XS, Deng XW. 2009. Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *The Plant Cell* **21**, 1053–1069.

Wang L, Peterson RB, Brutnell TP. 2011. Regulatory mechanisms underlying C_4 photosynthesis. *New Phytologist* **190,** 9-20.

Wang P, Fouracre J, Kelly S, Karki S, Gowik U, Aubry S, Shaw MK, Westhoff P, Slamet-Loedin IH, Quick WP, Hibberd JM, Langdale JA. 2013a. Evolution of GOLDEN2-LIKE gene function in C_3 and C_4 plants. *Planta* **237,** 481-495.

Wang P, Kelly S, Fouracre JP, Langdale JA. 2013b. Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C_4 Kranz anatomy. *The Plant Journal* **75,** 656–670.

Wang L, Czedik-Eysenberg A, Mertz RA, Si Y, Tohge T, Nunes-Nesi A, *et al.* 2014. Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice. *Nature Biotechnology* **32,** 1158–1165.

Ward JM, Cufr CA, Denzel MA, Neff MM. 2005. The Dof transcription factor OBP3 modulates phytochrome and cryptochrome signaling in Arabidopsis. *The Plant Cell* **17,** 475-85.

Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, Langdale JA. 2009. GLK transcription factors coordinate expression of the photosynthetic apparatus in Arabidopsis. *The Plant Cell* **21,** 1109–1128.

Westoff P, Gowik U. 2004. Evolution of C_4 phospho*enol*pyruvate carboxylase. Genes and proteins: A case study with the genus *Flaveria*. *Annals of Botany* **93,** 13–23.

Windhövel A, Hein I, Dabrowa R, Stockhaus J. 2001. Characterization of a novel class of plant homeodomain proteins that bind to the C4 phosphoenolpyruvate carboxylase gene of *Flaveria trinervia*. *Plant Molecular Biology* **45,** 201-14.

Whitney SM, Houtz RL, Alonso H. 2011. Advancing our understanding and capacity to engineer nature's CO_2-sequestering enzyme, Rubisco. *Plant Physiology* **155,** 27-35.

Williams BP, Aubry S, Hibberd JM. 2012. Molecular evolution of genes recruited into C_4 photosynthesis. *Trends in Plant Science* **17,** 1360-1385.

Williams BP, Johnston IG, Covshoff S, Hibberd JM. 2013. Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis. *eLife* **2,** e00961.

Williams BP, Burgess SJ, Reyna-Llorens I, Knerova J, Aubry S, Stanley S, Hibberd JM. 2016. An untranslated *cis*-element regulates the accumulation of multiple C_4 enzymes in *Gynandropsis gynandra* mesophyll cells. *The Plant Cell* **28**, 454-465.

Wiludda C, Schulze S, Gowik U, Engelmann S, Koczor M, Streubel M, Bauwe H, Westhoff P. 2012. Regulation of the photorespiratory *GLDPA* gene in C_4 *Flaveria*: an intricate interplay of transcriptional and posttranscriptional processes. *The Plant Cell* **24,** 137–151

Wostrikoff K, Clark A, Sato S, Clemente T, Stern D. 2012. Ectopic expression of Rubisco subunits in maize mesophyll cells does not overcome barriers to cell type-specific accumulation. *Plant Physiology* **160**, 419-32.

