An untranslated *cis*-element regulates the accumulation of multiple C_4 enzymes in *Gynandropsis gynandra* mesophyll cells

Ben P. Williams[1], Steven J. Burgess[1], Ivan Reyna-Llorens, Jana Knerova, Sylvain Aubry, Susan Stanley and Julian M. Hibberd*

Department of Plant Sciences, University of Cambridge, U.K.

* Correspondence to jmh65@cam.ac.uk

[1]These authors contributed equally to this work

Synopsis: This work identifies a short sequence motif that regulates mesophyll specific accumulation of multiple enzymes used in C_4 photosynthesis.

ABSTRACT

C_4 photosynthesis is a complex phenotype that allows more efficient carbon capture than the ancestral C_3 pathway. In leaves of C_4 species, hundreds of transcripts increase in abundance compared with C_3 relatives and become restricted to mesophyll (M) or bundle sheath (BS) cells. However, no mechanism has been reported that regulates the compartmentation of multiple enzymes in M or BS cells. We examined mechanisms regulating *CARBONIC ANHYDRASE 4* (*CA4*) in C_4 *Gynandropsis gynandra*. Increased abundance is directed by both the promoter region and introns of the *G. gynandra* gene. A 9-nucleotide motif located in the 5' untranslated region (UTR) is required for preferential accumulation of GUS in M cells. This element is present and functional in three additional 5' UTRs and six 3' UTRs where it determines accumulation of two isoforms of CA and pyruvate,orthophosphate dikinase in M cells. Although the *GgCA4* 5' UTR is sufficient to direct GUS accumulation in M cells, transcripts encoding GUS are abundant in both M and BS. Mutating the *GgCA4* 5' UTR abolishes enrichment of protein in M cells without affecting transcript abundance. The work identifies a mechanism that directs cell-preferential accumulation of multiple enzymes required for C_4 photosynthesis.

INTRODUCTION

The ability to restrict synthesis of specific proteins to individual cell types is essential to the development of multicellular organisms. Understanding the molecular mechanisms generating and maintaining this cell-specificity in gene expression has therefore attracted wide interest (Brady et al., 2007; Heintzman et al., 2009). In plants, leaves differentiate into a number of cell types that together allow efficient photosynthesis, and this differentiation is particularly apparent in plants that use the C_4 assimilatory pathway. In contrast to the majority of angiosperms in which CO_2 is fixed directly into 3-carbon compounds by the enzyme Ribulose-Bisphosphate Carboxylase/Oxygenase (RuBisCO) in many cell types (a process termed C_3 photosynthesis), C_4 species have acquired a carbon concentrating mechanism which almost always relies on compartmentation of gene expression between mesophyll (M) and bundle sheath (BS) cells. By repositioning the photosynthetic process between these cells, C_4 species generate a metabolic pathway that concentrates CO_2 around RuBisCO and therefore increases the efficiency of photosynthesis (Hatch & Slack, 1970).

In C_4 leaves, CO_2 is initially converted to bicarbonate by carbonic anhydrases (CAs) and assimilated to form four-carbon acids by an alternative carboxylase, phospho*enol*pyruvate carboxylase (PEPC). This fixation by PEPC occurs specifically in M cells and produces high concentrations of four-carbon acids. These acids then diffuse into BS cells where they are decarboxylated and re-fixed by RuBisCO. All of the enzymes operating in C_4 metabolism are already present within C_3 leaves, but they are present within both M and BS cells and in most cases are much less abundant than in leaves of C_4 species (Aubry et al., 2011). The evolution of a two-celled C_4 pathway therefore requires the recruitment of mechanisms restricting C_4 photosynthesis enzymes to M or BS cells, as well as mechanisms that increase the abundance of these enzymes in C_4 compared with C_3 leaves (Hibberd & Covshoff, 2010). As C_4 photosynthesis is present within at least 60 independent lineages of angiosperm (Sage et al., 2011), the repeated evolution of these gene regulatory networks represents one of the most striking examples of convergence within biology.

A number of mechanisms responsible for the patterns of gene expression in C_4 leaves have been characterized. For example, the increased abundance of C_4 enzymes can be conferred by elements within the promoter (Matsuoka et al., 1994; Gowik et al., 2004; Kajala et al., 2012; Wiludda et al., 2012), intron (Nomura et al., 2005b) or untranslated regions (UTRs) (Ali & Taylor, 2001; Wiludda et al., 2012). Similarly, cell-specificity can be generated by the promoter (Matsuoka et al., 1994; Gowik et al., 2004), exon (Brown et al., 2011) or UTRs (Patel et al., 2004; Kajala et al., 2012). As these studies focused on a number of genes isolated from phylogenetically distinct C_4 species, they suggest that the mechanisms regulating genes of the C_4 pathway vary both within and between independent C_4 lineages. In the majority of cases, different sequences generate increased abundance and cell-specificity for the same gene (Marshall et al., 1997;

Akyildiz et al., 2007; Kajala et al., 2012; Wiludda et al., 2012), suggesting that the evolution of these two traits is not coordinated. Most mechanisms identified to date are also unique to each C_4 lineage studied, although a homologous mechanism has been co-opted to generate BS-specificity for multiple genes encoding malic enzyme within *Zea mays* and *Gynandropsis gynandra* (formerly designated as *Cleome gynandra,* Brown et al., 2011). However, other C_4 genes appear to be regulated by non-homologous mechanisms (Williams et al., 2012). This seems surprising, as comparative transcriptomics of congeneric C_3 and C_4 species suggest that increased or decreased expression may have evolved for thousands of genes (Brautigam et al., 2011; Gowik et al., 2011). Furthermore, transcriptomics of M and BS cells from C_4 species *Zea mays* and *Setaria viridis* suggests that over 5,000 transcripts are differentially abundant between the two cell types (Li et al., 2010a; Chang et al., 2012; John et al., 2014). High-throughput experimental studies have therefore highlighted a large gap in current knowledge about how two-celled C_4 photosynthesis is both established and maintained. Increased abundance and cell-specificity appears to have evolved for thousands of transcripts within C_4 leaves, yet mechanisms coordinating the regulation of multiple transcripts mostly remain uncharacterized. Identifying these mechanisms is important for understanding how a system as complex as the C_4 leaf could evolve.

CA is encoded by a multi-gene family and previous studies have observed very low CA activity in BS cells of multiple C_4 species from independent lineages (Burnell & Hatch, 1988). For low levels of CA to be maintained, multiple CA isoforms must be preferentially restricted to M-cells. Despite this, only individual *CA* genes have been studied to date. For example, in *Flaveria,* loss of a transit peptide repositions a highly abundant chloroplastic CA into the cytosol (Tanz et al., 2009), while in *G. gynandra*, M-specificity of *CA4* is mediated by the UTRs (Kajala et al., 2012). To better understand the precise alterations required to recruit *CA* genes into C_4 photosynthesis, we examined *GgCA4,* which is preferentially expressed in M cells of the C_4 model species *G. gynandra* (Kajala et al., 2012). We report two molecular events associated with the evolution of increased expression conferred by *GgCA4*, as well as a post-transcriptional mechanism conferred by a short RNA motif that generates preferential accumulation in M cells compared with BS cells. We further establish that relatively limited expression in BS cells of a carbonic anhydrase enzyme, CA2*,* as well as a third C_4 enzyme, PPDK, is mediated by the same motif. This is the first characterised mechanism coordinating the accumulation of multiple C_4 pathway enzymes in M cells.

RESULTS

Multiple mechanisms underlie the evolution of increased CA4 activity in C_4 leaves.

