The C_4Ppc promoters of many C4 grass species share a common regulatory mechanism for gene expression in the mesophyll cell

Shipan Das Gupta, Myles Levey, Stefanie Schulze, Shanta Karki, Jan Emmerling, Monika Streubel, Udo Gowik, W. Paul Quick, Peter Westhoff

the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2019) doi: 10.1111/tpj.14532

The *C_4Ppc* promoters of many C_4 grass species share a common regulatory mechanism for gene expression in the mesophyll cell

Shipan Das Gupta[1,2], Myles Levey[1], Stefanie Schulze[1], Shanta Karki[3,4], Jan Emmerling[1], Monika Streubel[1], Udo Gowik[1,5], W. Paul Quick[3,6] and Peter Westhoff[1,7,*]

[1]*Institute of Plant Molecular and Developmental Biology, Universitätsstrasse 1 Heinrich-Heine-University, 40225 Düsseldorf, Germany,*

[2]*Department of Biotechnology and Genetic Engineering, Noakhali Science and Technology University, Noakahli 3814, Bangladesh,*

[3]*International Rice Research Institute, Los Banos, Laguna 4031, Philippines,*

[4]*National Citrus Development Program, Kirtipur, Kathmandu, Nepal,*

[5]*Department of Biology and Environmental Sciences, Carl Von Ossietzky University, D-26129, Oldenburg, Germany,*

[6]*Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK, and*

[7]*Cluster of Excellence on Plant Sciences from Complex Traits towards Synthetic Modules, 40225 Duesseldorf and, 50923 Cologne, Germany*

Received 25 June 2019; revised 7 August 2019; accepted 14 August 2019.
*For correspondence (e-mail west@uni-duesseldorf.de).

SUMMARY

C_4 photosynthetic plants have evolved from C_3 ancestors and are characterized by differential expression of several hundred genes. Strict compartmentalization of key C_4 enzymes either to mesophyll (M) or bundle sheath cells is considered a crucial step towards the evolution of C_4 photosynthesis. In this study, we demonstrate that the 5′-flanking sequences of the C_4 type phospho*enol*pyruvate carboxylase (*Ppc*) gene from three C_4 grass species could drive M-cell-specific expression of a reporter gene in rice. In addition to that, we identified about 450 bp (upstream of their transcription start site) of the analyzed *C_4Ppc* promoters contain all the essential regulatory elements for driving M-cell-specific expression in rice leaves. Importantly, four motifs of conserved nucleotide sequences (CNSs) were also determined, which are essential for the activity of the promoter. A putative interaction between the CNSs and an unknown upstream element (s) is required for driving M-cell-specific expression. This work identifies the evolutionary conservation of *C_4Ppc* regulatory mechanisms of multiple closely related C_4 grass species.

Keywords: C_4 photosynthesis, mesophyll cells, phospho*enol*pyruvate carboxylase (*Ppc*) gene, conserved nucleotide sequence.

INTRODUCTION

C_4 photosynthesis is a special kind of adaptation that alleviates the detrimental effects of photorespiration, and has risen independently over 60 times from C_3 ancestors (Sage, 2004; Sage *et al.*, 2011). The majority of C_4 plants conduct two-celled C_4 photosynthesis, spatially confining the initial pre-fixation and final assimilation of CO_2 in the mesophyll (M) and bundle sheath (BS) cells, respectively. The complex process of the C_4 photosynthetic cycle is ensured by compartmentalization of C_4 enzymes in either M cells or BS cells of the leaf (Hatch, 1987). All proteins involved in C_4 photosynthesis already existed in C_3 ancestors, but are involved in other aspects of plant growth and development (Monson, 2003; Gowik and Westhoff, 2011). It is widely accepted that evolution of the C_4 photosynthetic enzymes has occurred through changes in regulation, kinetics and cell specificity of their C_3 isoforms (Gowik and Westhoff, 2011). Tissue-specific regulation of C_4-specific enzymes appears to have been a crucial step in the evolution of C_4 photosynthesis (Hibberd and Covshoff, 2010). Evolution of the M-cell-specific C_4 type phospho*enol*pyruvate carboxylase (PEPC, EC.4.1.1.31) gene is an excellent example, which has been well studied in the dicot genus Flaveria (Gowik *et al.*, 2004; Akyildiz

et al., 2007). In plants, several different isozymes of PEPC have been recognized that are involved in various metabolic processes (Ku *et al.*, 1996). The C_4 type PEPC serves the initial fixation of CO_2 in the C_4 pathway and is exclusively expressed in the M cells of C_4 leaves (Ku *et al.*, 1996; Svensson *et al.*, 2003). The photosynthetic C_4 isoform of the *Ppc* gene has evolved independently several times from the ancestral C_3 non-photosynthetic isoforms (Svensson *et al.*, 2003; Christin *et al.*, 2009). There is compelling evidence that cell-specific expression of the *C_4Ppc* might have evolved by changes in the transcriptional control of C_3 isoforms (Stockhaus *et al.*, 1997; Gowik *et al.*, 2004). Studies on the genus Flaveria revealed that the M-cell-specific expression of the *C_4Ppc*A gene is indeed determined by a 41-bp segment located in the distal region of the promoter, known as M expression module-1 (MEM-1). A similar MEM-1 homolog can also be found in the *Ppc*A promoter of the C_3 Flaveria species. However, the C_3- and C_4-specific MEM-1 differ at two positions, a G to A transition and an insertion of the tetranucleotide CACT in the C_4 MEM-1 in comparison to the C_3 one. These two changes in the MEM-1 sequence are mandatory for the M-cell-specific expression of the *C_4Ppc*A in the genus Flaveria (Gowik *et al.*, 2004; Akyildiz *et al.*, 2007).

In contrast to C_4 dicots, little is known about the regulatory element(s) and evolutionary mechanism of *C_4Ppc* genes in grasses. Previous work showed that the *C_4Ppc* promoter of *Zea mays* could drive strong M-cell-specific expression of a reporter gene in the C_3 plant rice (Matsuoka *et al.*, 1994; Kausch *et al.*, 2001). This demonstrated that, like in dicots, M-cell-specific expression of the *C_4Ppc* gene of *Z. mays* is mainly determined at the transcription level. Comparison of the C_3 and C_4 type 5′-flanking sequences of *Z. mays Ppc*s revealed that the sequences are identical at a range from a putative TATA box element to the translational start codon ATG (Schaffner and Sheen, 1992). Diverse sequences upstream of their TATA box indicate that the *C_4Ppc* gene of *Z. mays* might have evolved from an ancestral *C_3Ppc* gene after an unequal recombination event near the putative TATA box element (Schaffner and Sheen, 1992; Sheen, 1999). This recombination event might have brought all the *cis*-elements necessary for the M-cell-specific expression of *C_4Ppc*. Phylogenetic data of the grasses showed that all the C_4 grass species are clustered in the PACMAD clade comprising of 20–22 independent C_4 origins, while the evenly sized sister group BEP clade does not contain any C_4 species (Christin *et al.*, 2009). However, the M-cell-specific activity of maize *C_4Ppc* promoter (PACMAD) in C_3 rice plants (BEP) suggests the evolutionary conservation of *trans*-regulatory element(s) present between the C_4 species maize and the C_3 species rice. Until today it still remains elusive whether this evolutionary conservation is also maintained across other C_4 grass species of the PACMAD clade.

Crop plants that use the C_4 photosynthetic cycle are more productive, and use less water and nitrogen than their C_3 counterparts (Sage, 2004; Sage *et al.*, 2012). These advantages of C_4 photosynthesis, along with the demand for future crop security, have spurred the development of a program to install C_4 traits into the C_3 plant rice in order to increase the latter's agricultural productivity (Sheehy *et al.*, 2008). To establish a C_4 metabolism in rice, a set of genes that determine C_4 leaf anatomy and biochemical processes have to be transformed into rice and expressed in an appropriate manner. For this, the expression of several genes has to be redirected to either M or BS cells. Therefore, a set of cell-specific promoter units is highly needed. But the current toolbox available to drive cell-specific expression of foreign genes in rice might not be sufficient for this huge project (Schuler *et al.*, 2016). Using the same promoter, several genes for the development of multi-transgenic C_4 rice may cause silencing effects by repetitive elements (Assaad *et al.*, 1993; Hsieh and Fire, 2000). Thereby, alternative M- and BS-cell-specific promoters are needed for the development of C_4 rice. In an ideal situation, it would be more useful if a set of cell-specific *cis*-regulatory elements could be identified to develop synthetic cell-specific promoters by taking the advantage of a synthetic biology approach.

To better understand the changes underlying the evolution of *C_4Ppc* in grasses and to determine the M-cell-specific regulatory elements for C_4 rice engineering, a detailed promoter-reporter gene analysis was carried out in transgenic rice. In the present study, we aimed to characterize the anatomy of *C_4Ppc* promoters from phylogenetically closely related C_4 species of *Z. mays*. Those species could be found in the Panicoideae family. The Panicoideae is one of the subfamilies of the PACMAD clade and contain diverse sets of C_4 as well as C_3 grass species (Christin *et al.*, 2009). Most of the commercially important C_4 crop plants like *Z. mays*, *Sorghum officinare*, etc. belong to the Panicoideae. Here, we investigated the 5′-flanking sequences of the *C_4Ppc* genes from the Panicoid C_4 grass species *Urochloa maxima*, *Panicum miliaceum* and *Setaria viridis*. The three 5′-flanking sequences showed M-cell-specific activity in rice when fused with the *uid*A (GUS) reporter gene. A detailed promoter-reporter gene analysis in rice revealed that the proximal 500 bp of these *C_4Ppc* 5′-flanking sequences are sufficient to drive M-cell-specific expression. Further analysis of the proximal 500 bp identified four motifs with conserved nucleotide sequence (CNS), which are essential for the *C_4Ppc* promoter activity. The presence of the CNSs in the six analyzed Panicoid C_4 grass species indicates a high degree of conservation of the *C_4Ppc* regulatory mechanism. We also address what are the putative core regulatory element(s) of these promoters responsible for directing the activity and M-cell specificity.