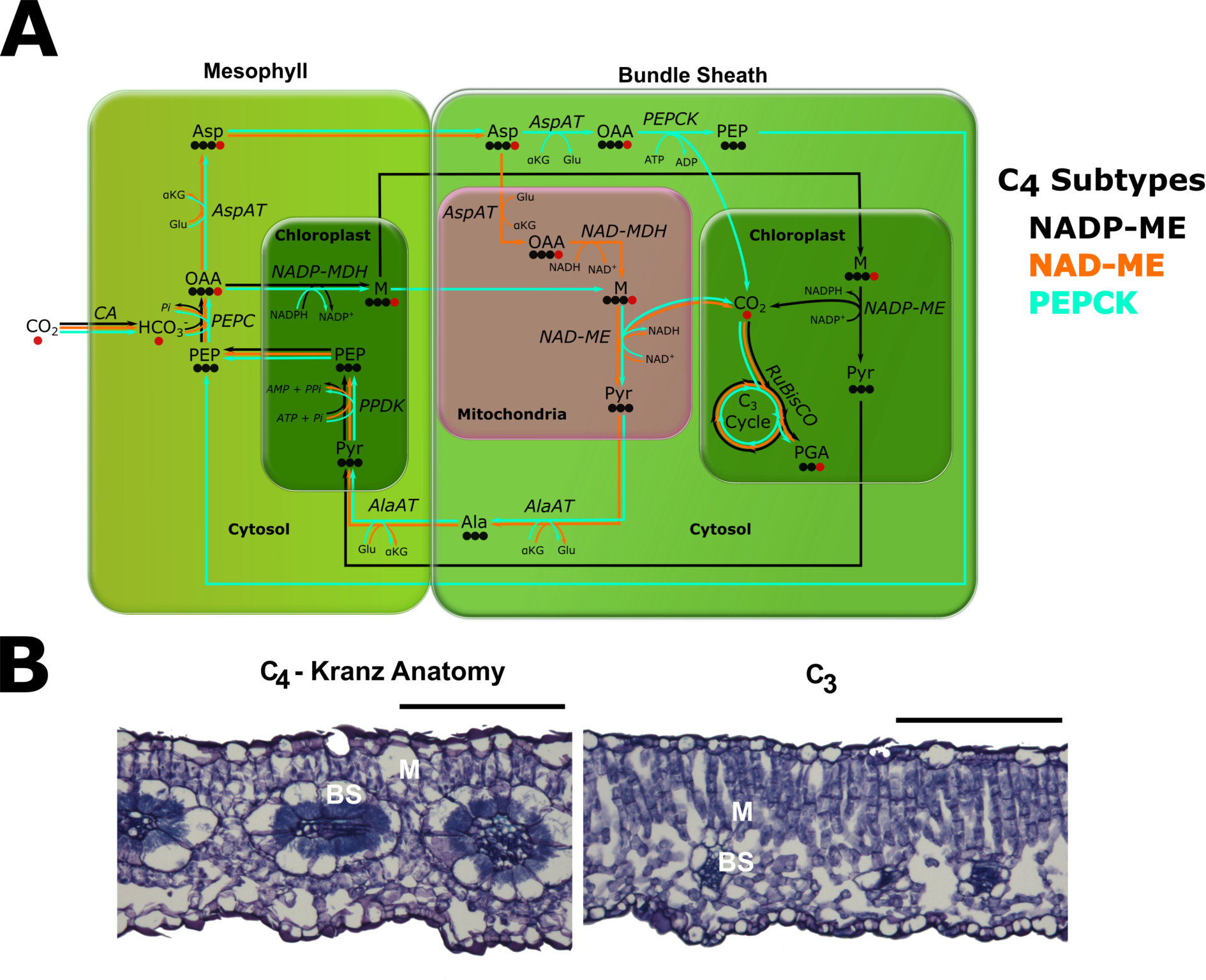
Yerramsetty P, Stata M, Siford R, Sage TL, Sage RF, Wong GK, Albert VA, Berry JO. 2016. Evolution of RLSB, a nuclear-encoded S1 domain RNA binding protein associated with post-transcriptional regulation of plastid-encoded rbcL mRNA in vascular plants. *BMC Evolutionary Biology* **16,** 141.