The promoters of many C_4 genes confer strong, light-activated expression in leaves of C_4 monocotyledons and dicotyledons (Sheen, 1999). In *Zea mays* (maize) and *Flaveria*, this recruitment to photosynthetic metabolism evolved via the acquisition of *cis*-elements that are recognized by pre-existing *trans*-acting factors in C_3 leaves (Ku et al., 1999; Nomura et al., 2000; Akyildiz et al., 2007; Engelmann et al., 2008). Despite this, it is unclear whether these increases in expression compared with orthologous C_3 genes are mediated by one or multiple evolutionary events. We therefore investigated whether the *G. gynandra CA4* gene evolved *cis*-elements that are recognized by pre-existing *trans*-factors in the C_3 relative, and if this was the case, how many evolutionary events likely underlie this increased expression. To test this, promoter, exon, intron and UTR sequences from the *CA4* gene were isolated from both the C_3 model species *A. thaliana* and its closest C_4 relative *G. gynandra,* fused to the *uidA* gene encoding the GUS reporter (Figure 1) and used to transform *A. thaliana*. The localization and activity of GUS were then determined to characterize mechanisms that have evolved in *cis* to increase CA4 abundance in C_4 leaves. Promoter regions were defined as the entire intergenic region between the transcriptional start site of *CA4* and the upstream locus (AT1G70420 in *A. thaliana*). In *G. gynandra*, genome walking isolated a 695-bp open reading frame with 74% identity to AT1G70420 that is present 1023 bp upstream of the translation start site of Gg-*CA4*. This implies that this region is syntenic between *A. thaliana* and *G. gynandra*. Histological staining for GUS activity revealed that the promoter regions of At-*CA4* and Gg-*CA4* are both sufficient to generate increased activity compared with the control Cauliflower Mosaic Virus 35S (CaMV35S) promoter (Figure 1A-C). This increase in activity was most evident in young leaves. Quantitative fluorometric assays (Jefferson et al., 1987) established that GUS activities directed by the promoters of At-*CA4* and Gg-*CA4* were 200- and 380-fold higher than the CaMV35S control, respectively (Figure 1O). The high-level expression of *CA4* conferred by the promoter from *A. thaliana* indicates that the ancestral C_3 state is for strong expression. However, the promoter from *G. gynandra* has evolved to increase *CA4* expression almost 2-fold further ($p = 0.007$). The intergenic regions upstream of At-*CA4* and Gg-*CA4* are 2019 and 1023 bp respectively, so it is possible that *cis*-elements reducing expression in the ancestral state were lost in the evolution of C_4 photosynthesis.

Neither the exons nor introns of At-*CA4* had a statistically significant impact on GUS accumulation in *A. thaliana* compared to the CaMV35S control (Figure 1D,E,O). However, the expression conferred by exons and introns combined compared to exons alone was marginally higher ($p = 0.06$, student's two-tailed t-test). Alternatively, introns of *GgCA4* led to increased activity of the reporter compared to the CaMV35S control ($p = 0.02$, student's two-tailed t-test) (Figure 1F,G,O). The combined intron length of Gg-*CA4* is much shorter

than At-*CA4* (Kajala et al., 2012), so the enhanced expression conferred by Gg-*CA4* introns may have evolved via the loss of *cis*-elements that act to reduce expression in the ancestral C_3 state. Quantification of GUS activity directed by At-*CA4* UTRs also identified two additional sequences reducing expression. First, GUS activity directed by the At-*CA4* 5' UTR alone was higher than when both UTRs were present ($p = 0.003$, student's two-tailed t-test) (Figure 1H,I,O), suggesting that the 3' UTR of At-*CA4* negatively regulates expression. Second, the 5' UTR of At-*CA4* contains a 625-bp intron that also acts to repress expression compared to the spliced 5' UTR ($p = 0.0002$, student's two-tailed t-test) (Figure I,J,O). Interestingly, this intron is absent from the 5'UTR of Gg-*CA4*. However, as both UTRs of Gg-*CA4* and At-*CA4* direct similar levels of GUS activity ($p = 0.68$), loss of these elements within UTRs appears not to have led directly to the overall increased expression of *CA4* in the C_4 lineage. We therefore conclude that increased abundance of CA4 in leaves of *G. gynandra* likely evolved via two events; loss of *cis*-elements within the promoter, as well as loss of elements within introns that act to reduce *CA4* expression in the ancestral C_3 leaf.

CA4 UTRs contain a cis-element that also regulates other C_4 transcripts

We previously demonstrated that either the 5' or 3' UTR from *CA4* from both *G. gynandra* and *A. thaliana* is sufficient to generate M-specificity within leaves of *G. gynandra* (Kajala et al., 2012). To test whether this is also the case for the other highly expressed *CA* gene in leaves of *G. gynandra* (Brautigam et al., 2011), we isolated UTRs from Gg-*CA2*, as well as the homologous UTRs from *A. thaliana*. We fused these UTRs to the *uidA* reporter, with the CaMV35S promoter driving expression of the fusion, and used microprojectile bombardment (Supplemental Figure 1 online) to determine if the UTRs were sufficient to generate M-specificity. In our microprojectile bombardment assay, discrete foci of GUS activity represent independent transformation events. For the BS, GUS staining was always restricted to individual cells, allowing us to count GUS-positive BS cells independently. For the M, GUS staining sometimes spread from the highly-expressing transformed cell to adjacent M cells (see Supplemental Figure 1B). In these instances, only one M-cell was counted, to eliminate false positives from spreading of the GUS stain. The CaMV35S promoter alone directed GUS accumulation in equal numbers of M and BS cells (595 and 613 cells respectively, totalled across all replicates; see Supplemental Table 1 online). This high efficiency and transformation rate effectively control against variation in transgene expression caused by the insertion location, as each individual transformation event has a very low statistical effect. Fusion of the 5' UTRs of Gg-*CA2* and At-*CA2* did not affect the ratio of M and BS cells accumulating GUS (Supplemental Figure 2 online), but the 3' UTRs of both homologs were sufficient to generate strong preferential accumulation in M-cells (Supplemental Figure 2 online). This result supported the hypothesis that sequences within *CA* UTRs coordinate M-

specificity for multiple functionally related transcripts. These data combined with previous work (Kajala et al., 2012) indicated that four 5' UTRs and six 3' UTRs from *CA2, CA4* and *PPDK* derived from both C_4 *G. gynandra* and C_3 *A. thaliana* are capable of preferentially directing accumulation in M cells.

To gain insight into the mechanisms responsible for generating this M-specificity, two independent approaches were taken. First, a deletion analysis on the 5' UTR of Gg-*CA4* was used to identify sequences necessary for M-specificity (Figure 2A). Second, computational motif detection was undertaken to identify *cis*-elements that may be present in all of the *A. thaliana* and *G. gynandra* UTRs demonstrated to generate M-specificity. Deletion analysis was performed by generating 7 to 20 bp incremental deletions from either the 5' or 3' end of the UTR and testing the effect of each deletion on M-specificity using microprojectile bombardment. We identified a thirteen nucleotide region from -99 to -86 bp, as well as a 7 nucleotide sequence from the 3' end that were individually necessary for M-specificity (Figure 2B).

Computational motif prediction using ten UTR sequences from both *A. thaliana* and *G. gynandra* identified a nine-nucleotide motif present in all ten UTRs (Figure 2C). Within this motif, three nucleotides were conserved in all ten UTRs, while a further two were conserved between at least seven sequences (Figure 2C-D). This nine-nucleotide motif overlaps with the 5' deletions that abolished the M-preferential expression of GUS, suggesting that this motif and nucleotides directly upstream may be critical for directing accumulation of Gg-*CA4* in M cells (Figure 2E).