RESULTS

The *C_4Ppc* 5′-flanking sequences of C_4 Panicoid grasses are capable of driving M-cell-specific expression in transgenic rice

In order to determine how the *C_4Ppc* evolved in grasses, the 5′-flanking sequences of the *C_4Ppc* genes were isolated through vectorette polymerase chain reaction (PCR; see Experimental procedures) from the Panicoid C_4 grasses *S. viridis* (*SvPpc*), *P. miliaceum* (*PmPpc*) and *U. maxima* (*UmPpc*), which exhibit NADP-ME, NAD-ME and PCK types of C_4 photosynthesis, respectively (Siebert *et al.*, 1995). However, little is known about the regulatory elements and M-cell-specific regulatory mechanism of the three Panicoid *C_4Ppc* genes. Therefore, to increase the probability to retain all the essential *cis*-regulatory elements of the promoters, we aimed to isolate about 2–3.5 kb of 5′-flanking sequences of the three Panicoid *C_4Ppc* genes for primary analyses. The length of the three isolated 5′-flanking sequences of *SvPpc*, *PmPpc* and *UmPpc* were 2233 bp, 3327 bp and 3531 bp, respectively (Figure 1a). To test the activity of these 5′-flanking sequences in rice, the three putative promoters were fused to the coding region of the *uid*A (GUS) gene linked to the terminator of the nopaline-synthase (Nos) gene (Figure 1a). All the chimeric constructs were stably transformed into rice for analysis.

GUS histochemical analysis of these three putative *C_4Ppc* promoters showed that they were able to keep their C_4 characteristic M-cell-specific expression pattern in rice (Figure 1b), similar to what has been found for the *C_4Ppc* promoter of *Z. mays* (Figure S1). No blue GUS staining was observed in the upper or lower epidermis, vascular tissue and BS cells of the rice leaf blade across several sections (Figure 1b). Although the GUS expression pattern of the three putative promoters in rice was very similar, the levels of GUS activity driven by the three putative promoters differed slightly from each other. In comparison to the *SvPpc* 5′-flanking sequences, the *PmPpc* and the *UmPpc* 5′-flanking sequences showed 2.7- and 4.9-fold lower GUS activities (Figure 1c). Nevertheless, the GUS histochemical and quantitative analysis data suggested that the *trans*-regulatory system responsible to direct M-cell-specific expression of *C_4Ppc* genes in C_4 grass plants is also present in the C_3 grass rice. Therefore, the isolated 5′-flanking sequences can be used as M-cell-specific drivers for rice leaf tissue.

The proximal 500 bp of the 5′-flanking regions of *UmPpc*, *PmPpc* and *SvPpc* are capable to drive M-cell specificity in rice leaf

To better understand the anatomy of a grass *C_4Ppc* promoter, we used the *UmPpc* promoter as a starting point aiming to identify the minimal set of *cis*-regulatory element(s) that are essential for its M-cell specificity. The available 5′-flanking sequence of *UmPpc* was 3531 bp long.

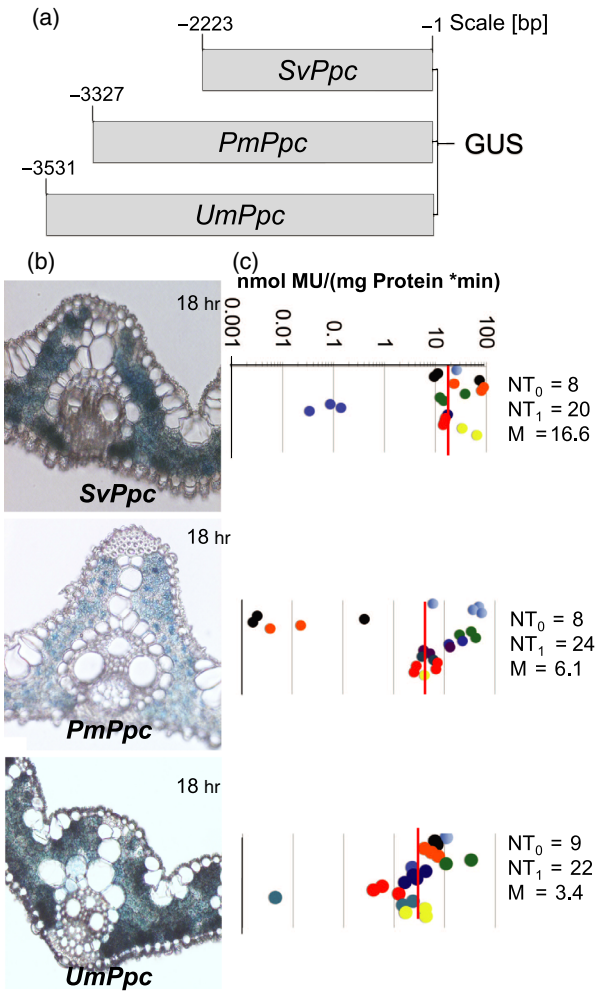

Figure 1. Analysis of the *C_4Ppc* 5′-flanking sequences from *Setaria viridis*, *Panicum miliaceum* and *Urochloa maxima* in transgenic rice.
(a) Schematic presentation of the *C_4Ppc* 5′-flanking sequence GUS reporter gene constructs from Panicoid grass species *S. virdis*, *P. miliaceum* and *U. maxima*. Nucleotide numbers refer to the translational start codon (ATG). Here, all three chimeric constructs were stably transformed in the Nipponbare rice cultivar.
(b) Histochemical analysis of GUS in leaf cross-sections of transgenic rice plants transformed with *SvPpc*:GUS, *PmPpc*:GUS and *UmPpc*:GUS constructs, respectively. The incubation period was 18 h.
(c) Quantitative measurement of GUS activities in leaves of transgenic rice plants. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per mg of protein per minute. Each single dot represents one T_1 transgenic plant, and the same color in dots indicates the plants belong to the same transgenic line. The total number of transgenic plants investigated (NT_1) from the independent transgenic (NT_0) lines and the median value (M) of GUS activity are represented at the right side of each column. Relative positions of the median values in each column are marked with black lines.

To delimit it, the 5′-flanking sequence of *UmPpc* was successively truncated, and the resulting three fragments with lengths of 1500 bp, 1000 bp and 500 bp were fused to the GUS reporter gene. The chimeric constructs *UmPpc*-1500:GUS, *UmPpc*-1000:GUS and *UmPpc*-500:GUS were transformed into rice for analysis.

Mature leaves from the transgenic T_0 rice plants were analyzed for GUS activity. Histochemical staining of leaf cross-sections of the *UmPpc*-1500:GUS, *UmPpc*-1000:GUS and *UmPpc*-500:GUS transformants showed GUS activity only in the M cells (Figure 2b). The same staining pattern was previously observed in rice transformants containing the full-length *UmPpc* 5′-flanking sequences (Figure 1b). Fluorometrical quantification of GUS activity of the three constructs led to almost equal amounts of activity (Figure 2c). Taken together there was no notable difference among the three dissected promoters in terms of overall activity and M-cell specificity. We concluded from these results that the *cis*-regulatory element(s) of the *UmPpc* promoter responsible for M-cell specificity and activity are located within the first 500 bp upstream of the translational start codon ATG.

This finding prompted us to test whether the other grass *C*$_4$*Ppc* promoters also contain their M-cell-specific regulatory elements within their proximal part. To determine this, the proximal 500 bp of the *SvPpc* and *PmPpc* sequences were fused to the GUS reporter gene and subsequently transformed into rice for analysis (Figure 2a). Histochemical analysis of GUS staining in the leaves of *SvPpc*-500:GUS and *PmPpc*-500:GUS transformants revealed that the GUS expression pattern of both constructs as well as their promoter strength were essentially the same as that of the *UmPpc*-500:GUS construct (Figure 2b and c). With respect to the C4Ppc of maize, the proximal 500 bp of its 5′-flanking sequence *ZmPpc* promoter was found to be a very weak promoter in rice, but extending the sequence to 660 bp upstream – a similar promoter activity was observed as for the other three C_4 grasses (Figure S1). This is consistent with the findings of Taniguchi *et al*. (2000) who had previously shown that the proximal 650 bp of the C$_4$*Ppc* promoter was sufficient to drive M-cell-specific expression.

Taken together, the data of the *UmPpc*-500, *PmPpc*-500, *SvPpc*-500 and *ZmPpc*-660 constructs provided convincing evidence that the *cis*-regulatory element(s) driving M-cell specificity and activity are located within the proximal part of the 5′-flanking region of the analyzed grass *C*$_4$*Ppc* promoters.

Mapping of *cis*-regulatory elements in the proximal regions of the *C*$_4$*Ppc* promoters

The data presented so far clearly demonstrate that the proximal regions of our selected C_4 grass *Ppc* promoters contain all the relevant regulatory elements for M-cell-specific expression. In order to gain insight into and to identify putative *cis*-element(s), it would be helpful to have a detailed sequence map of *C*$_4$*Ppc*-500 fragments from grasses. To determine whether there are any CNSs present among the Panicoid C_4 grass *Ppc* promoters, we pursued a comparative analysis.