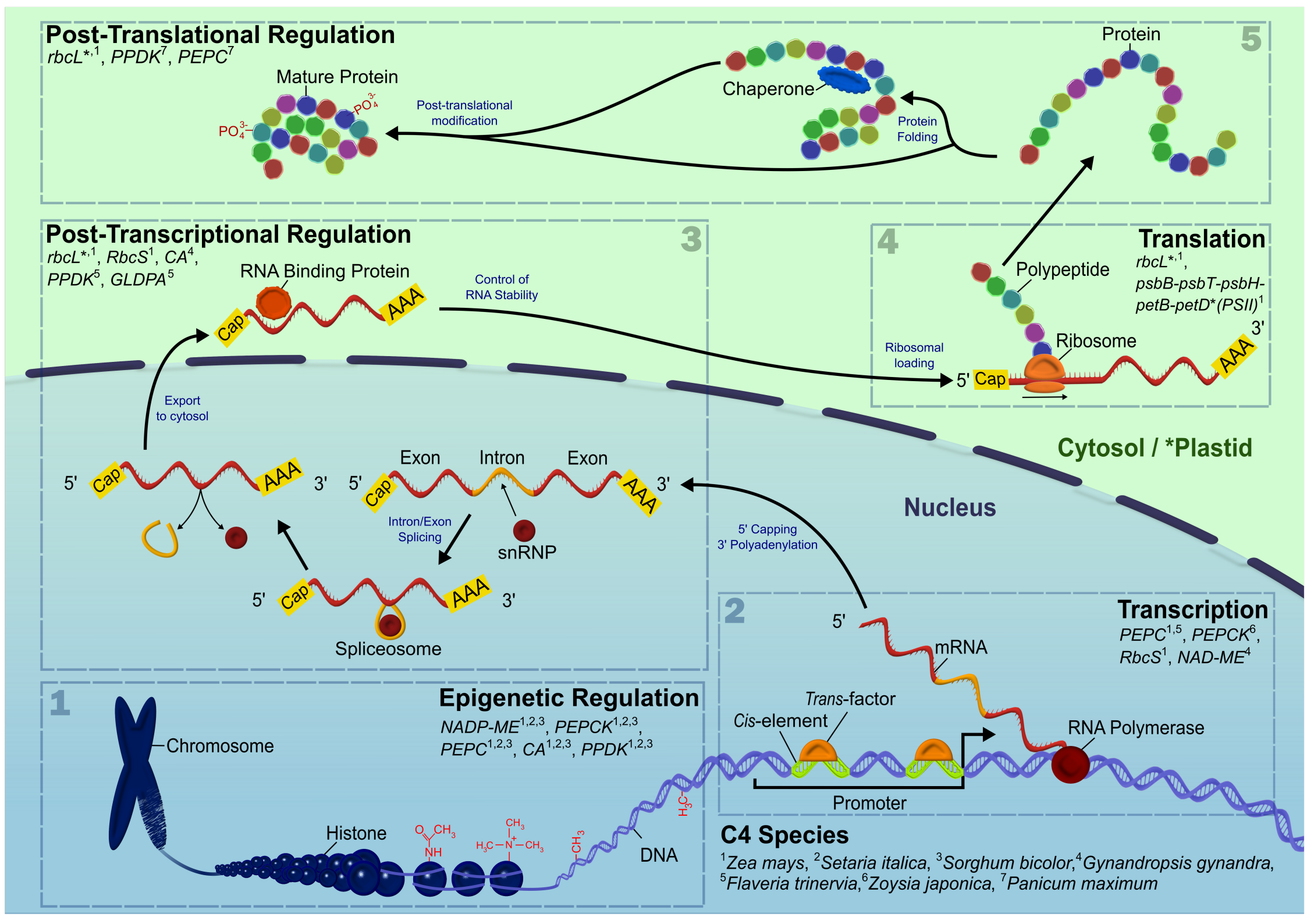
Yanagisawa S, Sheen J. 1998. Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. *The Plant Cell* **10,** 75-89.

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.

Yu CP, Chen SC, Chang YM, Liu WY, Lin HH, Lin JJ, Chen HJ, Lu YJ, Wu YH, Lu MY, *et al*. 2015. Transcriptome dynamics of developing maize leaves and genome wide prediction of *cis*-elements and their cognate transcription factors. *Proceedings of the National Academy of Sciences USA* **112,** 2477–2486.

Zhang Y, Verhoeff NI, Chen Z, Chen S, Wang M, Zhu Z, Ouwerkerk PBF. 2015. Functions of *OsDof25* in regulation of *OsC4PPDK*. *Plant Molecular Biology* **89,** 229–242.


A
Mesophyll
Bundle Sheath
Asp
AspAT
αKG
Glu
Chloroplast
NADP-MDH
OAA
M
NADPH
NADP+
CA
CO2
HCO3-
Pi
PEPC
PEP
AMP + PPi
ATP + Pi
PPDK
Pyr
AlaAT
Ala
Cytosol
Glu
αKG
PEPCK
ATP
ADP
NAD-MDH
NADH
NAD+
NAD-ME
Mitochondria
NADP-ME
RuBisCO
C3 Cycle
PGA
C4 Subtypes
NADP-ME
NAD-ME
PEPCK
B
C4 - Kranz Anatomy
C3
M
BS


Post-Translational Regulation
rbcL*,1, PPDK7, PEPC7
Mature Protein
PO4 3-
Post-translational modification
Chaperone
Protein Folding
Protein
5
Post-Transcriptional Regulation
rbcL*,1, RbcS1, CA4, PPDK5, GLDPA5
3
RNA Binding Protein
Cap
AAA
Control of RNA Stability
4
Translation
rbcL*,1, psbB-psbT-psbH-petB-petD*(PSII)1
Polypeptide
Ribosome
Ribosomal loading
5' Cap
3'
Export to cytosol
Cytosol / *Plastid
Nucleus
Exon
Intron
Exon
Intron/Exon Splicing
snRNP
Spliceosome
5' Capping
3' Polyadenylation
2
Transcription
PEPC1,5, PEPCK6, RbcS1, NAD-ME4
mRNA
Trans-factor
Cis-element
RNA Polymerase
Promoter
1
Epigenetic Regulation
NADP-ME1,2,3, PEPCK1,2,3, PEPC1,2,3, CA1,2,3, PPDK1,2,3
Chromosome
Histone
DNA
C4 Species
1Zea mays, 2Setaria italica, 3Sorghum bicolor, 4Gynandropsis gynandra, 5Flaveria trinervia, 6Zoysia japonica, 7Panicum maximum