To verify that this nine-nucleotide *cis*-element is required for preferential GUS accumulation in M cells, site-directed mutagenesis was used to alter the last five nucleotides of the motif, which contained the three completely conserved nucleotides (Figure 3A). Stable transformants of *G. gynandra* confirmed that whereas a 99-nucleotide truncation of Gg-*CA4* 5' UTR was sufficient to generate M-specificity (Figure 3B), the mutated UTR sequence directed GUS expression in both M and BS cells (Figure 3C). This was observed across three independent transgenic lines for each construct (Supplemental Figure 3 online). Substitution of the same five nucleotides within 5' and 3' UTRs of other transcripts was also sufficient to abolish M-specificity in all cases (Figure 3D). This is the second *cis*-element directing M-specificity in C_4 leaves to be identified after MEM1 (Mesophyll Enhancing Module 1), which is present in the *ppcA* promoter of C_4 *Flaveria* species (Gowik et al., 2004; Akyildiz et al., 2007). We therefore named this nine-nucleotide motif present in Gg-*CA4* as MEM2.

MEM2 does not impact on abundance of transcripts in M or BS cells

As MEM2 is a common element in multiple UTRs that direct M-preferential expression, we sought to define the minimal sequence sufficient for preferential accumulation of GUS in M cells. We observed that

deletions from the 5' or the 3' ends of the minimal Gg-*CA4* 5' UTR both disrupted M-specificity. We therefore hypothesized that the 5' UTR region directly upstream of the start codon is also important for MEM2 function. In eukaryotic genes, this region is important for recognition of the translation start codon AUG by the ribosome and translation initiation complex (Jackson et al., 2010). To test the importance of this sequence in the generation of M-specificity by the Gg-*CA4* 5' UTR, we generated two chimeric constructs, fusing a short 35-bp sequence around MEM2 to 8-bp or 13-bp sequences from immediately upstream of the translational start codon. To reflect the distance between these sequences in the endogenous UTR, randomly generated linker sequences of either 57 or 52 bp in length were inserted between the two (Figure 4A). Microprojectile bombardment suggested that both constructs were sufficient to direct M-preferential expression, but the effect was stronger when the larger (13 bp) sequence upstream of the translational start site was included (Figure 4B). Further evidence supporting these results was obtained by generating two independent *G. gynandra* stable lines expressing this chimaeric construct. To assess the extent to which this construct generated M-specificity, the percentage of M and BS cells expressing GUS was calculated by examining 22 transverse sections of GUS-stained leaves from these stable lines. Although there was some variability between sections, on average MEM2 combined with 13 nucleotides upstream of the start codon led to a reduction in GUS accumulation in BS cells, similar to the unmodified Gg-*CA4* 5' UTR alone (Figure 4C). A representative example of a transverse section is shown in Figure 4D and further examples of both transgenic lines are shown in Supplemental Figure 3 online. We therefore conclude that MEM2 is sufficient to direct strong preferential expression in M cells when combined with a second element, the sequence immediately 5' of the start codon from the same UTR.

Given the importance of nucleotides 5' of the start codon in the initiation of translation (Jackson et al., 2010), we sought to test the hypothesis that MEM2 acts post-transcriptionally. To investigate this, we examined the abundance of Gg-*CA4* transcripts in M and BS cells from cell-specific transcriptome datasets (Aubry et al., 2014a) (Supplemental Table 2 online). Gg-*CA4* transcripts were highly and similarly abundant in both M and BS cells (log2 fold change M:BS = -0.05, p-value = 0.73, Fisher's exact test). We next used the stable transformants containing *uidA* fused to the endogenous and mutated versions of the Gg-*CA4* 5' UTR to investigate whether MEM2 affected the abundance of *uidA* transcripts and GUS protein in each cell type. M and BS preparations were generated from these lines and both RNA and protein isolated. RT-qPCR on two transcripts known to be enriched in either cell type (*PHOSPHOENOLPYRUVATE CARBOXYLASE* (*PPC*)*,* enriched in M cells and *NAD-DEPENDENT MALIC ENZYME 2 (NADME2),* enriched in BS cells) indicated that there was clear enrichment of the desired cell type in both M and BS fractions from leaves expressing the Gg-*CA4* 5' UTR fused to GUS (Supplemental Figure 4 online). In our BS preparations from

leaves expressing the mutated Gg-*CA4* 5’ UTR, the *PPC* marker for M cells was clearly depleted, suggesting that the BS preparation had minimal contamination from M cells. However, we did detect the BS marker *NADME* in the M preparations from these leaves, suggesting that enrichment of M cells was less successful in these lines (Supplemental Figure 4 online). We note that this does not affect the conclusions drawn about transcript or protein abundance in M or BS cells from lines expressing the unmutated Gg-*CA4* 5’ UTR.

We found that *uidA* transcripts were not enriched in the M preparation, and tended in fact to be more abundant the BS when driven by the Gg-*CA4* 5' UTR (Figure 4E). This implies that the Gg-*CA4* 5' UTR does not generate M-specificity by reducing the abundance of transcripts in BS cells. In lines expressing the MEM2-mutated UTR, *uidA* expression in M preparations was more variable between lines, but transcripts were still abundant in both M and BS preparations. This variability may be due to differential integrity of RNA between samples. Together these data suggest that MEM2 does not generate M-preferential expression by altering the abundance of transcripts in a cell-type-specific manner. As *GUS* transcripts were not enriched in our isolated M preparation, it is likely that MEM2 functions post-transcriptionally to enrich the quantity of GUS protein translated within M cells compared to BS cells. To confirm this, the activity of GUS in the same cell-type fractions used for RT-qPCR was determined using the quantitative MUG assay. GUS activity was highly enriched in M preparations compared to BS preparations when driven by the endogenous Gg-*CA4* 5' UTR (Figure 4F). Conversely, when driven by the mutated UTR sequence, the activity of GUS was much reduced in M cells, and marginally higher in BS preparations. These quantitative assays are consistent with our qualitative observation of GUS accumulation in transverse sections (Figure 3B-C) and suggest that when MEM2 is mutated, transcripts in both M and BS cells are translated to a similar extent, so that GUS is present in both cell types. However, when the MEM2 sequence is intact, the abundant transcripts in BS cells are translated to a much lesser extent than in M cells, leading to M-specific accumulation of GUS. We observed that mutation of MEM2 does not affect the rate of translation *in vitro* (Supplemental Figure 5 online), or the overall activity of GUS in *Arabidopsis* leaves (Supplemental Table 3 online). We therefore propose that regulating translation is not a general function of the MEM2 motif in all cells, but specific to its role in generating M-specificity in the C_4 *G. gynandra* leaf.

