Genetics and Molecular Biology, 46, 3 (suppl 1), e20230190 (2023)

DOI: https://doi.org/10.1590/1678-4685-GMB-2023-0190

Review
60 years of the PPGBM UFRGS – Special Issue

C_4 Phosphoenolpyruvate Carboxylase: Evolution and transcriptional regulation

Pedro Carvalho[1*], Célia Gomes[1*] and Nelson J.M. Saibo[1]

[1]*Universidade Nova de Lisboa, Instituto de Tecnologia Química e Biológica António Xavier, Oeiras, Portugal.*

Abstract

Photosynthetic phosphoenolpyruvate carboxylase (PEPC) catalyses the irreversible carboxylation of phosphoenolpyruvate (PEP), producing oxaloacetate (OAA). This enzyme catalyses the first step of carbon fixation in C_4 photosynthesis, contributing to the high photosynthetic efficiency of C_4 plants. PEPC is also involved in replenishing tricarboxylic acid cycle intermediates, such as OAA, being involved in the C/N balance. In plants, PEPCs are classified in two types: bacterial type (BTPC) and plant-type (PTPC), which includes photosynthetic and non-photosynthetic PEPCs. During C_4 evolution, photosynthetic PEPCs evolved independently. C_4 PEPCs evolved to be highly expressed and active in a spatial-specific manner. Their gene expression pattern is also regulated by developmental cues, light, circadian clock as well as adverse environmental conditions. However, the gene regulatory networks controlling C_4 *PEPC* gene expression, namely its cell-specificity, are largely unknown. Therefore, after an introduction to the evolution of PEPCs, this review aims to discuss the current knowledge regarding the transcriptional regulation of C_4 PEPCs, focusing on cell-specific and developmental expression dynamics, light and circadian regulation, as well as response to abiotic stress. In conclusion, this review aims to highlight the evolution, transcriptional regulation by different signals and importance of PEPC in C_4 photosynthesis and its potential as tool for crop improvement.

Keywords: C_4 photosynthesis, transcriptional regulation, PEPC, C_3 to C_4 evolution.

Received: July 01, 2023; Accepted: February 06, 2024.

Phosphoenolpyruvate carboxylase in plants and its rise to power

Phosphoenolpyruvate carboxylase (PEPC; EC 4.1.1.31) is a ubiquitous and cytosolic enzyme, responsible for the irreversible β-carboxylation of phosphoenolpyruvate (PEP), in the presence of HCO_3^-, producing oxaloacetate and inorganic phosphate (Pi) (O'Leary, 1982; Chollet *et al.*, 1996; O'Leary *et al.*, 2011). It can be found in non-photosynthetic bacteria, cyanobacteria, green algae, and in all land plants (O'Leary *et al.*, 2011).

In most organisms, PEPC plays an anaplerotic role being important to replenish intermediates, namely oxaloacetate, in the tricarboxylic acid cycle, by re-fixing the CO_2 released by respiration, thus allowing an increased flux throughout this cycle (Sánchez and Cejudo, 2003). In plants, it occupies a central place in the primary carbon metabolism, linking the carbon and nitrogen metabolism (Figure 1) (O'Leary *et al.*, 2011). In *Arabidopsis thaliana*, plants lacking PEPC1 and PEPC2 show growth arrest in control conditions, which is linked to a disrupted carbon-nitrogen balance. Double mutants not only show reduction of NH_4^+ fixation, by repression of the GOGAT/GS cycles, but also an accumulation of sucrose and starch granules in the chloroplasts, having impaired starch degradation (Shi *et al.*, 2015).

Send correspondence to Nelson J.M. Saibo. Universidade NOVA de Lisboa, Instituto de Tecnologia Química e Biológica António Xavier, Avenida da República, 2780-157 Oeiras, Portugal. E-mail: saibo@itqb.unl.pt

*These authors contributed equally to the article.

In C_4 and CAM plants, one of the PEPCs playing an anaplerotic role evolved to have a role in photosynthesis. For these plants, the irreversible carboxylation performed by PEPC is the first step of carbon assimilation, being therefore a key enzyme for the proper operation of C_4 and CAM photosynthesis. Since it is possible to distinguish between their anaplerotic and photosynthetic roles, plant PEPC isoforms are divided into photosynthetic and non-photosynthetic (O'Leary *et al.*, 2011).

PEPCs in plant genomes

In plants, the different PEPC enzyme isoforms are encoded by a small multigene family. Within this family, two major lineages can be distinguished: bacterial-type (BTPC) and plant-type (PTPC) PEPCs (O'Leary *et al.*, 2011). At least one copy of the *BTPC* gene can be found in most plant species sequenced to date (Figure 2). BTPCs found in both dicots and monocots are phylogenetically closer to PEPCs from bacteria than to PTPCs (O'Leary *et al.*, 2011). In addition to its different gene structure, BTPCs and bacterial PEPCs lack a N-terminal Serine residue, which can be phosphorylated, an important feature that distinguishes them from PTPCs (Sánchez and Cejudo, 2003). It has been proposed that when Viridiplantae (green plants) arose, PTPC originated from BTPC through gene duplication (Chang *et al.*, 2013).

Plant-type PEPCs typically can be found as homotetramers and traditionally they are divided as photosynthetic, for those involved in C_4 or CAM photosynthesis, or non-photosynthetic, PTPCs not involved in photosynthesis in either C_3 or C_4 species. Although diverse, all plant PEPCs are thought to have appeared from a single ancestral form (Svensson *et al.*, 2003).