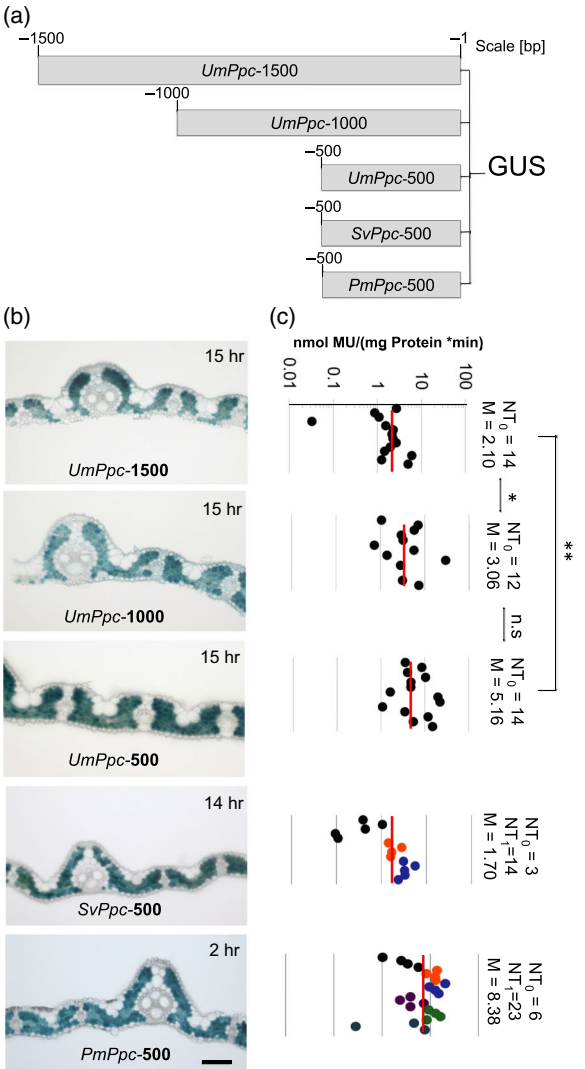

Figure 2. 5′-Deletion analysis of the *C*$_4$*Ppc* 5′-flanking sequences from *Urochloa maxima, Setaria viridis* and *Panicum miliacieum* in transgenic rice.
(a) Schematic presentation of the chimeric constructs of the *UmPpc, SvPpc* and *PmPpc* 5′-flanking sequences. Here, the *UmPpc*-1500:GUS, *UmPpc*-1000:GUS and *UmPpc*-500:GUS constructs were stably transformed in the Kitaake, whereas the *SvPpc*-500:GUS and *PmPpc*-500:GUS constructs were transformed in Nipponbare rice cultivar.
(b) Histochemical localization of GUS (blue staining) in leaf cross-sections of transgenic rice leaves. The incubation period for the *UmPpc* chimeric constructs was 15 h. In the cases of *SvPpc*-500:GUS and *PmPpc*-500:GUS constructs, the incubation periods were 14 h and 2 h, respectively. Scale bar: 50 μm.
(c) Fluorometrical quantification of GUS activity. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per mg of protein per minute. The number of transgenic plants (NT_1) analyzed from the independent transgenic (NT_0) lines and the median values (M) of GUS activity for each construct are represented at the right side of each column. The same color dots represent the transgenic T_1 plants originate of the same T_0 line. The positions of the median values are marked with red lines (*$P < 0.05$; **$P < 0.005$; n.s., not significant $P > 0.05$).

In order to broaden the sequence comparisons, two additional 5′-flanking sequences of the *C*$_4$*Ppc* gene from the C_4 Panicoid grasses *Digitaria sanguinalis* and

Setaria italica were isolated through vectorette PCR (genome walking; Siebert *et al.*, 1995). Sequence comparison of the proximal 500 bp of the *UmPpc* 5′-flanking sequence with the five selected grass *C_4Ppc* 5′-flanking sequences (*Z. mays*, *S. viridis*, *S. italica*, *P. miliaceum* and *D. sanguinalis*) identified four motifs of CNSs (Figure 3) that were named CNS-1, CNS-2, CNS-3 and CNS-4, respectively. In the case of *ZmPpc,* CNS-3 is divided into CNS3a and CNS3b, due to the insertion of six additional nucleotides in the CNS-3 region. An equally compelling possibility is that CNS-3 is actually shorter and may not contain the repetitive CACA (Figure 3) element and thus the CNS-3 module would not be divided in *Z. mays*. In the *UmPpc* promoter the four CNSs are located in the range of −159 bp to −73 bp (Figure 3). Downstream to these CNSs, a putative TATA box is also found to be highly conserved in all six *C_4Ppc* 5′-flanking sequences. In the 5′-flanking region of *UmPpc*, the putative TATA box element was detected at the position −72 to −64 bp. Rapid amplification of 5′-complementary DNA (cDNA) ends (5′-RACE) analysis was used to define the putative transcription site of the *UmPpc*, *SvPpc*, *SiPpc*, *PmPpc* and *DsPpc* *C_4Ppc* genes (Figure S2). In addition, information about the 5′-UTR regions of the *ZmPpc* was obtained from available Genbank data libraries (Accession No: X15239). Besides the CNSs regions and the putative TATA box sequence, there was no other noticeable sequence conservation observed among the six selected grass *C_4Ppc*-500 fragments. This observation suggests that the CNSs-harboring region could be the putative region that plays a crucial role within *C_4Ppc* promoters that might carry the M-cell-specific determinants.

Deletion of the CNSs comprising region from the *UmPpc*-500 fragment causes loss of activity

To unravel the potential role of the identified CNSs, it was necessary to perform promoter deletion and recombination tests with the analyzed *C_4Ppc* 5′-flanking sequences of the Panicoid grass, and we chose the *UmPpc*-500 fragment as an example. The fragment was subdivided into four different subfragments; 1, 2, 3 and 4, starting upstream of the TATA box (Figure 4a). Subfragment-1 contains all the CNSs that were identified in the six selected C_4 grass *Ppc* 5′-flanking sequences. Subfragments 2, 3 and 4 do not have any potential CNSs in the six C_4 grass *Ppc* promoters. With the four subfragments, five consecutive promoter deletion and recombination constructs were created and designated as *UmPpc*-500-del-4:GUS, *UmPpc*-500-del-3:GUS, *UmPpc*-500-del-2:GUS, *UmPpc*-500-del-1:GUS and *UmPpc*-500-region-1:GUS, respectively (Figure 4a). All five chimeric constructs were stably transformed into rice.

GUS histochemical analyses of transgenic rice leaf blade cross-sections showed that the individual removal of subfragments 4, 3 and 2 from the *UmPpc*-500 fragment had no effect on the spatial accumulation of GUS activity (Figure 4b). The GUS expression patterns were essentially the same as the pattern observed with the entire *UmPpc*-500 fragment. Differences were observed with respect to the promoter activities of the constructs del4 to del2 (Figure 4c). In contrast, deletion of subfragment-1 from the *UmPpc*-500 fragment resulted in a complete loss of GUS staining, indicating that subfragment-1 is essential for promoter activity (Figure 4b). Surprisingly, however, subfragment-1 together with its downstream sequences was not able to drive GUS expression in the M cell of rice leaf (Figure 4b). This implies the presence of additional *cis*-regulatory elements upstream of subfragment-1. To test this hypothesis, subfragment-1 was fused with either subfragment-2 or -3. Figure 5 shows that the addition of either of these two fragments to subfragment-1 restores the M promoter activity. This finding indicates a redundancy of *cis*-regulatory elements located in subfragments-2 and -3.

The CNSs-harboring region is also indispensable in the *Setaria viridis C_4Ppc* promoter to drive activity

The data obtained with the *UmPpc* proved that the CNSs-harboring subfragment-1 are absolutely essential for promoter activity. To test whether this is also the case for other Panicoid C_4 grass species, a promoter deletion and complementation assay was carried out with the *SvPpc*-500 fragment. First, the CNS1-4 harboring region (92 bp) of the *SvPpc*-500:GUS construct was deleted (Figure 6a). In the second construct the CNS1-4 region (92 bp) of *SvPpc* was exchanged with the CNS1-4 region (87 bp) of the *UmPpc* 5′-flanking sequence (Figure 6a). The two chimeric constructs were designated as *SvPpc*-500Δ92:GUS and *SvPpc*Δ92+*Um*87:GUS, respectively, and transformed into rice.

Removal of the 92-bp CNS1-4 region from the *SvPpc*-500:GUS construct reduced the promoter activity almost completely. Out of 15 transgenic plants, four plants produced weak GUS expression in the M cell (Figure 6b; the respective fluorometric value to the shown cross-section is marked with a black arrow in Figure 6c). The remaining 11 transgenic plants showed extremely low levels of GUS activity and did not reveal any blue staining in the histochemical assay. We therefore concluded that the CNS1-4 harboring region of the *SvPpc*-500 fragment is the essential cis-determinant of the proximal promoter region. When the deleted *SvPpc* CNS1-4 region (92 bp) was replaced by the *UmPpc* CNS1-4-containing region (87 bp) promoter activity was regained (Figure 6). Out of 15 independent transgenic plants analyzed, all displayed high GUS activity in the M cells (Figure 6b). We conclude from this deletion and heterologous substitution experiments that the CNS1-4-containing region is obligatory for *C_4Ppc* promoter activity of the analyzed C_4 Panicoid grass species. The conservation of the CNSs in other C_4 grasses of the Panicoids suggests their general importance in *C_4Ppc* promoter function.