DISCUSSION

Evolutionary events underlying the recruitment of CA4 into C_4 photosynthesis

The evolution of C_4 photosynthesis from the C_3 ancestral state requires mechanisms generating increased abundance as well as cell-specificity of a large number of enzymes (Marshall et al., 1997; Patel et al., 2004; Gowik et al., 2004; Wiludda et al., 2012). A recent study has suggested that the evolution of increased abundance can be complex, with multiple sequences within promoters and 5' UTRs contributing to the expression of C_4 genes (Wiludda et al., 2012). We examined mechanisms underlying the evolution of CA4, an enzyme catalyzing the first step of the C_4 pathway. Our data support the finding that the evolution of increased expression and cell-specificity of C_4 genes is complex, and in the case of *CA4* sequences within promoters, 5' and 3' UTRs and introns all contribute to the evolution of increased expression of Gg-*CA4* relative to At-*CA4*. Similar to the *PHOSPHOENOLPYRUVATE CARBOXYLASE A* and *GLYCINE DECARBOXYLASE P-SUBUNIT A (GLDPA)* genes of *Flaveria bidentis* (Gowik et al., 2004; Wiludda et al., 2012), we demonstrate that increased expression of Gg-*CA4* evolved via changes to *cis*-elements upstream of the transcriptional start site. However, we extend this analysis to show that sequences within introns and interactions between 5' and 3' UTRs can also generate increased CA4 abundance. In *Zea mays*, representing an independent C_4 lineage, introns direct increased expression of genes encoding NADP-dependent malic enzyme and Aspartate aminotransferase (Nomura et al., 2005a, 2005b). Together, our data and these previous studies suggest that multiple alterations in *cis* to non-coding regions are a primary route through which the increased abundance of C_4 pathway enzymes evolved in multiple C_4 lineages. We also found that the UTRs of *Gg-CA4* confer an increase in translational efficiency *in vitro*, as well as an increase in GUS accumulation *in vivo*, but that these effects are independent of the MEM2 element that restricts expression to M cells. UTRs of *NADP-ME1* from *Flaveria bidentis* (Ali & Taylor, 2001) and *RbcS* from *Amaranthus hypochondriacus* (Patel et al., 2004) are also sufficient to generate increased translational efficiency and so these combined data indicate that this appears to be an important mechanism that has been recruited repeatedly to regulate expression of C_4 genes. As we found that UTRs of the C_3 homologue At-*CA4* were sufficient to increase accumulation of GUS, this likely represents an ancestral character present in C_3 plants.

Together, these findings allow us to hypothesize evolutionary events underlying the recruitment of *CA4* to C_4 photosynthesis (Supplemental Figure 6 online). In *A. thaliana*, which is assumed to represent the ancestral C_3 state, the promoter of *CA4* directs strong expression; however, in C_4 *G. gynandra*, the promoter and introns increase *CA4* expression further (Supplemental Figure 6 online). As both the promoter and combined introns of Gg-*CA4* are half as long as the homologous sequences from At-*CA4*, we suggest that

the loss of *cis*-elements that reduce expression is a likely evolutionary scenario for this mechanism. Our data suggest that in leaves of C_4 *G. gynandra* M-preferential accumulation of CA4 is generated by post-transcriptional regulation mediated by the both the 5' and 3' UTR of *CA4* transcripts. This is consistent with the fact that *CA4* transcripts are equally abundant between M and BS cells of *G. gynandra* leaves (Supplemental Table 2), and that the Gg-*CA4* 5' UTR directs equal transcript abundance in M and BS cells (Figure 4E). Combining MEM2 with the nucleotides preceding the start codon from Gg-*CA4* is sufficient to confer accumulation of GUS in M cells. In eukaryotes, these nucleotides play an important role in recognizing the proper translation start site by the translation initiation complex (Jackson et al., 2010). Interestingly, both 5' and 3' UTRs are bound in this complex at the initiation of translation (Tarun & Sachs, 1995; Wilkie et al., 2003; Jackson et al., 2010). One possibility is that M-specificity evolved via the recruitment of *trans*-acting factors recognizing MEM2 within 5' and 3' UTRs at this complex within M or BS cells (Supplemental Figure 6 online). However, although we have defined the 5' UTR sequences sufficient to direct M-preferential accumulation (Figure 4A-D), we have not yet defined if adding the MEM2 element to a heterologous 3' UTR is sufficient to confer M-specific translation in the same way. This future experiment will be an important step in validating the roles and interactions between 5' and 3' UTRs containing MEM2 in generating M-specificity. In addition, it is also currently unclear whether MEM2 functions to promote translation in M cells, or repress translation in BS cells. Understanding this will be key in providing insight into the *trans*-acting factors that recognize MEM2.

Our data suggest that the evolution of increased abundance and M-preferential expression of Gg-*CA4* evolved independently of each other, and via distinct molecular mechanisms (Supplemental Figure 6 online). Mutation of MEM2 is sufficient to abolish M-preferential expression, but has no effect on the levels of GUS activity conferred in *cis* in a C_3 background (Supplemental Table 3 online). Conversely, increased abundance is conferred in *cis* by the Gg-*CA4* promoter region (Figure 1C). This promoter sequence is unlikely to be important in generating M-specificity, as Gg-*CA4* transcripts are equally abundant between M and BS cells (Supplemental Table 2 online). Computational modeling of the convergent evolutionary events generating C_4 photosynthesis in a wide variety of lineages predicts that increased abundance and cell specificity also evolved independently for at least five additional C_4 pathway enzymes (Williams et al., 2013). It therefore appears that the decoupling of changes to enzyme abundance and cell type localization may be a general trend in C_4 evolution, occurring in multiple lineages and in the evolution of both enzymes localized to either BS and M cells.

MEM2 - a motif coordinating cell-type-specific accumulation of multiple enzymes

Deletion analysis and computational detection identified MEM2, a nine-nucleotide element present within the 5' and 3' UTRs of multiple *A. thaliana* and *G. gynandra* transcripts. Several studies from yeast, animals and plants have identified functionally related groups of mRNAs whose regulation is coordinated by post-transcriptional processes (Keene, 2007; Zhu et al., 2007; Staiger & Köster, 2010). Furthermore, some regulons have also been shown to share small sequence-specific, or structural elements that confer similar rates of processing or decay (Chan et al., 2005; Goodarzi et al., 2012). As hundreds of transcripts can be coordinated by small elements within transcripts (Goodarzi et al., 2012), *CA2*, *CA4* and *PPDK* could belong to a much larger suite of transcripts whose regulation is coordinated by MEM2 in C_4 leaves.

MEM2 also represents a rare example of an element that performs the same function in either the 5' or the 3' UTR of transcripts. Other elements present in either 5' or 3' UTR are known, such as the iron responsive elements (IREs) present in animals and single-celled eukaryotes (Leipuviene & Theil, 2007). In contrast to MEM2, however, IREs perform different functions when present in the 5' or 3' UTRs (Muckenthaler et al., 2008). The adaptive significance of possessing multiple copies of MEM2 within transcripts is not yet clear, but both UTRs of *PPDK* and *CA4* are known to confer stronger M-preferential expression when combined compared with the 5' or 3' alone (Kajala et al., 2012). It is therefore possible that UTRs containing MEM2 function additively.

Distinct C_4 pathway enzymes share a common regulator

Since the discovery of the C_4 pathway (Hatch & Slack, 1966), only one mechanism that regulates the expression of multiple genes in specific cells has been reported (Brown et al., 2011), and in that case, the specific *cis*-element was not identified. In addition, although Brown et al. (2011) reported a shared mechanism regulating the expression of multiple genes, they all encoded sub-units of malic enzymes. MEM2 represents a *cis*-element that coordinates the regulation of transcripts encoding distinct C_4 pathway enzymes. Interestingly, three M-specific and two BS-specific genes in three C_4 monocot species may also share similar cell-type-enriched histone modifications (Heimann et al., 2013). These modifications include trimethylation of H3K4 residues as well as histone acetylation, both of which correlate with active transcription. This suggests transcriptional control at least in part contributes to the cell-type-specific accumulation of multiple enzymes in monocotyledons.

The complexity of C_4 metabolism currently represents a considerable challenge for attempts to engineer C_4 photosynthesis into C_3 crop species (Hibberd et al 2008). The discovery of MEM2 is therefore a key finding as it identifies a mechanism that co-ordinates the cell-specific synthesis of multiple C_4 enzymes. Defining the exact method by which MEM2 is recognized in C_4 leaves is an important next step. It is possible

that a micro RNA (miRNA) is the *trans*-acting factor that recognizes MEM2, as many miRNAs act at the ribosome to repress translation (Li et al., 2013). However, we were unable to find any known miRNAs that matched MEM2 or surrounding sequences within UTRs of *CA2*, *CA4* or *PPDK* in Arabidopsis or *G. gynandra*. It is also possible that an RNA-binding protein may recognize MEM2 in *G. gynandra* bundle sheath cells. PUF proteins recognize *cis*-elements present in mRNA that are typically within 3' UTRs, of a similar size to MEM2 and are dependent on secondary structure (Tadauchi et al., 2001; Francischini & Quaggio, 2009; Li et al., 2010b; Filipovska et al., 2011). However, PUF proteins typically recognize a sequence beginning with 5' UGUR (whereby R represents A or G) (Lu et al., 2009), which is not part of the conserved MEM2 sequence, or directly upstream in any of the MEM2-containing UTRs. A PUF recognition domain that does not require a target sequence containing UGUR at the 5' end has been engineered (Campbell et al., 2014), suggesting that the evolution of alternative recognition sites is at least possible, if unlikely.