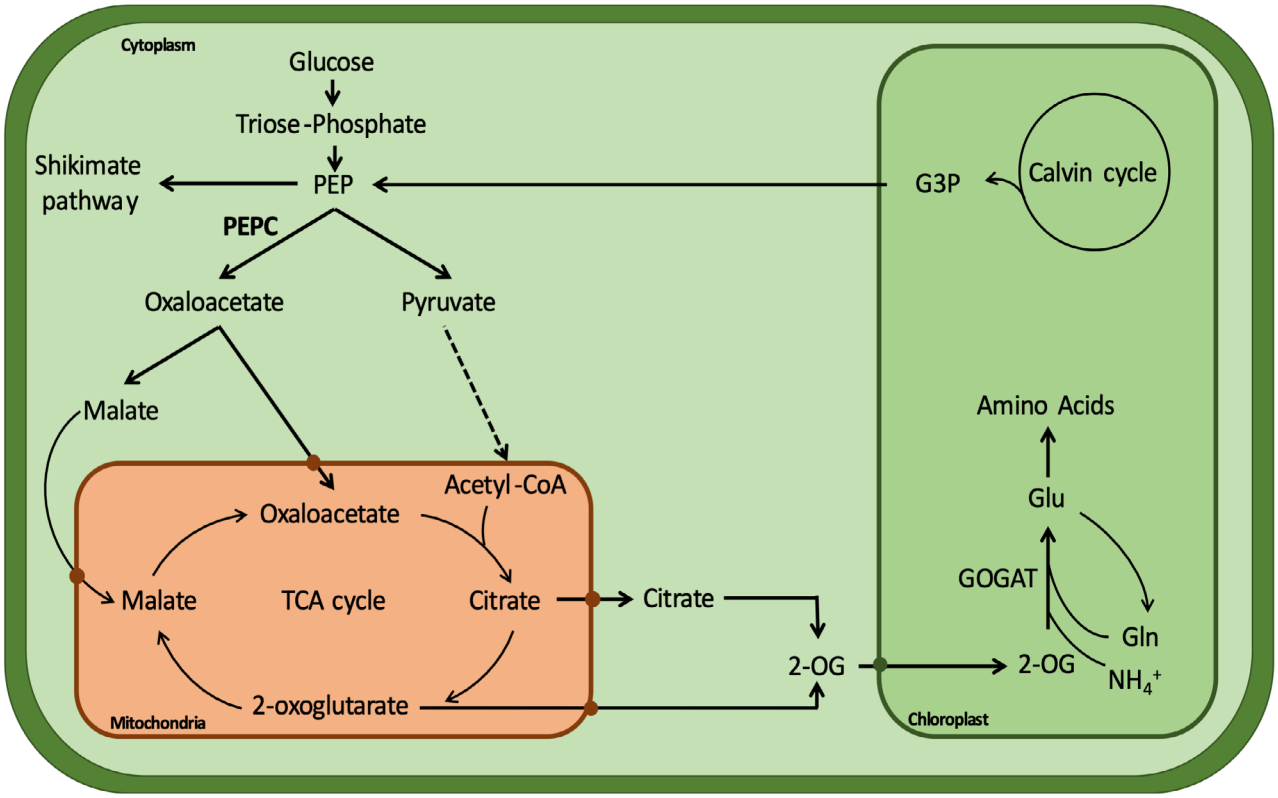


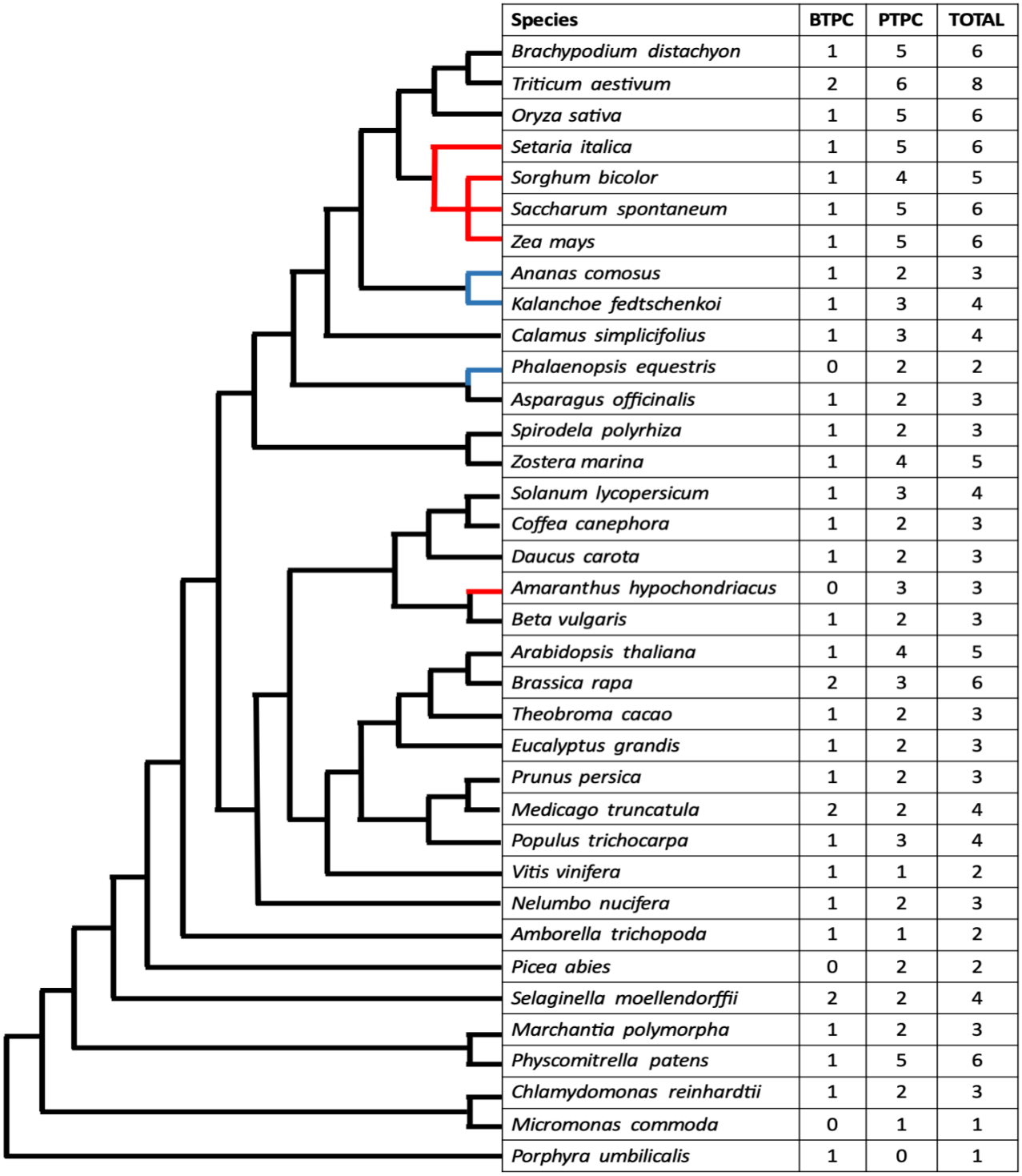
Figure 1 – Simplified schematic representation of the role played by non-photosynthetic PEPC in the carbon-nitrogen balance. The carboxylation of phosphoenolpyruvate (PEP) is an important step to replenish carbon skeletons to the TCA cycle, re-routing carbon (glycolysis products) into the TCA cycle. The link between the TCA and GOGAT/GS cycles is important for the carbon-nitrogen balance, making PEPC an important regulator of carbon partitioning.

The path to C_4 photosynthesis

To overcome the energy loss due to photorespiration, a process that metabolises a toxic compound generated when Rubisco acts as oxygenase, some plants have evolved a carbon concentration mechanism called C_4 photosynthesis. In most C_4 plants, CO_2 is first fixed in the mesophyll cells by PEPC, into a four-acid compound that is shuttled to the bundle sheath cells where it is decarboxylated, thus increasing the CO_2 concentration around Rubisco. In addition to the two-cell type C_4 photosynthesis, a few plants have developed C_4 photosynthesis in a single-cell, where the spatial separation of the carbon fixation reactions occurs inside one cell. For instance, in the single-cell C_4 species *Bienertia sinuspersici*, C_4 photosynthesis is based on an intracellular compartmentation including two physiologically and biochemically different chloroplast types (Caburatan *et al.*, 2019). Evolution of C_4 photosynthesis has occurred over 60 independent times, in both dicotyledons and monocotyledons, in one of the most amazing examples of convergent evolution known in nature (Sage *et al.*, 2011). Despite the broad evolutionary trajectories of C_4 photosynthesis, all C_4 species rely on PEPC for the first carboxylation step (Sage *et al.*, 2011). Many authors have tried to resolve the evolutionary origin of PEPCs and they have clearly shown that photosynthetic C_4 PEPCs from dicots and monocots evolved from different C_3 origins (Westhoff and Gowik, 2004; Christin *et al.*, 2007; Besnard *et al.*, 2009; Christin and Besnard, 2009).

In the dicot *Flaveria* genus, which contains C_3, C_4 and C_3-C_4 intermediate species, it is possible to distinguish 3 classes of PEPC genes (A, B, and C) (Westhoff and Gowik, 2004). PEPCs from class A are present in both C_3 and C_4 species and class A C_4 PEPCs originated from a duplication of class B PEPCs. The photosynthetic PEPCs belong to class A and originated from a duplication of class B PEPCs. Class A C_4 PEPCs (ppcA) are present in both C_3 and C_4 species, however, although these genes show variable transcript levels among species, in C_4-like intermediate species, ppcA transcript levels are higher and similar to C_4 plants (Engelmann *et al.*, 2003). Therefore, C_4 PEPC isoforms seem to have evolved in a stepwise fashion, with the increase of gene expression preceding amino acid changes (Westhoff and Gowik, 2004; Engelmann *et al.*, 2003).

In the clade PACMAD (named based on its subfamilies Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae, Danthonioideae), which comprises all the grass C_4 species, PEPCs have evolved over eight independent times, recruiting different C_3 PEPC isoforms to acquire the C_4 function (Christin *et al.*, 2007; Christin and Besnard, 2009). In most grass species, the recruited isoform was *ppc-B2*, while in the case of *Stipagrostis* genus, it was *ppc-A1b* isoform (Christin and Besnard, 2009). In the case of sedges (Cyperaceae), the PEPC isoform recruited for C_4 photosynthesis is sister of the *ppc-A1a* and *ppc-A1b* isoforms from grasses, evolving five independent times (Besnard *et al.*, 2009; Christin and Besnard, 2009).