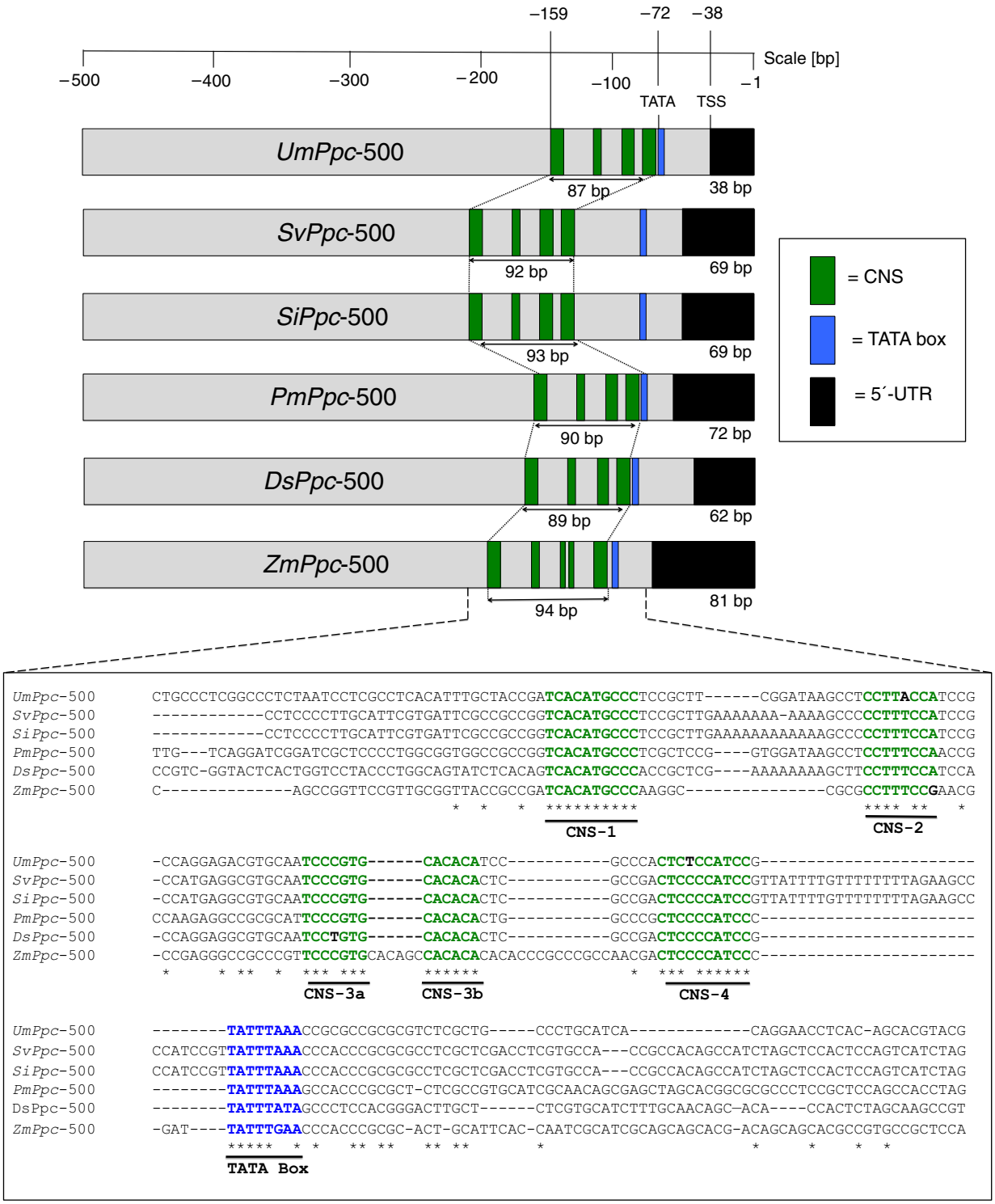

Figure 3. Comparison of the proximal 500 bp of the 5′-flanking regions of the *C₄Ppc* genes from the Panicoid C_4 grass species *Urochloa maxima* (*UmPpc*), *Setaria viridis* (*SvPpc*), *Setaria italica* (*SiPpc*), *Panicum miliaceum* (*PmPpc*), *Digitaria sanguinalis* (*DsPpc*) and *Zea mays* (*ZmPpc*).
The schematic figure of the top of the figure depicts the relative positions of the conserved nucleotide sequences (CNSs), putative TATA box and the 5′-untranslated region (5′-UTR). The CNSs are indicated in green, the putative TATA box in blue and the putative 5′-UTR in black squares. The relative lengths of the CNS harboring regions and the 5′-UTRs are indicated in bp below each respective square. Nucleotide numbers refer to the translational start codon ATG (+1), and the positions are related to the black scale bar on top. The complete sequence of the analyzed *C₄Ppc* 5′-flanking sequence can be found in the GenBank data library under the accession numbers: MH675616 (*UmPpc*), MH675618 (*SvPpc*), AF495586 (*SiPpc*), MH675617 (*PmPpc*), MH675615 (*DsPpc*) and X15239 (*ZmPpc*). TSS, putative transcription start site. The asterisk (*) indicates fully conserved base pairs.

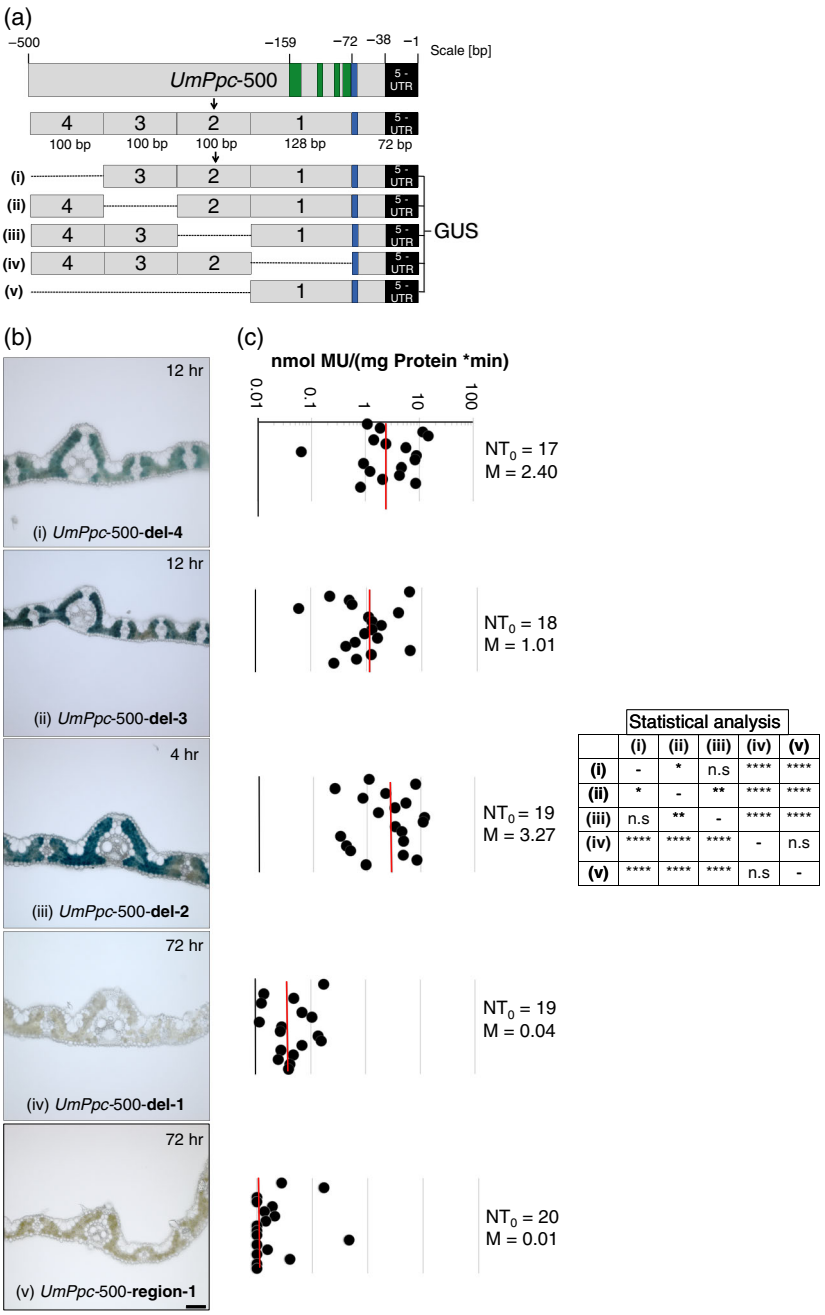

Figure 4. Deletion and recombination analysis of the *UmPpc*-500 proximal promoter fragments.
(a) Schematic presentation of the subdivision of the *UmPpc*-500 promoter fragment into four subfragments. Subfragment-1 (128 bp) consists of all four putative conserved nucleotide sequences (CNSs), while subfragments-2, 3 and 4 (100 bp each) do not have any conserved sequence. Based on the division of regions, five promoter deletion and recombination constructs: (i) *UmPpc*-500-del-4:GUS; (ii) *UmPpc*-500-del-3:GUS; (iii) *UmPpc*-500-del-2:GUS; (iv) *UmPpc*-500-del-1:GUS; and (v) *UmPpc*-500-region-1:GUS were prepared. The TATA box and the 5′-UTR are represented in a blue and black square, respectively. All five chimeric constructs were stably transformed into Kitaake rice cultivar.
(b) Histochemical localization of GUS (blue staining) in leaf cross-sections of transgenic Kitaake rice. Incubation times for (i) *UmPpc*-500-del-4:GUS, (ii) *UmPpc*-500-del-3:GUS and (iii) *UmPpc*-500-del-2:GUS were 12, 12 and 4 h, respectively, and for constructs (iv) *UmPpc*-500-del-1:GUS and (v) *UmPpc*-500-region-1:GUS were 72 h. Scale bar: 50 μm.
(c) GUS activities in leaves of transgenic rice plants. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per mg of protein per minute. The number of transgenic plants analyzed (NT_0) and the median (M) values of the GUS activity for each construct are represented at the right side of each column. The positions of median values are marked in red lines (*$P < 0.05$; **$P < 0.005$; ** **$P < 0.00005$; n.s., not significant $P > 0.05$).

The nucleotide sequence between the CNSs does not have any impact on directing activity and M-cell specificity

The comparative analysis of the six selected C_4*Ppc* 5′-flanking sequences showed there is some degree of conservation of the nucleotide sequences between the sequences of the four identified CNSs (Figure 3). One might speculate that not only the four putative CNSs are important, but also the nucleotides between the CNSs may have an impact on directing the activity and specificity. To address this question, we selected the nucleotide sequences between CNS-1 and CNS-2 (18 bp; Figure 3), and CNS-2 and CNS-3 (19 bp; Figure 3) of *UmPpc*-500-del-4:GUS and replaced them by 18 or 19 bp, respectively, of random sequences of the coding region of the green fluorescent protein (GFP) gene. GUS histochemical and quantitative analyses of 16 independent transgenic rice plants containing the *UmPpc*-fGFP:GUS construct (Figure S3) showed a comparable GUS expression to the *UmPpc*-500-del-4:GUS construct (Figure 4). This indicates that the CNS spacer nucleotide sequences, at least between CNS-1 and CNS-2 as well as between CNS 2- and CNS-3, have no major impact in directing activity and M-cell specificity.