Another large family of RNA-binding proteins in Arabidopsis is the PENTATRICOPEPTIDE REPEAT (PPR) proteins, which appear not to be restricted to any consensus in target sequence, but are mostly active in chloroplasts and mitochondria (Barkan & Small, 2014). The MEM2 consensus is similar to that bound by RNA editing PPR proteins ORGANELLE TRANSCRIPT PROCESSING 84 (OTP84) (Hammani et al., 2009) and MITOCHONDRIAL EDITING FACTOR 14 (MEF14) (Verbitskiy et al., 2011; Barkan et al., 2012) and so it is possible that MEM2 is recognised by a related protein, which lost organellar targeting. However, as fewer than 1% of Arabidopsis PPR proteins are known to localize outside of the nucleus (Colcombet et al., 2013) this seems unlikely. The MEM2 sequence itself therefore provides few clues into the likely *trans*-acting factor(s) that recognizes MEM2 in BS cells. Identifying the trans-acting factor recognizing MEM2 is therefore an important aim for future research. MEM2 is the first *cis*-element identified that is necessary to direct cell-type-specific accumulation of multiple C_4 enzymes. Our study suggests that recruitment of this pre-existing *cis*-element parsimoniously explains how multiple C_4 enzymes evolved the same localization. Its discovery therefore represents an important breakthrough towards understanding how a complex phenotype such as C_4 photosynthesis can evolve over relatively short evolutionary timescales.

MATERIALS & METHODS

Plant growth

Sterilised *A. thaliana* ecotype Col-0 seeds were spread on plates containing 0.5 x Murashige-Skoog basal salts (MS) and 1% (w/v) agar, pH 5.8, then stratified for 48 h at 4°C in the absence of light. Following stratification, seeds were transferred to a long-day growth room at 22°C, relative humidity of 65%, ambient CO_2 concentration and a light intensity of 200 µmol m^{-2} s^{-1} photon flux density (PFD) for 7 days. Seedlings were then transferred to a 1:1 mixture of Levington's M3 potting compost:fine vermiculite. *G. gynandra* seeds were germinated in the dark at 30°C for 30 h on a bed of wet filter papers. Seeds were then transferred to a medium containing 1 x MS, 1% (w/v) sucrose and 1% (w/v) agar. Seedlings were then grown in 16 h light, 8 h dark, at 20°C, ambient CO_2 concentration and a light intensity of 200 µmol m^{-2} s^{-1} PFD.

Vector construction and stable plant transformation

cDNA sequences were obtained using 5′ and 3' Rapid Amplifcation of cDNA Ends (RACE), coding regions were amplified from both cDNA and genomic DNA, and promoter regions were isolated by genome walking as described previously (Kajala et al., 2012). UTR reporter constructs containing the *uidA* gene (encoding GUS) were generated by ligation of UTRs into a modified vector containing a *gfp:uidA:nosT* cassette (Brown et al., 2011). 3′ UTRs were inserted between *uidA* and *nosT*. 5′ UTRs were fused to the CaMV35S promoter by Polymerase Chain Reactions and inserted in front of the cassette. Vectors containing promoter and coding regions were assembled into the same cassette using Gibson assembly (Gibson et al., 2009). Two constructs were synthesized using the coding regions of Gg-*CA4* and At-*CA4*. Firstly, the genomic coding sequence containing all exons and introns intact were fused to the open reading frame (ORF) of *uidA*, to test regulation conferred by exons and introns combined. Secondly, the spliced ORF of Gg-*CA4* and At-*CA4* was amplified from cDNA and fused to *uidA* to test regulation conferred in the absence of introns.

Assembled constructs were used for microprojectile bombardment and also placed in binary vectors to generate stable *A. thaliana* or *G. gynandra* transformants. Site-directed mutagenesis was performed using the Quickchange™ method. Chimaeric constructs were generated by site-directed insertion or fusion of sequences by overlapping PCR. UTRs and mutated UTRs fused to *uidA* were ligated into the pTNT vector (Promega) for *in vitro* transcription. Transgenic *A. thaliana* Col-0 plants were generated by floral dipping (Clough & Bent, 1998) using *Agrobacterium tumefaciens* strain GV3101. Primary transformants were identified by selection on 50 µg ml^{-1} kanamycin for 7 days. Stable transformation of *G. gynandra* was performed as previously described (Newell et al., 2010), with modifications to the co-cultivation medium (MS,

3% (w/v) sucrose, 1 mg/l thidiazuron (TDZ), and 0.2 mg/l indole-3-acetic acid (IAA), 100 µM acetosyringone), regeneration medium (MS 3% (w/v) sucrose, 1 mg/l TDZ and 0.2 mg/l IAA) and shoot regeneration medium (MS, 3% (w/v) sucrose 1 mg/l BAP and 0.2 mg/l IAA).

Microprojectile bombardment of *G. gynandra*

Transient expression of the constructs in *G. gynandra* was achieved by microprojectile bombardment using a Bio-Rad PDS-1000/He particle delivery system as described by Brown et al., (2011). *G. gynandra* seeds were incubated at 30°C for 30 h on moistened filter paper to achieve uniform germination. *G. gynandra* seeds were transferred to 0.5 x MS medium supplemented with 1% (w/v) sucrose and 1% (w/v) agar in sterile conditions. Seedlings were then grown in a growth room with 16-h days, at 22°C, relative humidity of 65%, ambient CO_2 concentration and a light intensity of 200 µmol m^{-2} s^{-1} PFD. After 14 days of growth, seedlings were prepared for microprojectile bombardment by removing all root tissue below the medium surface, leaving only the stems, cotyledons and two-expanding leaves, with the primary leaflet measuring approximately 110 mm from base to tip and 75-80 mm across at the widest point (Supplemental Figure 1A). Leaf size was found to be an important factor determining the ratio of mesophyll and bundle transformed using the ubiquitously expressed CaMV35S control.

To transform individual cells within *G. gynandra* leaves, 350 ng M-17 tungsten particles (1.1 µm diameter, BioRad) were washed with 100% ethanol and resuspended in ultrapure water. Then, 1.5 µg of plasmid DNA was adhered to the tungsten particles as described by Patel et al. (2006), adding DNA to the tungsten particles while vortexing at a slow speed by a Vortex-Genie 2 (Scientific Industries). After addition of the DNA, 50 µl 2.5 M calcium chloride (Fisher Scientific) and 10 µl 100 mM spermidine (Sigma-Aldrich) were added to the particle suspension to facilitate binding of the DNA to the tungsten particles. The tungsten-DNA suspension was incubated for 10 minutes on ice, with frequent agitation to prevent pelleting of the tungsten particles. The particles were then pelleted by centrifugation at 6,000 rpm for 2 s and washed once and resuspended in 100 µl 100% ethanol. Immediately prior to the bombardment of *G. gynandra*, 10 µl aliquots of tungsten/DNA were transferred to plastic macrocarriers (BioRad) and the ethanol allowed to evaporate. Three macrocarriers were used for each transformation. After microprojectile bombardment (Bio-Rad PDS-1000/He particle delivery system), seedlings were placed upright in a sealed petri dish, with the base of their stems immersed in 0.5 x MS medium. The abaxial leaf surface was placed on Whatman grade 1 filter paper moistened with 0.5 x MS, to prevent leaves from drying. Seedlings were left for 40 h after bombardment prior to staining for GUS activity.