Species	BTPC	PTPC	TOTAL
Brachypodium distachyon	1	5	6
Triticum aestivum	2	6	8
Oryza sativa	1	5	6
Setaria italica	1	5	6
Sorghum bicolor	1	4	5
Saccharum spontaneum	1	5	6
Zea mays	1	5	6
Ananas comosus	1	2	3
Kalanchoe fedtschenkoi	1	3	4
Calamus simplicifolius	1	3	4
Phalaenopsis equestris	0	2	2
Asparagus officinalis	1	2	3
Spirodela polyrhiza	1	2	3
Zostera marina	1	4	5
Solanum lycopersicum	1	3	4
Coffea canephora	1	2	3
Daucus carota	1	2	3
Amaranthus hypochondriacus	0	3	3
Beta vulgaris	1	2	3
Arabidopsis thaliana	1	4	5
Brassica rapa	2	3	6
Theobroma cacao	1	2	3
Eucalyptus grandis	1	2	3
Prunus persica	1	2	3
Medicago truncatula	2	2	4
Populus trichocarpa	1	3	4
Vitis vinifera	1	1	2
Nelumbo nucifera	1	2	3
Amborella trichopoda	1	1	2
Picea abies	0	2	2
Selaginella moellendorffii	2	2	4
Marchantia polymorpha	1	2	3
Physcomitrella patens	1	5	6
Chlamydomonas reinhardtii	1	2	3
Micromonas commoda	0	1	1
Porphyra umbilicalis	1	0	1

Figure 2 – Cladogram representing the amount of PEPC isoform present in plant genomes. Species are organised considering their phylogenetic relationships, with representatives of important evolutionary groups. Sequences were obtained from PLAZA and NCBI databases, using different PEPC protein sequences for BLASTp. Incomplete or unrelated sequences were removed by protein alignment and phylogenetic analysis. Red lines represent C_4 species, blue lines represent CAM species, and black lines represent C_3 species.

It is yet to be defined which amino acid changes are responsible for the evolution from a C_3 to a C_4 isoform. Despite some amino acid positions having been proposed as being under positive selection for C_4 function (Christin *et al.*, 2007), only one amino acid substitution has been conclusively linked to the C_4 isoform of PEPC (Bläsing *et al.*, 2000). The substitution of an Alanine to a Serine can be found in C_4 PEPCs of several dicots and monocots, making it a key criterion for C_4 isoform definition. It occurs in position 780 in maize (Christin *et al.*, 2007), and 774 in *Flaveria* species and significantly influences PEPC kinetic properties (Bläsing *et al.*, 2000). Besides the specific protein features, PEPC

transcriptional regulation in C_4 plants is tightly controlled and its essential for the proper functioning of C_4 metabolism.

Transcriptional regulation of C_4 PEPC

Developmental regulation

In monocots and dicots, leaves differentiate following a gradient, in which younger cells are present at the leaf base, while older and more mature cells are present at the leaf tip (Nelson and Langdale, 1989; Stockhaus *et al.*, 1997; Aubry *et al.*, 2014). During leaf development, C_4 *PEPC* gene expression is regulated by developmental cues, increasing gradually from leaf base to leaf tip (Martineau and Taylor, 1985; Stockhaus *et al.*, 1997; Pick *et al.*, 2011; Aubry *et al.*, 2014; Tao *et al.*, 2022). In maize and *Cleome gynandra*, C_4 *PEPC* transcript level is higher in mature than in younger leaves (Kausch *et al.*, 2001; Aubry *et al.*, 2014). Since mature leaves have more differentiated M cells than younger leaves, it seems that C_4 *PEPC* expression level follows M cells differentiation. In fact, maize *PEPC* was recently identified as a target of COL8, a transcription factor (TF) co-regulated with *PEPC* during M cell development (Tao *et al.*, 2022). This suggests that COL8 might regulate *PEPC* expression during leaf development, however further investigation is required to validate this TF as a *PEPC* gene regulator. A developmental regulation of C_4 *PEPC* gene expression was also observed in the single-cell type C_4 species *Bienertia sinuspersici.* In this species, gene expression analysis of PEPC isoforms showed that C_3 *PEPC* is more expressed in the younger leaves or early stages of development, while C_4 *PEPC* is upregulated in the mature stages of leaf development (Caburatan *et al.*, 2019). However, C_4 *PEPC* gene expression does not follow a developmental pattern in all species. In the particular case of Amaranth, C_4 *PEPC* is highly expressed since the beginning of leaf development, namely in leaf primordia and in the apical meristem and surrounding regions (Ramsperger *et al.*, 1996).

C_4 PEPC protein accumulates at different leaf development stages in a species-dependent manner (Mayfield and Taylor, 1984; Martineau and Taylor, 1985; Dengler *et al.*, 1995; Soros and Dengler, 2001; Voznesenskaya *et al.*, 2003; Wakayama *et al.*, 2003; Majeran *et al.*, 2010; Koteyeva *et al.*, 2014) and, in general, C_4 PEPC accumulation goes along with M cells differentiation (Voznesenskaya *et al.*, 2003; Wakayama *et al.*, 2003; Majeran *et al.*, 2010; Koteyeva *et al.*, 2014). Nevertheless, the mechanisms coordinating C_4 *PEPC* gene expression and protein accumulation during leaf development differ among species (Langdale *et al.*, 1988; Wang *et al.*, 1992; Wang *et al.*, 1993; Dengler *et al.*, 1995; Ramsperger *et al.*, 1996; Soros and Dengler, 2001; Voznesenskaya *et al.*, 2003; Wakayama *et al.*, 2003; Koteyeva *et al.*, 2014). In the case of amaranth, in early developmental stages, C_4 *PEPC* gene expression does not occur in a cell-specific way, however, the expressed protein is only present in the M cell precursors (Ramsperger *et al.*, 1996). This pattern is also observed in cotyledons and maintained in later stages of leaf development, namely during leaf unfolding (Wang *et al.*, 1992; Wang *et al.*, 1993). Although no information is available regarding the regulatory mechanisms underlying C_4 *PEPC* gene expression in amaranth during leaf development, post-transcriptional or translational regulation mechanisms seem to play the main role in regulating cell-specific C_4 PEPC protein accumulation (Wang *et al.*, 1992; Wang *et al.*, 1993; Ramsperger *et al.*, 1996). In contrast, maize C_4 *PEPC* is expressed in a cell-specific way throughout leaf development (Langdale *et al.*, 1988; Majeran *et al.*, 2010). Hence, transcriptional regulatory mechanisms seem to be the most important to establish a C_4 *PEPC* cell-specific expression pattern in maize. Other species known to accumulate C_4 PEPC only in M cells, regardless of developmental stage, are *Atriplex rosea*, *Arundinella hirta* and two *Cleome* species (Dengler *et al.*, 1995; Wakayama *et al.*, 2003; Koteyeva *et al.*, 2014), however, the regulatory mechanisms underlying this feature are not known. A different example is *Salsola richteri*, in which C_4 PEPC protein starts to accumulate in a non-cell specific way at early stages, being present in BS and M cells, and other leaf cells albeit at lower levels, but, in later stages of leaf development, C_4 PEPC is detected exclusively in M cells (Voznesenskaya *et al.*, 2003). The mechanisms regulating *S. richteri* C_4 PEPC cell-specific accumulation are also unknown. Similarly to *Salsola richteri*, in two Cyperaceae species, *Pycreus polystachyos* and *Eleocharis retrofiexa*, C_4 PEPC accumulation only becomes cell-specific later in leaf development (Soros and Dengler, 2001). In *Eleocharis retrofiexa* C_4 PEPC accumulation is also present in the parenchymatous BS (PBS), suggesting that PBS and M cells have similar functions (Soros and Dengler, 2001). In the particular case of *Rhynchospora rubra*, another Cyperaceae species, C_4 PEPC never accumulates in a cell-specific way throughout leaf development, suggesting that *Rhynchospora rubra* may have a different version of C_4 photosynthesis (Soros and Dengler, 2001). Although these three species belong to the same family, the differences regarding C_4 PEPC accumulation may be related to the different C_4 origins they represent and to the differences in the anatomical features between species (Soros and Dengler, 2001).

The fact that C_4 *PEPC* gene expression and protein accumulation patterns during leaf development differ among species shows that different species acquired different developmental regulatory mechanisms during C_4 evolution, which is not surprising given the evolutionary convergence of C_4 photosynthesis. To better understand these regulatory mechanisms, more information regarding C_4 *PEPC* transcriptional regulation in different species during their leaf development is needed.