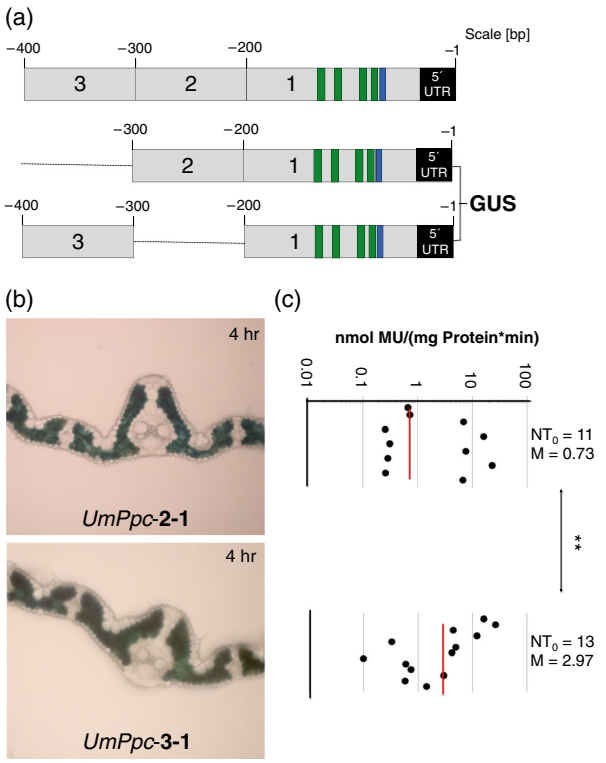

Figure 5. Functional analysis of subfragments-1, 2 and 3 fusion constructs in leaves of transgenic rice.
(a) Schematic presentation of *UmPpc*-2-1:GUS and *UmPpc*-3-1:GUS constructs transformed into Kitaake.
(b) Histochemical localization of GUS (blue staining) in leaf cross-sections of transgenic Kitaake rice. The incubation time was 16 h.
(c) GUS activities in leaves of transgenic rice plants. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per mg of protein per minute. The number of transgenic plants analyzed (NT_0) and the median (M) values of the GUS activity for each construct are represented at the right side of each column.

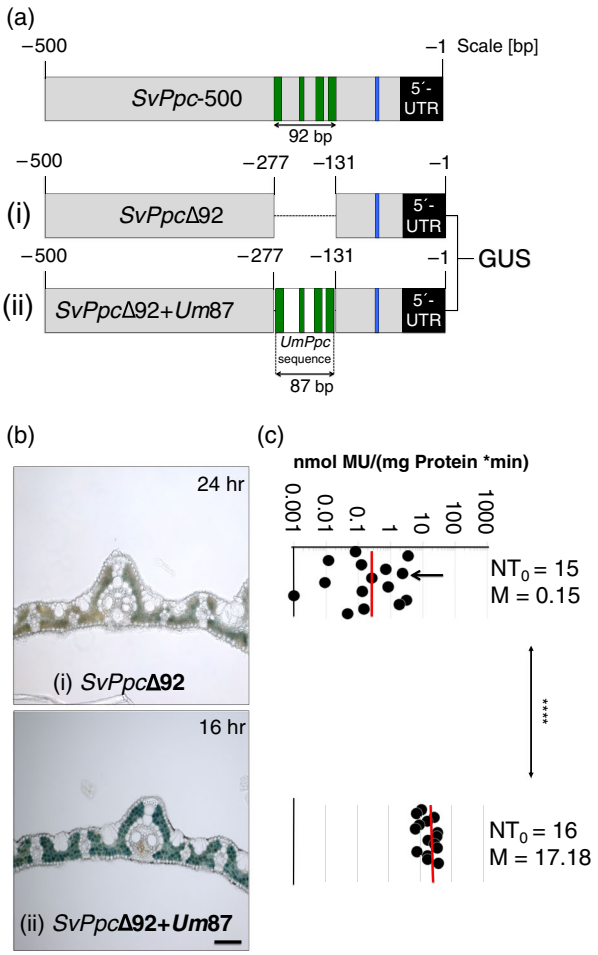

Figure 6. Deletion and complementation analysis of conserved nucleotide sequences (CNS)1-4 containing region of the *SvPpc*-500.
(a) Overview of the chimeric deletion and complementation constructs of the CNS1-4 harboring region of the *SvPpc*-500:GUS construct. Here, both the chimeric constructs were transformed in the Kitaake rice cultivars.
(b) Histochemical analysis of transgenic rice plants transformed with *SvPpc*-500Δ92:GUS and *SvPpc*Δ92+*Um*87:GUS, respectively. Scale bar: 50 μm. The incubation times were 24 and 18 h for *SvPpc*-500Δ92:GUS and *SvPpc*Δ92+*Um*87:GUS transformants, respectively.
(c) GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per mg of protein per minute. The number (NT_0) of independent transgenic plants analyzed and the median (M) values of GUS activity for each construct are represented at the right side of each column. The positions of median values are marked in red lines (****$P < 0.00005$).

The TATA and 5′-UTR segment of the *UmPpc* promoter can be replaced with a minimal 35S promoter

It was tempting to speculate whether the TATA box and 5′-UTR segment (72 bp) of the *UmPpc* 5′-flanking sequence contain any important *cis*-regulatory element(s) and, if so, it may have an effect on directing M-cell-specific expression. Therefore, we replaced this 72-bp sequence from the *UmPpc*-500-del-4:GUS construct (Figure 4) with the −60 bp minimal 35S promoter (M35S) of the cauliflower mosaic virus, resulting in the formation of the *UmPpc*-M35S:GUS construct (Figure S4a). It is known that this segment of the 35S promoter is not able to drive expression of the GUS gene (Ow *et al.*, 1987; Wei *et al.*, 2006; Kirschner *et al.*, 2018).

After examining 18 independent transgenic rice plants of the *UmPpc*-M35S:GUS construct, we did not find any substantial difference in terms of activity and M-cell specificity between the *UmPpc*-M35S: GUS construct (Figure S4) and the *UmPpc*-500-del-4:GUS (Figure 4). This clearly indicates that neither the putative TATA box nor the 5′-UTR of the *UmPpc* promoter contain any relevant *cis* element(s) for directing activity and M-cell specificity.

DISCUSSION

C_4 photosynthesis is a complex biochemical trait, which has evolved over 60 times in the angiosperms and at least 22 times independently in the grasses (Sage, 2016). The

polyphyletic origin of C4 photosynthesis implies that it must have been relatively easy, in genetic terms, to evolve this pathway and that evolution could rely on already existing modules. In 1993, Matsuoka *et al*. (1993) reported that the *C_4Ppc* promoter of maize maintained its M-cell-specific expression in transgenic C_3 rice. This finding implies that the gene regulatory system acting in the leaves of rice and maize must share common *trans*-regulatory factors that recognize cell-specific *cis*-regulatory modules in the C_4 maize promoter and interpret them correctly in rice. Maize is a member of the subfamily Panicoideae, which is part of the PACMAD clade (Christin *et al*., 2009) in which all the C_4 grasses are clustered. A complete molecular phylogenetic tree developed for the grass family showed that rice belongs to the BEP clade, which diverged from the PACMAD clade more than 50 million years ago (Christin *et al*., 2008). Despite this large evolutionary distance, maize and rice share common *trans*-regulatory element(s) for M-cell-specific expression of *C_4Ppc* gene. One could assume that the *trans*-regulatory factors that are necessary for M-cell-specific expression of the *C_4Ppc* gene already existed in a common ancestor of the PACMAD and BEP clades. To gain support for this hypothesis, we wanted to explore whether the *C_4Ppc* 5′-flanking sequences of other C_4 grasses also function as mesophyll-specific promoters in rice. Therefore, we isolated *C_4Ppc* 5′-flanking sequences from three further C_4 Panicoid grass species (i.e. *S. viridis*, *P. miliaceum* and *U. maxima*) and subsequently analyzed their promoter potentiality in rice.

All the three *C_4Ppc* 5′-flanking sequences were able to maintain their C_4-type cell-specific expression when tested in rice (Figure 2), thus supporting the proposed evolutionary scenario of Matsuoka *et al*. (1994). A similar evolutionary conservation of a transcription regulatory system has also been observed for the BS cells of distantly related dicotyledonous C_3 and C_4 plants. The BS-specific promoter of the *GLDPA gene* (encoding the P subunit of glycine decarboxylase) of the C_4 Asteracean species *Flaveria trinervia* maintains its promoter specificity in the Brassicaceae C_3 species *Arabidopsis thaliana* (Engelmann *et al*., 2008; Wiludda *et al*., 2012). Conversely, the promoter of the sulfate transporter 2;2 gene (*SULTR2;2*) from Arabidopsis has been found to be BS cell specific in the Asteracean C_4 species *Flaveria bidentis* (Kirschner *et al*., 2018). These findings indicated that the *trans*-regulatory factors of the gene transcription system of BS cells are at least partially evolutionary conserved in dicotyledonous plants. Therefore, the evolutionary conservation of the *trans*-regulatory systems of M cells in the C_3 and C_4 grass species is not an uncommon phenomenon. However, it cannot be neglected that, in many cases, a parallel evolution of *trans*-acting factors also had a major impact. Nomura *et al*. (2005) could demonstrate that in case of two BS-specific promoters of the C_4 species maize tested, only one kept its specificity when expressed in the C_3 species rice. Also, for the *C_4Ppc* gene of the C_4 Flaveria species a *cis*-element was identified that underwent minor changes from C_3 to C_4 Flaveria species, but the C_4 variant could not confer mesophyll-specificity in Arabidopsis (Gowik *et al*., 2004; Akyildiz *et al*., 2007). Both studies show that the necessary *trans*-factors are either not present in the tested C3 species or not regulated in the appropriate way for cell-specific expression. Brown *et al*. (2011) could demonstrate that for the BS-cell-specific expression of the NADP malic enzyme of *Cleome gynandra* (C_4), part of the coding sequence is responsible for the cell-specific expression. The necessary sequence is already present in the C_3 species Arabidopsis, but only in the C_4 plant the *trans*-factors function appropriately to confer the BS-cell-specific expression (Brown *et al*., 2011). An in-depth analysis of the expression behavior of transcription factors in both *C. gynandra* and *Z. mays* revealed that many of them show a very similar expression pattern in various stages of leaf development (Aubry *et al*., 2014). This indicates that in both independent evolutions of C_4 photosynthesis the same set of transcription factors was recruited to achieve the same outcome (Aubry *et al*., 2014).