β-glucuronidase (GUS) assays and *in vitro* translation assay

Staining for GUS activity was performed on *A. thaliana* or *G. gynandra* plants as described (Jefferson et al., 1987). Tissue was fixed in 3:1 ethanol/acetic acid for 30 minutes at room temperature, and chlorophyll was cleared using 70% (v/v) ethanol at 37°C for 24 h and then 5% (w/v) NaOH at 37°C for 2 h. *G. gynandra* seedlings subjected to microprojectile bombardment were incubated in X-glcA solution for 24 h. *A. thaliana* plants transformed with constructs containing the CaMV35S promoter were incubated in X-glcA solution for 18 h. At least six independent T_1 stable transgenic lines were stained for each construct. Plants derived from independent T_1 lines transformed with constructs containing the At-*CA4* or Gg-*CA4* promoters were incubated for 6 h at 37°C. Tissue was then fixed in 3:1 ethanol/acetic acid for 30 minutes at room temperature, and chlorophyll was cleared using 70% (v/v) ethanol at 37°C for 24 h and then 5% (w/v) NaOH at 37°C for 2 h. Transverse leaf sections were obtained from at least 3 independent T_1 transgenic *G. gynandra* lines by embedding fresh leaf tissue in 5% (w/v) agarose and isolating 60 µm sections using a vibratome. Sections were stained for GUS activity for 1 and 2 h.

The activity of GUS was also quantified by measuring the rate of 4-Methylumbulliferone-glucuronide (MUG) conversion to 4-Methylumbulliferone (MU) as described (Jefferson et al., 1987). Soluble protein was extracted from transgenic *A. thaliana* plants by freezing in liquid nitrogen and maceration, followed by addition of an extraction buffer (50 mM NaH_2PO_4, 0.007% (v/v) β-mercaptoethanol and 0.01% (v/v) Triton X-100). Extracts underwent centrifugation for 5 minutes at 4°C. Diluted protein extracts were incubated with 1 mM MUG at 37°C for 5, 10, 20, and 30 minutes in a 96-well plate. GUS activity was terminated at the end of each time point by the addition of 200 mM Na_2CO_3 and MU fluorescence was measured by exciting at 365 nm and measuring emission at 455 nm, averaging five light pulses for each well. The concentration of MU/unit fluorescence in each sample was interpolated using a concentration gradient of MU from 1.5 to 800 µM MU. MU/unit fluorescence was found to be linear over this range. Pilot experiments were performed with two to three independent T_1 transgenic lines expressing each transgenic construct to determine the optimal protein dilution required to retain linearity. Subsequent assays were performed using between 10-20 independent T_1 lines.

In vitro transcription was initiated from the T7 promoter using a MEGAShortScript™ T7 Kit (Ambion), with the addition of 1 U RNase OUT™ (Invitrogen). Reactions were incubated for 2 h at 37°C and transcripts were purified using MEGAClear™ and the integrity of transcribed products was verified using a Bioanalyzer™ 2100 (Agilent Technologies). Quantitative *in vitro* translation was performed using a Fluorotect™ GreenLys *in vitro* translation labeling system (Promega) in conjunction with a wheat germ extract (Promega). Time-course and concentration gradient experiments were performed to ensure

fluorescent protein production was linear and unsaturated relative to incubation time and RNA concentration, respectively. After these optimisation steps, reactions were incubated for 45 minutes at 25°C, with 200 nM RNA templates, using equimolar concentrations of each transcript.

RNA-seq analysis and qPCR analysis of GUS accumulation in *G. gynandra* transgenic lines

RNA-seq data from *G. gynandra* bundle sheath and mesophyll fractions extracted by laser capture microdissection (SRA066236, Aubry et al., 2014b) and from ribosomal pulldown (using a procedure analogous to the one described in Aubry et al. 2014a for *A. thaliana*) were quality trimmed and the adapters removed using Trimmomatic v0.30 (Bolger et al., 2014). Reads shorter than 60 bp after trimming were discarded. The remaining reads were aligned against the *CA4* gene using TopHat v2 (Kim et al., 2013). Gene and transcript expression was quantified using RSEM v1.2.7 (Li and Dewey, 2011).

Isolation of mesophyll (M) extracts from *G. gynandra* was performed as described previously (Covshoff et al., 2012) and bundle sheath (BS) cells according to Markelz et al. (2003). For each replicate, between six to eight leaves representing 350-400 mg of tissue were initially rolled to extract M sap, and then blended to separate BS strands. To quantify GUS activity, sap containing M cells was transferred directly into extraction buffer consisting of 100 mM NaH_2PO_4 (Sigma-Aldrich), 0.007% (v/v) β-mercaptoethanol (Fisher Scientific) and 0.01% (v/v) Triton X-100 (Sigma-Aldrich). Ground bundle sheath powder was re-suspended in extraction buffer and samples were spun down at 13000 x g for 5 minutes and the pellet discarded. Protein concentration was determined using the Qubit(R) Protein Assay Kit (ThermoFisher Scientific) and the MUG assay was performed immediately as described in the previous section. To quantify transcript abundance in M and BS preparations, samples were extracted using the mirVana miRNA Isolation Kit (Ambion), stopping the protocol after the extraction of total RNA (prior to the optional purification of small RNAs). Total RNA was eluted in RNAse-free water and samples treated with DNAse-I (Promega). Although leaf rolling gives RNA of good quality (Covshoff et al 2012; John et al 2014), due to the small quantities of RNA obtained, we note that variability in the integrity of some transcripts may exist between samples. qPCR was performed as described by Aubry et al. (2014) with the exception that data are presented as an average of four replicates. Primer sequences were as follows: GUS *(Forward: 5'-ACCTCGCATTACCCTTACGCTGAA-3', Reverse:* 5'-*GCCGACAGCAGCAGTTTCATCAAT-3'),* Actin7 *(Forward: 5'-TCCGACCCGATGTGATGTTATGGT-3', Reverse: 5'-CAATCACTTTCCGGCTGCAACCAA-3')*.

***cis*-element prediction**

cis-element detection was performed using the Multiple Em for Motif Elucidation (MEME) suite v.4.8.1. (Bailey et al., 2009). Unaligned transcribed UTR sequences were used as the input with the following conditions imposed: 1. No limit to the minimum or maximum motif length, 2. A minimum of one motif present in each UTR in the dataset and 3. Motif detection was not applied to the complementary strand. The extent of conservation of nucleotides within detected motifs was plotted using the MEME suite.

^

Accession Numbers

Gg-*CA4:* NCBI Genbank Accession KU517022

Gg-*CA2:* NCBI Genbank Accession KU517023

Gg-*PPDK:* NCBI Genbank Accession KU517024

At-*CA4:* AT1G70410

At-*CA2:* AT5G14740

At-*PPDK:* AT4G15530

Supplemental Data

Supplemental Figure 1: Transformation of *G. gynandra* M and BS cells

Supplemental Figure 2: 3' UTRs of Gg-*CA2* and At-*CA2* generate M-specificity

Supplemental Figure 3: Representative transverse sections from stable transgenic lines

Supplemental Figure 4: Quantification of marker transcripts in isolated M and BS cells

Supplemental Figure 5: Mutation of MEM2 does not affect translation *in vitro*

Supplemental Figure 6: Hypothesis for mechanisms regulating the abundance and cell-specificity of CA4

Supplemental Table 1: Total number of transformed M and BS cells analyzed for each construct

Supplemental Table 2: Gg-*CA4* transcript abundance measured by transcriptome sequencing of M and BS cells:

Supplemental Table 3: GUS activity conferred by mutated Gg-*CA4* UTRs Arabidopsis seedlings.