Spatial regulation

In most C_4 plants, photosynthetic reactions are divided into two different cell types, M and BS cells. As stated in section 1b, C_4 PEPC first fixes CO_2 in M cells, where it is highly and specifically expressed (Sage, 2004). This expression pattern required the development of a complex regulatory network during C_4 evolution. It has been suggested that the transcriptional mechanisms regulating non-photosynthetic *PEPC* gene expression were modified to reach a high and cell-specific transcript level (Williams *et al.*, 2012). The recruitment of *cis*-elements and TFs regulating C_3 genes was essential to achieve this purpose (Williams *et al.*, 2012).

In maize, the C_4 *PEPC* promoter (C_4 *ZmPEPC* promoter) drives a leaf-specific expression. Despite some gene expression in some leaf-like organs, the C_4 *ZmPEPC* promoter shows a very high activity in leaves as compared with other mature tissues, such as roots and stems, in which no activity is detected (Kausch *et al.*, 2001). Dof1 and Dof2 are two TFs identified as putative regulators of C_4 *PEPC* organ-specific gene expression in maize (Yanagisawa and Sheen, 1998) (Figure 3). Dof1 is a ubiquitously expressed TF, working as a light-dependent activator, while Dof2 is only expressed in roots and stems and acts as a repressor (Yanagisawa and Sheen, 1998). *In vivo* experiments demonstrated that when Dof2 is expressed, it binds to the C_4 *PEPC* promoter, impairing Dof1 binding and consequently promoter activation (Yanagisawa and Sheen, 1998). Therefore, it was hypothesised that, in stems and roots, Dof2 binds to the C_4 *PEPC* promoter, blocking Dof1 DNA interaction and, consequently, down-regulating C_4 *PEPC* transcript levels in these tissues (Figure 3A). In leaves, Dof1 is free to bind to the C_4 *PEPC* promoter, thus activating it (Figures 3B and 3C) (Yanagisawa and Sheen, 1998). However, contrasting with this hypothesis, the knockout of *Dof1* does not affect C_4 *PEPC* expression levels, implying that this TF does not have a prominent role in C_4 *PEPC* transcriptional regulation (Cavalar *et al.*, 2007). Another possibility is the existence of transcriptional redundancy by other Dof TFs or even TFs from other families. If this is true, the knockout of *Dof1* may not be sufficient to affect C_4 *PEPC* expression levels. Hence, the identification of other TFs regulating C_4 *PEPC* gene expression will be useful to understand how TFs regulate C_4 *PEPC* expression in a tissue-specific way.

Recently, three additional maize TFs, ZmbHLH80, ZmbHLH90, and ZmOrphan94 have been identified as putative regulators of C_4 *PEPC* cell-specific gene expression, having binding sites in the promoter regions known to be crucial to establish this expression pattern (Górska *et al.*, 2019; Gupta *et al.*, 2020; Górska *et al.*, 2021) (Figures 3A and 3B). ZmbHLH90 was shown to act as an activator of C_4 *ZmPEPC,* while ZmbHLH80 and ZmOrphan94 act as repressors (Górska *et al.*, 2019; Górska *et al.*, 2021). It was proposed that both repressors, ZmbHLH80 and ZmOrphan94, play an important role in C_4 *PEPC* cell-specific gene expression keeping its expression low in the BS cells, where they are preferentially expressed. The high *ZmbHLH80* and *ZmOrphan94* gene expression in the BS cells may lead to the formation of heterodimers with the activator ZmbHLH90, thus impairing its function (Górska *et al.*, 2019; Górska *et al.*, 2021) (Figure 3B). In M cells, ZmbHLH80 and ZmOrphan94 are less expressed and, therefore, ZmbHLH90 is free to form homodimers and thus activate C_4 *ZmPEPC* expression (Górska *et al.*, 2019; Górska *et al.*, 2021). We must however emphasise that, though it was clearly shown that ZmbHLH80 and ZmOrphan94 transcript levels are higher in BS as compared with M cells, nothing is known about their protein abundance. In addition to the negative regulation by heterodimerization, we may have other regulation mechanisms between activators and repressors, such as competition for the same binding site, interaction after DNA binding or a stronger regulatory effect of repressors over activators (Górska *et al.*, 2021) (Figure 3). It would be interesting to investigate whether these new identified TFs interact with the TFs previously identified and, if they interact, how they function to regulate C_4 *PEPC* gene expression. One could also hypothesise that a double mutant Dof1/ZmbHLH90 might be needed to affect C_4 *ZmPEPC* gene expression.

In addition to TFs, *cis*-elements in the C_4 *PEPC* promoter have also been associated with the mesophyll cell-specific gene expression (Gowik *et al.*, 2004; Akyildiz *et al.*, 2007; Gupta *et al.*, 2020). Interestingly, it has been reported that C_4 *PEPC* promoter regions underpinning cell-specific expression are different between dicots and monocots (Gowik *et al.*, 2004; Akyildiz *et al.*, 2007; Engelmann *et al.*, 2008; Gupta *et al.*, 2020). In dicots, such as *Flaveria* species, a region of the distal promoter (2141 to 1566 bps before ATG) of C_4 *PEPC* is responsible to establish the spatial expression pattern, while the proximal promoter region (570 bps before ATG) works as an enhancer of C_4 *PEPC* expression, being both necessary for high and cell-specific expression levels (Gowik *et al.*, 2004; Akyildiz *et al.*, 2007; Engelmann *et al.*, 2008). When the C_4 *PEPC* proximal promoter region was isolated, no cell-specificity was observed. On the other hand, when the proximal promoter region was replaced by its C_3 counterpart, although cell-specificity was maintained a decrease in promoter strength was observed (Gowik *et al.*, 2004; Akyildiz *et al.*, 2007; Engelmann *et al.*, 2008). Although some *cis*-elements have been identified as putative enhancers within the proximal promoter, their role in C_4 *PEPC* expression was never proven (Engelmann *et al.*, 2008). Deletions in the distal promoter, however, showed that a *cis*-element designated mesophyll expression module 1 (MEM1) is essential for a cell-specific expression. Without this element, or when it is replaced by its C_3 counterpart, the M cell specificity is lost (Gowik *et al.*, 2004; Akyildiz *et al.*, 2007). In contrast to *Flaveria* species, the C_4 *PEPC* proximal promoter (~500 bps) from grasses (monocots) is sufficient to drive a high M cell-specific expression, thus having all the necessary *cis*-elements to achieve cell-specificity (Schaffner and Sheen, 1992; Taniguchi *et al.*, 2000; Gupta *et al.*, 2020). Within this region, four conserved nucleotide sequences (CNSs) were identified as essential *cis*-elements for an M cell-specific expression (Gupta *et al.*, 2020). When the CNSs were eliminated from the C_4 *PEPC* promoter, the promoter activity was almost eliminated, being rescued when the original CNSs were replaced by equivalent sequences from a different C_4 grass species (Gupta *et al.*, 2020).

In addition to the *cis* and *trans* factors, some epigenetic modifications might be involved in C_4 *PEPC* gene expression regulation. Tri-methylation (H3K4me3) and di-methylation (H3K4me2) states, found in C_4 *PEPC* proximal promoter and transcribed regions, seem to be associated with the establishment of C_4 *PEPC* cell-specific expression (Danker *et al.*, 2008; Heimann *et al.*, 2013). These epigenetic modifications seem to have antagonistic effects as an enrichment of H3K4me3 in M cells and of H3K4me2 in BS cells is observed in several grass species (Danker *et al.*, 2008; Heimann *et al.*, 2013). Based on this evidence, it was proposed that a methyltransferase is recruited in a cell-specific way to convert low histone methylation states, such as HeK4me2, established by default in C_4 *PEPC*, in HeK4me3 enabling promoter activation (Danker *et al.*, 2008).