One of the major interests of our present study was to define the minimal promoter of *C_4Ppc* that is capable of directing similar patterns of M-cell-specific expression as their full-length 5′-flanking sequences. In the present study, we found that the proximal 500 bp of the *C_4Ppc* 5′-flanking sequences of *SvPpc*, *PmPpc* and *UmPpc* showed strong M-cell-specific expression in transgenic rice leaves, suggesting all the *cis*-regulatory elements required for promoter strength and specificity are contained in about 450 bp (upstream of the 5′-UTR). This finding indicates that the M-cell-specific regulatory mechanism of the *C_4Ppc* gene of the Panicoid grasses differs from that of the *ppc*A gene of the C_4 Asteracean species *F. trinervia*, albeit both the genes were experimentally proven to be regulated at a transcriptional level (Matsuoka *et al*., 1994; Gowik *et al*., 2004). The M-cell-specific expression of the C_4 *ppc*A promoter of *F. trinervia* in the transformable, closely related C_4 species *F. bidentis* was found to be regulated by two important regions, a proximal fragment up to −570 bp (with respect to the ATG start codon) and a distal region from −1551 to −2141 bp. The proximal part of the promoter is mainly responsible for the quantitative expression of the *ppc*A gene. The distal part of the promoter contains a 41-bp enhancer-like expression module-1 (MEM-1), and is thought to be responsible for the M-cell-specific expression of the C_4 *ppc*A gene by suppressing its expression in the BS cells (Gowik *et al*., 2004; Akyildiz *et al*., 2007). In contrast to the C_4 *ppc*A promoters of dicot Flaverias, the analyzed *C_4Ppc* promoters of Panicoid grasses contain their all-regulatory elements in the proximal part of their promoters. These data clearly suggest that the length of the *ppc*A promoter of Flaveria is larger than the *C_4Ppc*

promoters of the Panicoid grasses. A cap analysis of gene transcription performed by Mejia-Guerra *et al.* (2015) predicted that many of the *Z. mays* genes, but by no means all, have sharp transcription start site clusters. In addition to that, a total of 38% of promoters with the sharp transcription start site have been reported to possess a broader TATA consensus motif and are assumed to associate with the tissue-specific expression of those respective genes (Mejia-Guerra *et al.*, 2015). Genome-wide chromatin immunoprecipitation experiments performed in the dicot plant *A. thaliana* showed that transcription factors bind with the highest frequency close to the transcription start site of their target gene promoters (Heyndrickx *et al.*, 2014). Concerning the position of transcription factor binding sites with respect to the gene, it is assumed that the core promoter elements of most maize genes are located up to 1 kbp upstream of their transcription start sites (Mejia-Guerra *et al.*, 2015). A further study of a gene regulatory network for the phenolic pathway supported this assumption (Yang *et al.*, 2017). Taken together these results indicate that the length of most maize promoters is relatively small, and our findings with the *C_4Ppc* promoters are completely in line with this hypothesis. It is still unclear at present which selective pressure is favoring the short promoter sizes of maize and other Panicoid grass species.

To localize regulatory elements within the *C_4Ppc* promoters of the Panicoid grasses, we compared the proximal 500 bp of *C_4Ppc* 5′-flanking sequences from six different Panicoid C_4 grasses, including *Z. mays*, for which genome sequences were available. We observed that all *C_4Ppc* 5′-flanking sequences of the six Panicoid grasses share four motifs of CNSs (Figure 3). To the best of our knowledge, the role of these CNSs in grass *C_4Ppc* promoters is still elusive. Although a recent publication by Gorska *et al.* (2019) identified the binding sites of two bHLH TFs for the C_4*Ppc* promoter of *Z. mays* that contribute to its cell-specific expression, and it is intriguingly identical with CNS1 we identified in this study. Downstream to these CNSs, a putative TATA box can be detected in all the six analyzed 5′-flanking sequences. For metazoan transcription, it has been suggested that promoters (Type-I promoters) possessing a TATA motif are usually involved in tissue-specific expression of the respective genes (Engstrom *et al.*, 2007; Lenhard *et al.*, 2012). The presence of a putative TATA motif in *C_4Ppc* promoters is, therefore, in line with this suggestion. Because the proximal parts of the *ZmPpc*, *SvPpc*, *PmPpc* and *UmPpc* promoters gave comparable GUS expression in transgenic rice leaves, the existence of a conserved TATA box and CNSs in their sequences indicate a common regulatory mechanism among these *C_4Ppc* genes.

The necessity of the CNSs for promoter activity was proven by the deletion of the region containing the CNSs from the *UmPpc-500* and *SvPpc*-500 fragments (Figures 4 and 6). This indicates the existence of a quantitative module in the CNSs-harboring region. Interestingly, all four CNSs were also detected in the proximal 500 bp of a *ppc* homolog gene (*ppc1*) from the C_3 Panicoid grass *Dichanthelium oligosanthes* (Figure S5). However, in comparison to the expression of *C_4Ppc* of C_4 Panicoid grass, the expression level of the *ppc1* gene was almost undetectable in *D. oligosanthes* (Studer *et al.*, 2016). When testing the 5′-flanking sequence of the *ppc*1 gene of *D. oligosanthes* for promoter activity in rice, no M-cell-specific expression was observed; GUS expression could also be detected in the vascular bundle and BS cells (Figure S5). This demonstrates that the CNSs are not the only players for regulating the M-cell-specific expression of the *C_4Ppc* gene of Panicoid grasses. This assumption was proven when we identified that the CNSs-harboring region – together with its basal sequences (TATA and 5′-UTR segment) of the *UmPpc* promoter – was not sufficient to drive GUS expression in the M cells (Figure 5) of a rice leaf; only when either region 2 or 3 were added was M-cell-specific GUS expression observed (Figures 4 and 5). Additional regulatory elements upstream to the CNSs-harboring region are therefore essential for driving M-cell specificity of transcription. Because both region 2 or region 3 could provide these additional *cis*-regulatory elements they must be redundant. Interestingly, a sequence comparison of these two fragments in the *C_4Ppc* promoters Panicoid grasses did not detect any sequence similarities between the two regions.

It has been found in several studies that cell-specific regulators are often redundant in nature without possessing identical or repeated sequences (Leyva *et al.*, 1992; Hatton *et al.*, 1995; Kirschner *et al.*, 2018). The phenylalanine ammonia-lyase-2 (*PAL2*) promoter of *Phaseolus vulgaris* L. and the sulfate transporter-2;2 (*AtSULTR*2;2) promoter of Arabidopsis are possibly the best examples of such a scenario in plants (Leyva *et al.*, 1992; Hatton *et al.*, 1995; Kirschner *et al.*, 2018). This study adds the *C_4Ppc* promoters of Panicoid grasses to this list.

Engineering the C_4 pathway into current C_3 crops, that is rice, requires the transformation of quite a large number of genes. This in turn necessitates a toolbox filled with various cell-specific promoters for driving the expression of those transgenes. The results presented in this study add three additional mesophyll-specific promoters for rice to this toolbox. The identification of *cis*-regulatory elements/modules within these promoters should also aid in the construction of synthetic minimal promoters, which would greatly ease the genetic engineering efforts.

EXPERIMENTAL PROCEDURES

Rapid amplification of 5′-cDNA ends (5′-RACE PCR)

The 5′-UTR of the *SvPpc*, *PmPpc*, *UmPpc* and *DsPpc* was determined using the 5′-RACE PCR following the manufacturer's

manual of either the SMART cDNA Synthesis Kit or the SMARTer™ RACE cDNA Amplification Kit (Clontech Laboratories, Mountain View, USA). At the first stage of this process, total RNA from the leaves of respective wild-type plants of *S. viridis*, *P. miliaceum*, *U. maxima* and *D. sanguinalis* were extracted with the RNeasy® Plant Mini Kit (Qiagen, Hilden, Germany); 1 μg RNA was used for the cDNA first strand synthesis in all cases. Afterwards, the PCR amplification of the 5′-UTRs was performed using the Advantage® 2 DNA Polymerase Mix (Clontech Laboratories, Mountain View, USA) or the Phusion HF DNA Polymerase (Thermo Fisher, Waltham, USA). The 3′-gene-specific oligonucleotides that were employed in combination with either the SMART-II-A 5′-Primer or the SMARTer™ Universal Primer A Mix (UPM) in the 5′-RACE PCR are listed in Table 1.

The resulting products of the 5′-RACE PCR were cloned into the pJET1.2/Blunt vector (Thermo Fisher, Waltham, USA) and sequenced (LGC Genomics, Berlin, Germany). The 5′-UTR sequence of *ZmPpc* (Accession No: X15239) was taken from Hudspeth and Grula (1989), and the sequence for *SiPpc* (Accession No: AF495586) was assumed to be the same length as the one determined for *S. viridis*.

Isolation of *C_4Ppc* 5′-flanking sequences of *Setaria viridis, Setaria italica, Panicum miliaceum, Urochloa maxima* and *Digitaria sanguinalis*

The 5′-flanking regions of *C_4Ppc* genes of *S. viridis*, *S. italica*, *P. miliaceum*, *U. maxima* and *D. sanguinalis* were isolated from total DNA by vectorette PCR (Siebert *et al.*, 1995) using the Universal Genome Walker Kit (Clontech Laboratories, Mountain View, USA) and following the manufacturer's protocol. For each plant species, DNA libraries with the restriction enzymes *Dra*I, *Eco*RV, *Pvu*II, *Stu*I, *Ehe*I, *Ksp*AI, *Sma*I *Sml*I and *Afe*I (Thermo Fisher Scientific, Waltham, USA) were prepared. For the primary and secondary PCR of the 1st walking step, the gene-specific oligonucleotides (GW1 and GW2) were designed that bind specifically to the 5′-end of the coding sequence of *C_4Ppc* gene of the respective species (Table S1). Together with the adapter primers AP1and AP2, the gene-specific primers were used for the 1st walking step. The sequence information obtained in this step was then used for a subsequent 2nd walking step, and the next gene-specific oligonucleotides (GW3 and GW4) were designed accordingly (Table S1). Subsequent walking steps were continued till the desirable length of the 5′-flanking sequenced was achieved for each of the *C_4Ppc* genes.