ACKNOWLEDGEMENTS

We would like to thank Steve Kelly (Plant Sciences, Oxford) for advice. We would like to thank the BBSRC for funding.

AUTHOR CONTRIBUTIONS

BPW, SJB and JMH conceived the study and designed experiments. BPW, SJB, IRL, JK, SA and SS performed the experiments and BPW & SJB analysed the data and produced figures. JMH supervised the whole study and wrote the paper together with BPW and SJB.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

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.

Ali, S. and Taylor, W.C. (2001). Quantitative regulation of the *Flaveria Me1* gene is controlled by the 3'-untranslated region and sequences near the amino terminus. Plant Mol. Biol. **46:** 251-261.

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., Li, W.W. and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. **37:** W202-W208.

Barkan, A., Rojas, M., Fuji, S., Yap, A., Chong, Y.S., Bond, C.S. and Small, I. (2012) A combinatorial amino acid code for rna recognition by pentatricopeptide repeat proteins. PLoS Genet. **8:** e1002910.

Barkan, A., Small, I. (2014) Pentatricopeptide repeat proteins in plants. Annu. Rev. Plant Biol. **65**: 415-442

Brady, S.M., Orlando, D.A., Lee, J-Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U. and Benfey, P.N. (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. Science **318:** 801-806.

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 Physiol. **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.

Burnell, J.N. and Hatch, M.D. (1988). Low bundle sheath carbonic anhydrase is apparently essential for effective C_4 pathway operation. Plant Physiol. **86:** 1252-1256.

Campbell, Z.T., Valley, C.T., Wickens, M. (2014). A protein-RNA specificity code enables targeted activation of an endogenous human transcript. Nat. Struct. Mol. Biol. **21:** 732-738.

Chan, C.S., Elemento, O. and Tavazoie, S. (2005). Revealing post-transcriptional regulatory elements through network-level conservation. PLoS Comput. Biol. **1:** e69.

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., et al (2012). Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. Plant Physiol **160**: 165-177.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J. **16:** 735-743.

Colcombet, J., Lopez-Obando, M., Heurtevin, L., Bernard, C., Martin, K., Berthome, R., Lurin, C. (2013) Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles. RNA Biology **10**: 1557-1575.

Covshoff, S. Furbank, R.T., Leegood, R.C. and Hibberd, J.M. (2012) Leaf rolling allows quantification of mRNA abundance in mesophyll cells of sorghum. *Journal of Experimental Botany* 10.1093/jxb/ers286.

Curatola, A.M., Nadal, M.S. and Schneider, R.J. (1995). Rapid degradation of AU-rich element (ARE) mRNAs is activated by ribosome transit and blocked by secondary structure at any position 5' to the ARE. Mol. Cell. Biol. **15:** 6331-6340.

Engelmann, S., Wiludda, C., Burscheidt, J., Gowik, U., Schlue, U., Koczor, M., Streubel, M., Cossu, R., Bauwe, H. and 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 Physiol. **146:** 1773-1785.

Filipovska, A., Razif, M.F.M., Nygård, K.K.A. and Rackham, O. (2011) A universal code for RNA recognition by PUF proteins. Nat. Chem. Biol. **7:** 425-427.

Francischini, C.W. and Quaggio, R.B. (2009). Molecular characterization of *Arabidopsis thaliana* PUF proteins--binding specificity and target candidates. FEBS J. **276:** 5456-5470.

Gibson, D.G., Young, L., Chuang, R-Y., Venter, J.C., Hutchison, C.A. and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods **6:** 343-345.

Goodarzi, H., Najafabadi, H.S., Oikonomou, P., Greco, T.M., Fish, L., Salavati, R., Cristea, I.M. and Tavazoie, S. (2012). Systematic discovery of structural elements governing stability of mammalian messenger RNAs. Nature **485:** 264-268.

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 phosphoenolpyruvate carboxylase gene. Plant Cell **16:** 1077-1090

Hammani, K., Okuda, K., Tanz, S.K., Chateigner-Boutin, A-L., Shikanai, T. and Small, I. (2009) A study of new *Arabidopsis* chloroplast RNA editing mutants reveals general features of editing factors and their target sites. Plant Cell **21:** 3686-3699.

Hatch, M.D. and Slack, C.R. (1966). Photosynthesis by sugar-cane leaves. Biochem J. **101:** 103-111.

Hatch, M.D. and Slack, C.R. (1970). Photosynthetic CO_2-fixation pathways. Annu. Rev. Plant Physiol. **21:** 141-162.

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

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., Ching, K.A., Antosiewicz-Bourget, J.E., Liu, H., Zhang, X., Green, R.D., Lobanenkov, V.V., Stewart, R., Thomson, J.A., Crawford, G.E., Kellis, M., et al., (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature **459:** 108-112.

Hibberd, J.M. and Covshoff, S. (2010). The regulation of gene expression required for C_4 Photosynthesis. Annu. Rev. Plant Biol. **61:** 181-207.

Jackson, R.J., Christopher, U., Hellen, T. and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev. Mol. Cell Biol. **11**: 113-127.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. **6:** 3901-3907.

John C.R., Smith-Unna R.D., Woodfield H., Covshoff S., Hibberd J.M. (2014) Evolutionary convergence of cell-specific gene expression in independent lineages of C_4 grasses. Plant Physiol **165**: 62-75.

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 J. **69:** 47-56.

Keene, J.D. (2007). RNA regulons: coordination of post-transcriptional events. Nat. Rev. Genet. **8:** 533-543.

Ku, M.S.B., Agarie, S., Nomura, M., Fukayama, H., Tsuchida, H., Ono, K., Hirose, S., Toki, S., Miyao, M. and Matsuoka, M. (1999). High-level expression of maize phosphoenolpyruvate carboxylase in transgenic rice plants. Nat. Biotechnol. **17:** 76-80.

Leipuviene, R. and Theil, E.C. (2007). The family of iron responsive RNA structures regulated by changes in cellular iron and oxygen. Cell. Mol. Life Sci. **64:** 2945-2955.

Li, S., Liu, L., Zhuang, X., Yu, Y., Liu X., Cui, X., Ji, L., Pan, Z., Cao, X., Mo, B., Zhang, F., Raikhel, N., Jiang, L., Chen X. (2013). MicroRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in Arabidopsis. Cell **25**: 153:562-74

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

Li, X., Quon, G., Lipshitz, H.D. and Morris, Q. (2010b). Predicting *in vivo* binding sites of RNA-binding proteins using mRNA secondary structure. RNA **16:** 1096-1107.

Marshall, J.S., Stubbs, J.D., Chitty, J.A., Surin, B. and Taylor, W.C. (1997) Expression of the C_4 *Me1* gene from *Flaveria bidentis* requires an interaction between 5' and 3' sequences. Plant Cell **9:** 1515-1525.

Mathews, D.H. (2004). Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. RNA **10:** 1178-1190.

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

Muckenthaler, M.U., Galy, B. and Hentze, M.W. (2008). Systemic iron homeostasis and the iron-responsive element/iron-regulatory protein (IRE/IRP) regulatory network. Annu. Rev. Nutrition **28:** 197-213.

Newell, C.A., Brown, N.J., Liu, Z., Pflug, A., Gowik, U., Westhoff, P. and Hibberd, J.M. (2010) *Agrobacterium tumefaciens*-mediated transformation of *Cleome gynandra* L., a C_4 dicotyledon that is closely related to Arabidopsis thaliana. J. Exp. Bot. **61:** 1311-1319.