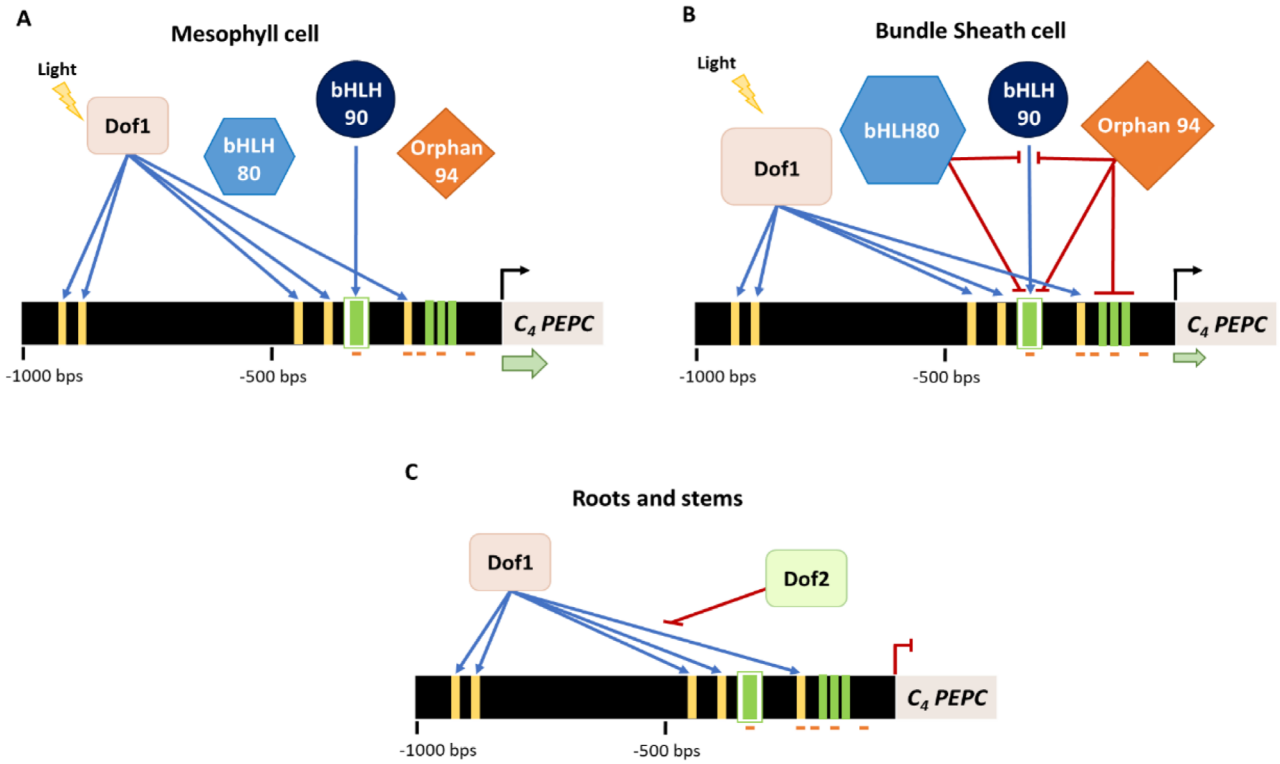


Figure 3 - Schematic representation of the different mechanisms proposed to regulate the transcription of C_4 *ZmPEPC* in an organ- and cell-specific way. (A) Regulation of C_4 *ZmPEPC* gene expression in M cells. The repressors ZmbHLH80 and ZmOrphan94 are less expressed than in BS cells, therefore there is a high gene expression activation by ZmbHLH90. (B) Regulation of C_4 *ZmPEPC* gene expression in BS cells. ZmbHLH80 and ZmOrphan94 are preferentially expressed in BS cells, working as repressors of ZmbHLH90, leading to a down-regulation of C_4 ZmPEPC expression. ZmbHLH80 and ZmOrphan94 can impair ZmbHLH90 function through heterodimerization or competitive binding for the same binding site. In addition, ZmOrphan94 may also impair ZmbHLH90 through its binding to CACA motifs, close to ZmbHLH90 binding site. In leaves, Dof1 is activated by light, allowing its binding and consequent activation of C_4 *ZmPEPC* gene expression (A and B). (C) Regulation of C_4 *ZmPEPC* in stems and roots by Dof1 and Dof2. These TFs are both expressed in these tissues, however, while Dof1 bind to the respective cis-elements in the C_4 *ZmPEPC* promoter to activate gene expression, Dof2 binds them to block Dof1 DNA-interaction, thus impairing C_4 *ZmPEPC* expression. The black arrows and the red lines represent activation and repression of gene expression, respectively. The thickness of the green arrow represents the expression levels of C_4 *PEPC* in each cell type. Activation and repression by the different TFs are represented as blue arrows and red lines, respectively. The different sizes of Dof1, ZmbHLH80 and ZmOrphan94, between A and B denote their gene expression levels in each cell type. The yellow rectangles represent the binding sites of Dof1 and Dof2 (Yanagisawa and Sheen, 1998) and the green rectangles represent the ZmOrphan94 binding sites. The binding site of ZmbHLH80 and ZmbHLH90 (E-box) is represented by a white rectangle. Within this E-box, there is a CACA motif, which is represented by a green rectangle, similar to the other binding sites of ZmOrphan94. The orange lines underneath the promoter represent the CNSs identified by Gupta *et al.* (2020).

A few studies have identified unmethylated CpG islands in the C_4 *PEPC* promoter (Langdale *et al.*, 1991; Tolley *et al.*, 2012). These regions, along with H3K4me3 may maintain an open chromatin state. Despite these CpG islands being unmethylated in both M and BS cells, a similar hypothesis regarding the recruitment of a methyltransferase has been proposed (Tolley *et al.*, 2012). This way, an open chromatin conformation is maintained, and transcription can be induced in M cells (Tolley *et al.*, 2012). Nevertheless, the identification and functional characterization of such methyltransferase(s) or de-methylase(s) is still to be carried out.

Although progress has been made over the last years towards a better understanding of the gene regulatory mechanisms underlying C_4 *PEPC* cell-specific gene expression, there is still a lot more to be unveiled. More progress has been done regarding the characterization of important *cis*-elements than in the identification and characterization of key *trans*-factors regulating C_4 *PEPC* cell-specificity. Although some TFs have been identified as binding to the C_4 *PEPC* promoter and as putative regulators of C_4 *PEPC* cell-specific gene expression, the key players are still missing. It is still to be identified the key TF or TFs that promote or impair C_4 *PEPC* cell-specific gene expression. Therefore, we believe that more effort is necessary to identify new TFs regulating C_4 *PEPC* gene expression and to understand the signalling pathways and the regulatory networks involved.

Diel regulation

The circadian clock is an internal mechanism that regulates several biological processes, including C_4 photosynthesis (Khan *et al.*, 2010). Although the effects of the circadian clock on C_4 *PEPC* gene expression remain largely unknown, a few studies have shown that similarly to other C_4 genes, C_4 *PEPC* gene expression has a circadian regulation (Horst *et al.*, 2009; Khan *et al.*, 2010). C_4 *PEPC* is an early morning phasing gene and, despite its light regulation, it presents an oscillatory rhythm under constant light (Horst *et al.*, 2009; Khan *et al.*, 2010; Xu *et al.*, 2016).

In the maize C_4 *PEPC* distal promoter region (1300 bps before ATG), some histone acetylation sites, such as H3K9ac, which has a high correlation with transcription activation, show circadian oscillation, maintaining its rhythmicity and high amplitude levels under constant light (Horst *et al.*, 2009). These observations show that, though regulators of C_4 *PEPC* cell-specific gene expression are located within the first 500 bp upstream of the translational start codon (Gupta *et al.*, 2020), the distal promoter region (1300 bps before ATG) might be more related to the C_4 *PEPC* gene expression level, as well as with the circadian regulation.

It was shown that, during the night period of a diel cycle, histone acetylation is not totally removed (Offermann *et al.*, 2006). These intermediary histone acetylation levels found during this period, contrast with the low acetylation levels found in this gene after a long period of dark exposure (Offermann *et al.*, 2006). Therefore, it was proposed that light regulates histone acetyltransferases (HATs), being also active under dark conditions to maintain steady-state acetylation levels (Offermann *et al.*, 2006). Therefore, one can hypothesise that HATs' activity or expression levels may also be regulated by the circadian clock. Nevertheless, it was shown that high histone acetylation of the C_4 *PEPC* promoter may not be enough to induce transcription. In maize, the treatment of darkened plant leaves with a histone deacetylase (HDAC) inhibitor did not alter C_4 *PEPC* gene expression (Offermann *et al.*, 2006).