The resulting PCR products from the Genome walking were cloned into pJET1.2/Blunt vector using the CloneJET PCR cloning Kit (Thermo Fisher, Waltham, USA) following the manufacturer's protocol. Afterwards the cloned plasmids were transferred into chemically competent *Escherichia coli* DH5α cells (Invitrogen™, Carlsbad, USA), and the positive clones were identified via colony PCR using the pJET1.2 Forward Sequencing Primer and the pJET1.2 Reverse Sequencing Primer of the kit. Next, the plasmids from the positive colonies were extracted using the QIAprep® Spin Miniprep Kit (Qiagen, Hilden, Germany). Finally, the 5′-flanking sequences from the respective plasmids were determined by means of Sanger sequencing. In order to identify the CNSs among the 5′-flanking sequences, multiple sequence alignment was carried out using the online tool Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

Cloning of promoter-reporter gene constructs

All the chimeric promoter-reporter gene constructs used in this study were cloned into a slightly modified Gateway pMDC164 vector (Curtis and Grossniklaus, 2003). In the original pMDC164 vector the marker gene hygromycin phosphotransferase was expressed under control of the cauliflower mosaic virus 35S promoter. It was reported that the strong 35S promoter commonly used in a selectable marker gene of plant transformation vector could affect the expression pattern of the adjacent tissue-specific gene promoters (Zheng *et al.*, 2007). Thereby, in our pMDC164-modified vector, the 35S promoter of the hygromycin cassette has been replaced with a ubiquitin promoter of *Z. mays*. In this study, the full-length promoter and the deletion constructs were generated using the Gateway cloning technique (Hartley *et al.*, 2000). The gateway *att*B sites were added with PCR to the respective sequences. Oligonucleotide combinations that have been used for PCR amplification of the promoters' sequence are listed in Tables S2 and S3. Next the purified PCR products were cloned into the entry vector pDONR207 or pDONR221 following the instruction of the Gateway® BP-Clonase II enzyme mix manual (Invitrogen, Thermo Fisher Scientific, USA). The resulting entry clones were verified by sequencing and then subjected for LR reaction to the destination vector pMDC164-modified. LR reaction was performed according to the manual instruction of the Gateway® LR-Clonase II enzyme mix (Thermo Scientific, Waltham, USA). Positive clones from the LR reaction were verified by means of restriction enzyme test digestion. Internal promoter deletion and recombination constructs (listed in Table S3) were synthesized by Biomatik using the Custom Gene Synthesis service (Biomatik, Ontario, Canada) (http://www.biomatik.com/services/gene-synthesis-9.html). Gateway overhang *att*L sides were added during the synthesis of the constructs and the company cloned them in pUC57 as an entry vector. Next the synthesized constructs from the pUC57 entry clones were delivered to pMDC164-modified destination vector via LR reaction.

Transformation of *Oryza sativa*

Two different cultivars of *Oryza sativa* ssp. *japonica* rice, Kitaake and Nipponbare, were used for the transformation in this project.

Table 1 List of oligonucleotides used to determine the 5′-UTR sequences

Species	5′-RACE oligonucleotides	3′-RACE gene-specific oligonucleotides (5′ to 3′)	Kit used
Setaria viridis	UPM (Universal Primer A Mix)	*viridis* RACE1: TCAGCAGTGGCCGCTTGCCG	SMARTer™ RACE cDNA Amplification Kit
Digitaria sanguinalis	UPM	*sanguinalis* RACE1: CAGCTGGAGTGCACGTTCGCGTGTGT	SMARTer™ RACE cDNA Amplification Kit
Urochloa maxima	SMART-II-A 5′-Primer	*maxima* RACE1: AGTCGGAGTAGCCGACCATC	SMART cDNA Synthesis Kits
Panicum miliaceum	SMART-II-A 5′-Primer	*miliaceum* RACE1: GTAGGGGCCGAAGCTGTCGCGG	SMART cDNA Synthesis Kits

The chimeric promoter-reporter gene constructs that were transformed in the respective Kitaake and Nipponbare rice cultivars are listed in Table S4. The constructs of interest for the transformation in the Kitaake and Nipponabare rice were first transformed into the *Agrobacterium tumefaciens* strain AGL1 (Lazo *et al.*, 1991) and LBA4404 (Wise *et al.*, 2006), respectively, via electroporation and the freeze-thaw method. Next, callus induction, transformation, selection, regeneration and rooting of Kitaake rice were done according to the slightly modified protocol of Toki *et al.* (2006). The adopted protocol can be found at https://langdalelab.files.wordpress.com/2015/07/kitaake_transformation_2015.pdf. Hygromycin-resistant rice plantlets with well-developed roots from the *in vitro* culture were transferred to hydroponic culture in the greenhouse for 2 weeks according to the protocol described in Gregorio *et al.* (1997). Afterwards, the transgenic rice plants were grown to maturity in soil (J. Arthur Bower's, JOHN INNES No.2, Sinclair Pro, Cheshire, UK) at 30°C. In contrast, transformation of Nipponbare was performed according to the protocol described by Hiei and Komari (2006). Finally, the presence of the T-DNA in the transgenic T_0 and T_1 plants was confirmed by genotyping via PCR.

GUS histochemical and MUG quantitative assay

Mature leaves from positively PCR-tested transgenic plants were harvested for histochemical analysis and quantitative GUS fluorometric assay. Regenerated T_0 plants of the Kitaake rice cultivar and T_1 plants of the Nipponbare cultivar harboring the respective constructs were used for the analyses. Fully developed 2nd or 3rd leaves from the top of the plant with a length of at least 10 cm were harvested before the onset of flowering. For the histochemical analysis, thin cross-sections of mature leaves of the transgenic plants were prepared using a razor blade. Then the cross-sections were incubated in a reaction tube with 100–200 μl incubation buffer [100 mM Na-phosphate pH 7.0, 10 mM EDTA, 0.5 mM $K_4Fe(CN)_6$, 0.5 mM $K_3Fe(CN)_6$, 0.1% (v/v) Triton X-100, 1 mM X-Gluc (5-bromo-4-chlor-3-indoyl-β-D-glucuronid)] at 37°C for 2–72 h (Jefferson *et al.*, 1987). After removing the incubation buffer, the reaction was stopped by adding 100 μl fixation buffer [75% (v/v) acetic acid, 25% (v/v) ethanol]. The cross-sections were rinsed in 70% ethanol to remove the chlorophyll. Finally, the cross-sections were analyzed, and images taken with a Carl Zeiss Axiopot equipped with Axio Cam ICc5. Fluorometric quantification of the GUS activity (MUG assay) was performed according to the protocols of Jefferson *et al.* (1987) and Kosugi *et al.* (1990). Activity of GUS expression was measured with the help of a plate reader (Synergy/HTX-multi mode reader, BioTek, Vermont, USA). To determine the statistical difference between data sets, a Mann–Whitney U-test (Prism 6, Graph Pad Software, La Jolla, USA) was employed.

ACCESSION NUMBERS

Sequence data presented in this article have been submitted to GenBank data library under the following accession numbers MH675618 (*SvPpc* 5′-flanking sequences), MH675617 (*PmPpc* 5′-flanking sequences), MH675616 (*UmPpc* 5′-flanking sequences), and MH675615 (*DsPpc* 5′-flanking sequences).

ACKNOWLEDGEMENTS

This work was supported by grants from the Deutsche Forschungsgemeinschaft (IRTG1525) through the International Research Training Group IRTG1525 'The dynamic response of plants to a changing environment' (to PW) and the Cluster of Excellence in Plant Sciences EXC 1028 (to PW), and the Bill and Melinda Gates Foundation in the framework of the C_4 Rice Project (to WPQ and PW). The authors thank Jean Melgar and Justina De Palma-Davis for helping with the rice transformation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. 5′-Deletion analysis of the *ZmPpc* 5′-flanking sequences in transgenic rice plant.

Figure S2. 5′-RACE experiment to determine the 5′-UTR of the C_4*Ppc* cDNAs.

Figure S3. Functional analysis of the *UmPpc*-fGFP promoter in leaves of transgenic rice.

Figure S4. Functional analysis of the TATA and 5′-UTR segment of the *UmPpc* promoter.

Figure S5. Functional analysis of the proximal 500 bp of *DoPpc* 5′-flanking sequences in leaves of transgenic rice.

Table S1. List of oligonucleotides used for genome walking.

Table S2. List of oligonucleotides used to amplify promoter sequences.

Table S3. Promoter-reporter constructs, and oligonucleotide combinations used for amplification.

Table S4. List of the constructs transformed in the rice cultivars.

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 C4 dicot *Flaveria trinervia*. *Plant Cell*, **19**, 3391–3402.

Assaad, F.F., Tucker, K.L. and Signer, E.R. (1993) Epigenetic repeat-induced gene silencing (RIGS) in Arabidopsis. *Plant Mol. Biol.* **22**, 1067–1085.

Aubry, S., Kelly, S., Kumpers, B.M., Smith-Unna, R.D. and Hibberd, J.M. (2014) Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of C4 photosynthesis. *PLoS Genet.* **10**, e1004365.

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.

Christin, P., Besnard, G., Samaritani, E., Duvall, M., Hodkinson, T., Savolainen, V. and Salamin, N. (2008) Oligocene CO_2 decline promoted C4 photosynthesis in grasses. *Curr. Biol.* **18**, 37–43.