Nomura, M., Higuchi, T., Ishida, Y., Ohta, S., Komari, T., Imaizumi, N., Miyao-Tokutomi, M., Matsuoka, M., and Tajima, S. (2005a). Differential expression pattern of C_4 bundle sheath expression genes in rice, a C3 plant. Plant Cell Physiol. **46:** 754-761.

Nomura, M., Higuchi, T., Katayama, K., Taniguchi, M., Miyao-Tokutomi, M., Matsuoka, M. and Tajima, S. (2005b). The promoter for C_4-type mitochondrial aspartate aminotransferase does not direct bundle sheath-specific expression in transgenic rice plants. Plant Cell Physiol. **46:** 743-753

Nomura, M., Sentoku, N., Nishimura, A., Lin, J.H., Honda, C., Taniguchi, M., Ishida, Y., Ohta, S., Komari, T., Miyao-Tokutomi, M., Kano-Murakami, Y., Tajima, S., Ku, M.S.B. and Matsuoka, M. (2000). The evolution of C_4 plants: acquisition of *cis*-regulatory sequences in the promoter of C_4-type pyruvate, orthophosphate dikinase gene. Plant J. **22:** 211-221.

Patel, M., Corey, A.C., Yin, L-P., Ali, S., Taylor, W.C. and Berry, J.O. (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 Physiol. **136:** 3550-3561.

Rabani, M., Kertesz, M. and Segal, E. (2008). Computational prediction of RNA structural motifs involved in post-transcriptional regulatory processes. PNAS **105:** 14885-14890.

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

Sheen, J. (1999). C_4 Gene Expression. Annu. Rev. Plant Physiol. Plant Mol. Biol. **50:** 187-217.

Staiger, D. and Köster, T. (2010). Spotlight on post-transcriptional control in the circadian system. Cell Mol. Life Sci. **68:** 71-83.

Tadauchi, T., Matsumoto, K., Herskowitz, I. and Irie, K. (2001). Post-transcriptional regulation through the HO 3'-UTR by Mpt5, a yeast homolog of Pumilio and FBF. EMBO J. **20:** 552-561.

Tanz, S.K., Tetu, S.G., Vella, N.G.F. and Ludwig, M. (2009). Loss of the transit peptide and an increase in gene expression of an ancestral chloroplastic carbonic anhydrase were instrumental in the evolution of the cytosolic C_4 carbonic anhydrase in *Flaveria*. Plant Physiol. **150:** 1515-1529.

Tarun, S.Z. and Sachs, A.B. (1995). A common function for mRNA 5' and 3' ends in translation initiation in yeast. Genes Dev. **9:** 2997-3007.

Verbitskiy, D., Härtel, B., Zehrmann, A., Brennicke, A. and Takenaka, M. (2011). The DYW-E-PPR protein MEF14 is required for RNA editing at site matR-1895 in mitochondria of *Arabidopsis thaliana*. FEBS Lett. **585:** 700-704.

Wilkie, G.S., Dickson, K.S. and Gray, N.K. (2003). Regulation of mRNA translation by 5' and 3' UTR-binding factors. Trends Biochem. Sci. **28:** 182-188.

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.

Williams, B.P., Johnston, I.G., Covshoff, S. and Hibberd, J.M. (2013). Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis. *eLife* 2:e00961

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

Zhu, J., Dong, C-H., and Zhu, J-K. (2007). Interplay between cold-responsive gene regulation, metabolism and RNA processing during plant cold acclimation. Curr. Opin. Plant Biol.**10:** 290-295.

FIGURE LEGENDS

Figure 1: Promoter and intron sequences generate increased abundance of Gg-*CA4*. Promoter, coding region and UTR sequences of At-*CA4* and Gg-*CA4* were fused to the *uidA* reporter encoding GUS. Exon and intron sequences were tested by fusing the genomic coding region to *uidA*. Exon sequences alone were tested by fusing the spliced open reading frame from cDNA to *uidA*. Multiple independent T_1 lines (represented by n) were assayed for each construct. Representative images were selected from a minimum of six independent T_1 lines (A-N). All seedlings were stained for 24 h, except (B) and (C), which were imaged after 6 h of staining. Scale bars = 5 mm. GUS activity was quantified by measuring the rate of 4-methylumbeliferone (MU) synthesis (O) in at least eleven independent lines for each construct. Both the promoter and introns of Gg-*CA4* directed increased GUS activity compared with either the CaMV35S control or homologous sequences from At-*CA4*. P-values are derived from two-tailed student's t-tests.

Figure 2: A *cis*-element for M-preferential expression identified by two independent approaches. (A) Deletion analysis was performed on the 5' UTR of Gg-*CA4*. Truncated UTRs missing sequence from the 5' or 3' ends were fused to the *uidA* reporter. (B) The number of M cells expressing GUS after microprojectile bombardment of each deletion construct is expressed as a percentage of all GUS positive cells observed. Numbers within histogram bars represent the number of independently transformed cells for each construct. 5' and 3' deletions identified 13 bp and 7 bp necessary for M-specificity, respectively. Computational prediction identified a nine nucleotide motif present in all ten UTRs (C) with three completely conserved nucleotides (D). This motif (blue font) is in close proximity to the 13 bp region identified as necessary by deletion analysis (red font) (E). Asterisks denote statistical significance compared to the control ($p = <0.005$, two-tailed student's t-test), and error bars denote one standard error.

Figure 3: A five-nucleotide sequence is necessary for M-specific accumulation of multiple C_4 transcripts. Site-directed mutagenesis was used to alter five nucleotides (red font) within the Gg-*CA4* 5' UTR (A). Grey arrows denote the position of the mutated nucleotides. Histological GUS staining of transverse leaf sections from stable transgenic *G. gynandra* lines showed that the 99 base pair Gg-*CA4* 5' UTR generated M-specific accumulation of GUS (B). When the five nucleotides predicted to be important were mutated, GUS was present in M and BS cells (C). Transverse sections in (B) and (C) are representative images from three independent transgenic lines each. Scale bars = 100 μm. The same five-nucleotide motif was mutated in additional 5' and 3' UTRs. The number of M cells expressing GUS after

microprojectile bombardment of each mutant construct is expressed as a percentage of all GUS positive cells observed (D). Numbers within histogram bars represent the number of independently transformed cells for each construct. Scale bars = 100 µm. Asterisks denote statistical significance compared to the control (p = <0.005, two-tailed student's t-test), and error bars denote one standard error.

Figure 4: MEM2 confers equal transcript abundance in M and BS cells. Two chimeric constructs were synthesized by placing the nucleotides preceding the start codon of Gg-*CA4* downstream of MEM2 (red; A), and tested using microprojectile bombardment (B). Numbers within histogram bars represent the number of independently transformed cells for each construct. Results from microprojectile bombardment were validated by generating three stable transgenic lines. Error bars represent one standard error. *p = <0.01 **p = <0.001, two-tailed student's t-test (C) The percentage of GUS positive M and BS cells was quantified in stable lines expressing the Gg-*CA4* 5' UTR (3 lines) , Gg-*CA4* 5' UTR mutant (3 lines), MEM2 Sufficiency 2 chimaera (2 lines) and the CaMV35S promoter (4 lines) fused to *uidA*. Error bars represent one standard error. (D) A representative transverse section expressing MEM2 Sufficiency 2. Scale bar = 100 µm. (E) RT-qPCR quantification of *uidA* transcripts from M and BS cells expressing *uidA* under the control of the endogenous and mutated Gg-*CA4* 5' UTR. Error bars represent one standard error. Three technical replicates were performed for each line and cell type. (F) The activity of GUS measured by the rate of 4-methylumbeliferone (MU) synthesis in the same cell fractions, expressed as fold-enrichment in M cells versus BS cells. Error bars represent one standard error. Three technical replicates were performed.