As described above, ZmbHLH80 and ZmbHLH90 participate in C_4 *PEPC* regulation (Górska *et al.*, 2019). Interestingly, in *Arabidopsis thaliana*, FBH1, a homologous TF to ZmbHLH80 and ZmbHLH90, is involved in the circadian rhythm regulation by repressing the *CCA1* gene expression (Nagel *et al.*, 2014). FBH1 is also involved in the *CCA1* regulation in response to warm temperatures (Nagel *et al.*, 2014). It would be interesting to understand if this mechanism is conserved in maize, and other C_4 species, and to unveil the regulators involved. This will help us to better understand how C_4 *PEPC* and, eventually, other C_4 genes are regulated by the circadian rhythm.

Light regulation

Light is an important environmental stimulus regulating the genes involved in C_4 photosynthesis, being C_4 *PEPC* one of the C_4 genes most responsive to light (Nelson *et al.*, 1984; Schaffner and Sheen, 1992; Kausch *et al.*, 2001; Offermann *et al.*, 2006; Offermann *et al.*, 2008; Burgess *et al.*, 2016; Xu *et al.*, 2016). In greening assays, C_4 *ZmPEPC* transcript level and promoter activity increase until several hours after illumination (Nelson *et al.*, 1984; Schaffner and Sheen, 1992; Kausch *et al.*, 2001; Xu *et al.*, 2016).

Despite the molecular mechanisms underlying C_4 *PEPC* light regulation being still unclear, this gene is known to be light-regulated at different levels. In C_4 *PEPC* distal promoter (between 3178 and 2908 bps before ATG) four cytosine residues were identified as differentially methylated in plants grown under different light conditions (Langdale *et al.*, 1991; Tolley *et al.*, 2012). These residues are less methylated in M cells of green leaves, compared with etiolated leaves or roots (Langdale *et al.*, 1991; Tolley *et al.*, 2012). In greening leaves, an increase in demethylation of two of these cytosine residues was also observed within 48h of light exposure (Langdale *et al.*, 1991). However, although the demethylation of these residues has a good correlation with the increase of C_4 *ZmPEPC* transcript levels, it does not seem to be important for the cell-specific transcription of this gene, since its proximal promoter region is sufficient to drive M cell-specific expression (Tolley *et al.*, 2012; Gupta *et al.*, 2020). Nevertheless, it is possible that upstream differentially-methylated regions can act as enhancers of C_4 *ZmPEPC* expression in M cells, being their contribution to C_4 *PEPC* expression still unclear (Tolley *et al.*, 2012).

In greening maize leaves, the chromatin of the proximal promoter region (500 bps before ATG) has an open state, compared with the chromatin of the same region in etiolated leaves, showing that light modulates chromatin dynamics of this region of C_4 *PEPC* promoter (Kalamajka *et al.*, 2003). In species from different C_4 evolution origins, some histone acetylation sites in both coding and promoter regions of C_4 *PEPC* are regulated by light (Table 1) (Offermann *et al.*, 2006;

Table 1 – Histone modifications found in C_4 *PEPC* gene promoter and regulated processes.

Associated process	Modification	PEPC promoter region	Specie	Reference
Cell-specificity	H3K4me3 H3K4me2	Proximal	*Zea mays*	Danker *et al.*, 2008
Circadian regulation	H3K9ac	Distal	*Zea mays*	Horst *et al.*, 2009
Light regulation	H3K9ac	Proximal	*Zea mays* *Setaria italica*	Offermann *et al.*, 2008; Horst *et al.*, 2009;
		Distal	*Zea mays* *Sorghum bicolor*	Heimann *et al.*, 2013
	H4K5ac	Proximal	*Zea mays* *Setaria italica*	Offermann *et al.*, 2008; Horst *et al.*, 2009;
		Distal	*Zea mays* *Sorghum bicolor*	Heimann *et al.*, 2013
	H4K16Ac	Proximal and Distal	*Zea mays*	Horst *et al.*, 2009
	H3K23Ac	Proximal and Distal	*Zea mays*	Horst *et al.*, 2009

Offermann *et al.*, 2008; Horst *et al.*, 2009; Heimann *et al.*, 2013). A comparison between both distal and proximal C_4 *ZmPEPC* promoter regions revealed that acetylation levels have a stronger light response and higher correlation with transcription in C_4 *ZmPEPC* distal promoter regions (Horst *et al.*, 2009). This further supports the idea that the distal promoter of C_4 *PEPC* may contribute as an enhancer of C_4 *PEPC* gene expression.

To control C_4 *PEPC* acetylation levels, light modulates histone deactylases' (HDACs) activity (Offermann *et al.*, 2006; Offermann *et al.*, 2008). During the night period, some HDACs are activated to deacetylate the C_4 *PEPC* promoter. During the day, although some HDACs are repressed, others are activated to maintain the steady-state histone acetylation levels (Offermann *et al.*, 2006; Offermann *et al.*, 2008). This shows that HDACs seem to be important to regulate the acetylation levels of C_4 *PEPC*, however the HDACs involved in this regulation remain to be identified. It has long been known that light has an important role in modulating the binding of proteins to the C_4 *PEPC* promoter (Kano-Murakami *et al.*, 1991). *In vitro* experiments showed that nuclear factors extracted from green maize leaves are able to bind to the C_4 *ZmPEPC* promoter, whilst the nuclear factors extracted from etiolated maize leaves are not. (Kano-Murakami *et al.*, 1991) A good example of a TF binding to the C_4 *PEPC* promoter in a light-dependent manner is Dof1, whose activity is modulated by light (Yanagisawa and Sheen, 1998). Dof1 can induce higher C_4 *PEPC* promoter activity in greening as compared with etiolated protoplasts (Yanagisawa and Sheen, 1998). Since both blue and red light induce the expression of C_4 *PEPC*, it seems that both phytochrome and the cryptochrome pathways contribute to the regulation of C_4 *PEPC* gene. However, the downstream players of this regulation remain to be unveiled (Hendron and Kelly, 2020). Being light an important stimulus regulating C_4 *PEPC* expression, it would be interesting to identify and characterize more TFs that regulate C_4 *PEPC* in response to light and unveil the regulatory mechanisms of the different photoreceptors.

Besides light playing a crucial role in regulating C_4 *PEPC* gene expression, the signals originated from the interplay between light and chloroplast development seem to be relevant for C_4 *PEPC* regulation (Kausch *et al.*, 2001; Burgess *et al.*, 2016). The inhibition of chloroplast development reduces the activation of the C_4 *ZmPEPC* promoter and an increase in C_4 *ZmPEPC* expression was observed in greening maize seedlings (Kausch *et al.*, 2001; Burgess *et al.*, 2016). Although one can hypothesise that chloroplast development is a relevant component of C_4 *PEPC* gene expression regulation, the regulatory mechanisms are still unknown.

Despite being a crucial environmental cue regulating C_4 *PEPC* gene expression, the regulatory mechanisms underlying light response need to be further investigated to better understand this topic. It would be interesting to unveil the regulatory mechanisms involved in the epigenetic modifications of C_4 *PEPC* promoter in response to light and understand their relevance for C_4 photosynthesis. The identification of TFs and *cis*-elements and downstream players of the different photoreceptor pathways involved in the regulation of C_4 *PEPC* is also important for understanding the light regulatory networks. Finally, retrograde signalling is a rather unexplored topic regarding C_4 *PEPC* expression. Since it seems to be a relevant component of C_4 *PEPC* regulation, it would be important to understand the regulatory mechanisms involved in this process and the interplay between light and retrograde signalling.

Response of C_4 PEPC to adverse environmental conditions

Plants are sessile organisms that cannot escape from adverse environmental conditions. To cope with such conditions, plants need to re-arrange their metabolism. Photosynthesis is a key process for life on Earth, being essential for many different ecosystems. Alterations in this metabolic pathway can lead to serious decreases in plant yield, which is detrimental to our current agricultural systems. It is of utmost importance to understand how the adverse environmental conditions modulate the photosynthetic metabolism. Given the importance of C_4 photosynthesis, it is particularly important to understand how this metabolism is affected by different environmental stresses. One of the key enzymes in C_4 photosynthesis is C_4 PEPC, but the mechanisms by which this protein is regulated under stress conditions remain unclear. Here we summarise the current knowledge regarding the effects of various stress conditions on C_4 *PEPC* gene expression. Table 2 summarises the reported effects of different abiotic stresses on C_4 PEPC levels.