Christin, P.A., Salamin, N., Kellogg, E.A., Vicentini, A. and Besnard, G. (2009) Integrating phylogeny into studies of C4 variation in the grasses. *Plant Physiol.* **149**, 82–87.

Curtis, M.D. and Grossniklaus, U. (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* **133**, 462–469.

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 C4 species Flaveria trinervia: analysis of transcriptional control in transgenic Flaveria bidentis (C4) and Arabidopsis (C3). *Plant Physiol.* **146**, 1773–1785.

Engstrom, P.G., Ho Sui, S.J., Drivenes, O., Becker, T.S. and Lenhard, B. (2007) Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res.* **17**, 1898–1908.

Gorska, A.M., Gouveia, P., Borba, A.R., Zimmermann, A., Serra, T.S., Lourenco, T.F., Margarida Oliveira, M., Peterhansel, C. and Saibo, N.J.M. (2019) ZmbHLH80 and ZmbHLH90 transcription factors act

antagonistically and contribute to regulate PEPC1 cell-specific gene expression in maize. *Plant J.* **99**, 270–285.

Gowik, U. and Westhoff, P. (2011) The path from C3 to C4 photosynthesis. *Plant Physiol.* **155**, 56–63.

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 C4 plant Flaveria trinervia, the promoter of the C4 phosphoenolpyruvate carboxylase gene. *Plant Cell*, **16**, 1077–1090.

Gregorio, G., Senadhira, D. and Mendoza, R. (1997) Screening rice for salinity tolerance, Vol 22, IRRI discussion paper series.

Hartley, J.L., Temple, G.F. and Brasch, M.A. (2000) DNA cloning using *in vitro* site-specific recombination. *Genome Res.* **10**, 1788–1795.

Hatch, M.D. (1987) C4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochem. Biophys. Acta*, **895**, 81–106.

Hatton, D., Sablowski, R., Yung, M.H., Smith, C., Schuch, W. and Bevan, M. (1995) Two classes of cis sequences contribute to tissue-specific expression of a PAL2 promoter in transgenic tobacco. *Plant J.* **7**, 859–876.

Heyndrickx, K.S., Van de Velde, J., Wang, C., Weigel, D. and Vandepoele, K. (2014) A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. *Plant Cell*, **26**, 3894–3910.

Hibberd, J.M. and Covshoff, S. (2010) The regulation of gene expression required for C4 photosynthesis. *Annu. Rev. Plant Biol.* **61**, 181–207.

Hiei, Y. and Komari, T. (2006) Improved protocols for transformation of Indica rice mediated by Agrobacterium tumefaciens. *Plant Cell, Tissue Organ Cult.* **85**, 271.

Hsieh, J. and Fire, A. (2000) Recognition and silencing of repeated DNA. *Annu. Rev. Genet.* **34**, 187–204.

Hudspeth, R.L. and Grula, J.W. (1989) Structure and expression of the maize gene encoding the phosphoenolpyruvate carboxylase isozyme involved in C4 photosynthesis. *Plant Mol. Biol.* **12**, 579–589.

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.

Kausch, A.P., Page Owen, T., Zachwieja, S.J., Flynn, A.R. and Sheen, J. (2001) Mesophyll-specific, light and metabolic regulation of the C4 PPCZm1 promoter in transgenic maize. *Plant Mol. Biol.* **45**, 1–15.

Kirschner, S., Woodfield, H., Prusko, K., Koczor, M., Gowik, U., Hibberd, J.M. and Westhoff, P. (2018) Expression of SULTR2;2, encoding a low-affinity sulphur transporter, in the Arabidopsis bundle sheath and vein cells is mediated by a positive regulator. *J. Exp. Bot.* **69**, 4897–4906.

Kosugi, S., Ohashi, Y., Nakajima, K. and Arai, Y. (1990) An improved assay for β-glucuronidase in transformed cells: methanol almost completely suppresses a putative endogenous β-glucuronidase activity. *Plant Sci.* **70**, 133–140.

Ku, M., Kano-Murakami, Y. and Matsuoka, M. (1996) Evolution and expression of C4 photosynthesis genes. *Plant Physiol.* **111**, 949–957.

Lazo, G.R., Stein, P.A. and Ludwig, R.A. (1991) A DNA transformation-competent Arabidopsis genomic library in Agrobacterium. *Biotechnology, (NY)* **9**, 963–967.

Lenhard, B., Sandelin, A. and Carninci, P. (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* **13**, 233–245.

Leyva, A., Liang, X., Pintor-Toro, J.A., Dixon, R.A. and Lamb, C.J. (1992) cis-element combinations determine phenylalanine ammonia-lyase gene tissue-specific expression patterns. *Plant Cell*, **4**, 263–271.

Matsuoka, M., Tada, Y., Fujimura, T. and Kano-Murakami, Y. (1993) Tissue-specific light-regulated expression directed by the promoter of a C4 gene, maize pyruvate, orthophosphate dikinase, in a C3 plant, rice. *Proc. Natl Acad. Sci. USA*, **90**, 9586–9590.

Matsuoka, M., Kyozuka, J., Shimamoto, K. and Kano-Murakami, Y. (1994) The promoters of two carboxylases in a C4 plant (maize) direct cell-specific, light-regulated expression in a C3 plant (rice). *Plant J.* **6**, 311–319.

Mejia-Guerra, M.K., Li, W., Galeano, N.F., Vidal, M., Gray, J., Doseff, A.I. and Grotewold, E. (2015) Core promoter plasticity between maize tissues and genotypes contrasts with predominance of sharp transcription initiation sites. *Plant Cell*, **27**, 3309–3320.

Monson, R. (2003) Gene duplication, neofunctionalization, and the evolution of C4 photosynthesis. *Int. J. Plant Sci.* **164**, S43–S54.

Nomura, M., Higuchi, T., Ishida, Y., Ohta, S., Komari, T., Imaizumi, N., Miyao-Tokutomi, M., Matsuoka, M. and Tajima, S. (2005) Differential expression pattern of C4 bundle sheath expression genes in rice, a C3 plant. *Plant Cell Physiol.* **46**, 754–761.

Ow, D.W., Jacobs, J.D. and Howell, S.H. (1987) Functional regions of the cauliflower mosaic virus 35S RNA promoter determined by use of the firefly luciferase gene as a reporter of promoter activity. *Proc. Natl Acad. Sci. USA*, **84**, 4870–4874.

Sage, R. (2004) The evolution of C4 photosynthesis. *New Phytol.* **161**, 341–370.

Sage, R.F. (2016) A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *J. Exp. Bot.* **67**, 4039–4056.

Sage, R.F., Christin, P.A. and Edwards, E.J. (2011) The C4 plant lineages of planet Earth. *J. Exp. Bot.* **62**, 3155–3169.

Sage, R.F., Sage, T.L. and Kocacinar, F. (2012) Photorespiration and the evolution of C4 photosynthesis. *Annu. Rev. Plant Biol.* **63**, 19–47.

Schaffner, A.R. and Sheen, J. (1992) Maize C4 photosynthesis involves differential regulation of phosphoenolpyruvate carboxylase genes. *Plant J.* **2**, 221–232.

Schuler, M.L., Mantegazza, O. and Weber, A.P. (2016) Engineering C4 photosynthesis into C3 chassis in the synthetic biology age. *Plant J.* **87**, 51–65.

Sheehy, J.E., Mitchell, P.L. and Hardy, B. (2008) *Charting New Pathways to C_4 Rice.* (First Edition), World Scientific Publishing Company, Singapore.

Sheen, J. (1999) C4 gene expression. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 187–217.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A. and Lukyanov, S.A. (1995) An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.* **23**, 1087–1088.

Stockhaus, J., Schlue, U., Koczor, M., Chitty, J.A., Taylor, W.C. and Westhoff, P. (1997) The promoter of the gene encoding the C4 form of phosphoenolpyruvate carboxylase directs mesophyll-specific expression in transgenic C4 Flaveria spp. *Plant Cell*, **9**, 479–489.

Studer, A.J., Schnable, J.C., Weissmann, S., Kolbe, A.R., McKain, M.R., Shao, Y., Cousins, A.B., Kellogg, E.A. and Brutnell, T.P. (2016) The draft genome of the C3 panicoid grass species Dichanthelium oligosanthes. *Genome Biol.* **17**, 223.

Svensson, P., Blasing, O. and Westhoff, P. (2003) Evolution of C_4 phosphoenolpyruvate carboxylase. *Arch. Biochem. Biophys.* **414**(2), 180–188.

Taniguchi, M., Izawa, K., Ku, M.S.B., Lin, J.-H., Saito, H., Ishida, Y., Ohta, S., Komari, T., Matsuoka, M. and Sugiyama, T. (2000) Binding of cell type-specific nuclear proteins to the 5′-flanking region of maize C4 phosphoenolpyruvate carboxylase gene confers its differential transcription in mesophyll cells. *Plant Mol. Biol.* **44**, 543–557.

Wei, H., Meilan, R., Brunner, A.M., Skinner, J.S., Ma, C. and Strauss, S.H. (2006) Transgenic sterility in Populus: expression properties of the poplar PTLF, Agrobacterium NOS and two minimal 35S promoters in vegetative tissues. *Tree Physiol.* **26**, 401–410.

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 C4 Flaveria: an intricate interplay of transcriptional and posttranscriptional processes. *Plant Cell*, **24**, 137–151.

Wise, A.A., Liu, Z. and Binns, A.N. (2006) Three methods for the introduction of foreign DNA into Agrobacterium. *Methods Mol. Biol.* **343**, 43–53.

Yang, F., Li, W., Jiang, N. *et al.* (2017) A maize gene regulatory network for phenolic metabolism. *Mol. Plant*, **10**, 498–515.

Zheng, X., Deng, W., Luo, K., Duan, H., Chen, Y., McAvoy, R., Song, S., Pei, Y. and Li, Y. (2007) The cauliflower mosaic virus (CaMV) 35S promoter sequence alters the level and patterns of activity of adjacent tissue- and organ-specific gene promoters. *Plant Cell Rep.* **26**, 1195–1203.