Osmotic stress

Different adverse environmental conditions alter the osmotic balance within the cell, leading to osmotic stress. These conditions include for instance water deficit, salt stress (osmotic component), or osmolyte pressure (e.g. PEG-mediated drought). Although some studies have investigated the impact of osmotic stress in C_4 plants it is still not clear its effect on the C_4 cycle, with many authors claiming that the CBB cycle is the major limiting step in osmotic stress tolerance in C_4 plants.

Several reports have shown a decrease in C_4 PEPC expression and activity in response to water deficit (Pelleschi *et al.*, 1997; Foyer *et al.*, 1998) but other authors have seen an increase of its activity under water deficit (Ghannoum, 2009). An increase in PEPC levels would raise the initial carboxylation of atmospheric CO_2 and increase the carbon flux to BS. If not accompanied by an increase of Rubisco-mediated carboxylation, this increase would lead to decreased net carbon fixation, and subsequent CO_2 leakage. Major effect of osmotic stress is the decrease of photosynthetic rate in both C_3 and C_4 plants. It has been proposed that, in C_4 plants, an increase of non-used CO_2 in the BS cells (i.e. $\uparrow[CO_2]_{BS}$) leads to CO_2 leakage and subsequent decrease in net photosynthesis (Ghannoum, 2009), which could be linked with the changes in PEPC levels described in some works.

Jeanneau *et al.*, 2002 tested the effect of overexpression of *Sorghum bicolor* C_4 *PEPC* in drought tolerance in maize. They observed an increase in carbon assimilation rates in lines with increased C_4 *PEPC* expression and a decrease in the lines with decreased C_4 *PEPC* expression, as it was expected. In terms of drought tolerance, no effect of the overexpression

Table 2 – Summary of the abiotic stress effects in C_4 PEPC levels.

Stress condition	Species	Regulatory effect	Reference
Osmotic stress	*Zea mays*	Decrease transcipt	Pelleschi *et al.*, 1997; Foyer *et al.*, 1998
		Increase activity	Ghannoum, 2009
	Sorghum bicolor	Decrease transcript	Buchanan *et al.*, 2005
Salt stress	*Zea mays*	Increase activity	Hatzig *et al.*, 2010)
	Sorghum bicolor	Increase transcript	Buchanan *et al.*, 2005
Cold	*Zea mays*	Decrease activity	Selinioti *et al.*, 1985; Angelopoulos and Gavalas, 1988; Chinthapalli *et al.*, 2003
Heat	*Zea mays*	Increase activity	Crafts-Brandner and Salvucci, 2002; Chinthapalli *et al.*, 2003
Nitrogen deficiency	*Zea mays*	Decrease transcript/protein	Sugiharto and Sugiyama, 1992; Sugiharto *et al.*, 1992; Sugiharto *et al.*, 1990
Cadmium excess	*Zea mays*	Decrease activity	Wang *et al.*, 2009
Ozone excess	*Zea mays*	Decrease transcript/protein	Leitao *et al.*, 2007b; Leitao *et al.*, 2007a

of C_4 *PEPC* in severe drought conditions was observed, but plants showed a higher water use efficiency in mild-drought conditions. Together, C_4 PEPC plays a role in regulating the carbon flux from M to BS cells, the increase of this flow may be beneficial in the early stages of drought but under more severe water deficit it becomes irrelevant. Overexpression of C_4 PEPC alone seems to lead to an increase in transported CO_2 that may not be efficiently used by Rubisco, either by Rubisco limitation or decarboxylation inefficiency, possibly due to a lack of increase in decarboxylation enzymes (e.g. NADP-ME).

Under salt stress, C_4 plants showed higher PEPC activity contrary to C_3 plants (Hatzig *et al.*, 2010). There are no insights showing that this increase is linked to upregulation of photosynthesis but rather for the anaplerotic role of PEPC. It would be interesting to understand which component of the salt stress (osmotic or ionic) is indeed responsible for the upregulation of PEPC and which PEPCs are regulated at transcriptional level.

Work on *Sorghum bicolor*, analysed the genome wide transcriptional response to salt, PEG and ABA stress in both shoot and roots (Buchanan *et al.*, 2005). In terms of C_4 *PEPC* transcripts, it was observed an upregulation upon salt stress in both roots and shoots, which is in agreement with previous work in maize (Hatzig *et al.*, 2010). PEG induced osmotic stress led to down regulation in roots but no changes in shoots, which is contrary to previous results in maize where either upregulation (Ghannoum, 2009) or downregulation (Pelleschi *et al.*, 1997; Foyer *et al.*, 1998) of C_4 *PEPC* was observed. Abscisic acid treatment, a key hormone in stress response, leads to no change in *PEPC* transcript.

Most genome wide studies in maize show no significant transcriptional response for C_4 *ZmPEPC*, in both biotic and abiotic stresses [data obtained via Genevestigator (https://genevestigator.com/)].

Temperature stress

High and low temperatures affect photosynthesis in both C_3 and C_4 plants. C_4 plants are considered to be more sensitive to cold stress than C_3 plants, due to the cold-labile feature of some C_4 enzymes (Long, 1983). Plants that are more tolerant to low temperature usually show a higher accumulation of photosynthesis related enzymes, like Rubisco (Yamori *et al.*, 2014). It was therefore expected that C_4 plants under cold stress accumulated C_4 related enzymes to counterbalance their reduced activity. Contrary to what was expected, C_4 plants seem to show a decrease in PEPC activity under cold (Selinioti *et al.*, 1985; Angelopoulos and Gavalas, 1988; Chinthapalli *et al.*, 2003). It would be important to understand the transcriptional regulation and how knock-out or overexpression of C_4 *PEPC* would affect temperature tolerance.

Although cold decreases C_4 PEPC activity, this effect is reversible when plants are placed back on optimal conditions. Though changes in activity its many times related to the phosphorylation of C_4 PEPC, (Chinthapalli *et al.*, 2003) showed that there are no changes in the phosphorylation status of C_4 PEPC when treated with different temperature conditions, thus refuting the hypothesis of regulation by phosphorylation. The same study showed that C_4 PEPC has increased activity at higher temperatures, in a way that is remarkably different from its C_3 counterpart. On the other hand, (Crafts-Brandner and Salvucci, 2002) showed that C_4 PEPC activity is rather insensitive to increase in temperature, although photosynthesis was reduced at temperatures higher than 40ºC. It would be important to investigate how different temperature conditions regulated C_4 *PEPC* gene expression and how this correlates with photosynthesis efficiency.

Nitrogen levels regulation

Nitrogen deficiency is well known to cause a down regulation of C_4 *PEPC* transcript and protein levels, in maize

leaves (Sugiharto *et al.*, 1990; Schlüter *et al.*, 2012). On the other hand, upon nitrogen treatment, regardless of the form supplied (nitrate or ammonium), C_4 *PEPC* transcript level and activity are significantly up regulated in maize (Sugiharto and Sugiyama, 1992; Suzuki *et al.*, 1994). This up regulation is thought to be mediated by Glutamic acid, as its addition leads to an upregulation of the C_4 *PEPC* gene expression and the inhibition of its synthesis leads to a down regulation (Sugiharto *et al.*, 1992). Nevertheless, the addition of ammonium does not affect the C_4 *PEPC* gene expression in sorghum (Arias-Baldrich *et al.*, 2017), indicating that regulation of C_4 *PEPC* gene expression by nitrate or ammonium treatment may differ even among close C_4 species. The fact that C_4 *PEPC* gene expression can be modulated by nitrogen levels shows an intrinsic interplay between carbon and nitrogen metabolism, which may have been co-opted during C_4 evolution.

Other stresses

It has been reported that cadmium affects the growth of maize plants by disturbing the light and carbon reactions of photosynthesis. High cadmium levels lead to a down regulation of C_4 PEPC activity in maize, with the dosage affecting the time needed to see the effects (Wang *et al.*, 2009). Whether this regulation takes place at the transcriptional level is not known.

Atmospheric conditions can also affect photosynthesis, namely the increase in ozone concentration. It has been shown that increase in atmospheric ozone led to impacts in maize growth and in its photosynthetic potential. Although the light harvesting complex is affected at relatively low increases of ozone, the carbon fixation reactions namely PEPC and Rubisco, are only affected at higher concentration with a reduction in protein amount and transcript (Leitao *et al.*, 2007a, b).

Concluding remarks

During plant evolution, PEPCs evolved from bacterial PEPCs, after an ancestral duplication, when Viridiplantae arose. In C_3 plants, PEPC is an important enzyme for plant development since it works as a link between carbon and nitrogen metabolism. Later, during C_4 evolution, PEPC was recruited independently several times to incorporate the C_4 cycle, by performing the first step of CO_2 fixation. However, to obtain the features required for C_4 photosynthesis operation, it was necessary to modify the mechanisms that regulate its gene expression, as well as protein accumulation and activity. Therefore, to engineer the C_4 metabolism, it is crucial to understand the C_4 PEPC regulatory network.

The regulation of C_4 PEPC is complex, being modulated at several levels. At the epigenetic level, patterns of histone methylation were associated with the establishment of cell specificity. However, the mechanisms that maintain this pattern remain unknown. It would be interesting to investigate if there are methyltransferases recruited to the promoter in a cell-specific way, to induce higher levels of histone methylation, contributing to gene activation. If this is true, it would also be important to know which methyltransferases are recruited and the mechanisms underlying this process. Similarly, a deeper understanding of the role of CpG islands for the establishment of cell-specificity of C_4 *PEPC* gene expression, would also be an interesting topic to investigate. Histone acetylation has been associated with light and circadian regulation and even not being crucial for C_4 *PEPC* regulation, it may contribute. It would be interesting to investigate if histone acetylation can function as prerequisite to enable C_4 *PEPC* transcription. In addition, it seems that different photoreceptors, may also be involved in C_4 *PEPC* transcriptional regulation, since blue and red light induce C_4 *PEPC* gene expression. In the future, it would be relevant to further characterise the regulatory mechanisms of C_4 *PEPC* by the different photoreceptors, to better understand C_4 *PEPC* light response.

To establish cell-specificity, *cis*-elements and *trans*-factors were recruited during C_4 evolution. Although some progress has been made to characterise C_4 *PEPC* promoters and to identify putative regulatory *cis*-elements, there is still a gap regarding the identification and characterization of new *trans*-factors. It would be interesting to know which TFs bind to MEM1, a crucial *cis*-element defining cell-specificity in *Flaveria* species. In monocots, some TFs have been identified as putative regulators of cell-specificity. However, their relevance to establish cell-specificity and to C_4 photosynthesis efficiency still needs to be demonstrated. The identification and characterization of key TFs to establish C_4 PEPC cell-specificity in both monocots and dicots would be crucial to better understand these mechanisms. Furthermore, in both dicots and monocots, there are certainly relevant *cis*-elements in C_4 *PEPC* gene promoter, involved in gene expression that remain to be identified.

The circadian regulation of C_4 *PEPCs* is the most unexplored regulatory mechanism presented in this review. It is known that the circadian clock regulates C_4 *ZmPEPC* at transcriptional level and its expression is regulated by ZmbHLH80 and ZmbHLH90. Since the Arabidopsis homologue for these two TFs, FBH1, regulates circadian clock through the transcriptional regulation of *CCA1*, it would be interesting to know if ZmbHLH80 and ZmbHLH90 could be involved in the circadian regulation of *ZmPEPC1* and if the regulation of *CCA1* is conserved.

Different species have distinct regulatory mechanisms to regulate developmental C_4 *PEPC* gene expression and protein accumulation, which is not surprising, given that C_4 photosynthesis is a convergent evolutionary event. Despite these differences, in all species, M cell differentiation seems to be important for a high C_4 *PEPC* gene expression and protein accumulation. However, the regulatory mechanisms underlying leaf development are still poorly understood. In the future, it would be interesting to identify the internal cues involved in establishing M cell specificity along the developmental gradient.

The photosynthetic metabolism underpins the synthesis of carbohydrates needed for plant growth and reproduction. Adverse environmental conditions that negatively affect photosynthesis will impair plant growth and yield. It is therefore important to understand how photosynthesis responds to environmental stresses and find ways to improve such responses. In C_4 photosynthesis, C_4 PEPC plays an important role in carbon fixation, being responsible for the first carboxylation step in the cycle. Because of this role, C_4 PEPC is tightly regulated and responds to environmental stimuli, such as water availability, light, nutritional signals,

and atmospheric conditions. The regulation of C_4 PEPC is poorly understood, but the effects of different environmental clues have been described. The regulation of C_4 PEPC levels in response to stress is important to regulate the carbon flux into the C_4 cycle, thus regulating the photosynthetic efficiency of the plant. It is difficult to distinguish between the role of C_4 PEPC in the C_4 cycle and its role in anaplerotic reactions. Being C_4 PEPC an important enzyme for the C/N balance, its regulation can impact several metabolic pathways, making it a good target for improvement of plant stress response.

In conclusion, C_4 evolution represents one of the most impressive cases of convergent evolution in Nature that has occurred independently over 60 times in very distant species. Nevertheless, their carbon concentration mechanisms always rely on a C_4 PEPC, which is tightly regulated by internal and environmental cues. Since the function of C_4 PEPC in C_4 photosynthesis, combined with its anaplerotic role, makes it an important modulator of plant growth and yield, it is of utmost importance to better understand the gene regulatory network (including its evolution) modulating its expression and function.

Acknowledgments

This work was funded by Fundação para a Ciência e a Tecnologia (Portugal) through the R&D unit "GREEN-IT Bioresources for Sustainability" (DOI 10.54499/UIDB/04551/2020 & DOI 10.54499/UIDP/04551/2020) and the fellowships to PC (ref. PD/BD/128403/2017) and CG (ref. 2020.08000.BD).

Conflict of Interests

The authors declare that there is no conflict of interest that could be perceived as prejudicial to the impartiality of the reported research.

Authors Contributions

PC, CG, and NJMS conceived the review. PC and CG wrote the manuscript. NJMS reviewed the manuscript. All authors read and approved the final version.

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

Angelopoulos K and Gavalas NA (1988) Reversible cold inactivation of C_4-phosphoenolpyruvate carboxylase: Factors affecting reactivation and stability. J Plant Physiol 132:714-719.

Arias-Baldrich C, de la Osa C, Bosch N, Ruiz-Ballesta I, Monreal JA and García-Mouriño S (2017) Enzymatic activity, gene expression and posttranslational modifications of photosynthetic and non-photosynthetic phosphoenolpyruvate carboxylase in ammonium-stressed sorghum plants. J Plant Physiol 214:39-47.

Aubry S, Kelly S, Kümpers BMC, Smith-Unna RD and Hibberd JM (2014) Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of C_4 photosynthesis. PLoS Genet 10:e1004365.

Besnard G, Muasya AM, Russier F, Roalson EH, Salamin N and Christin P-A (2009) Phylogenomics of C_4 photosynthesis in sedges (Cyperaceae): Multiple appearances and genetic convergence. Mol Biol Evol 26:1909-1919.

Bläsing OE, Westhoff P and Svensson P (2000) Evolution of C_4 phosphoenolpyruvate carboxylase in Flaveria, a conserved serine residue in the carboxyl-terminal part of the enzyme is a major determinant for C_4-specific characteristics. J Biol Chem 275:27917-27923.

Buchanan CD, Lim S, Salzman RA, Kagiampakis I, Morishige DT, Weers BD, Klein RR, Pratt LH, Cordonnier-Pratt MM, Klein PE *et al.* (2005) *Sorghum bicolor*'s transcriptome response to dehydration, high salinity and ABA. Plant Mol Biol 58:699-720.

Burgess SJ, Granero-Moya I, Grangé-Guermente MJ, Boursnell C, Terry MJ and Hibberd JM (2016) Ancestral light and chloroplast regulation form the foundations for C_4 gene expression. Nat Plants 2:16161.

Caburatan L, Kim J and Park J (2019) Expression profiles and post-translational modifications of phosphoenolpyruvate carboxylase isozymes of *Bienertia sinuspersici* during leaf development. Russ J Plant Physiol 66:738-747.

Cavalar M, Phlippen Y, Kreuzaler F and Peterhänsel C (2007) A drastic reduction in DOF1 transcript levels does not affect C_4-specific gene expression in maize. J Plant Physiol 164:1665-1674.

Chang YM, Chang CL, Li WH and Shih ACC (2013) Historical profiling of maize duplicate genes sheds light on the evolution of C_4 photosynthesis in grasses. Mol Phylogenet Evol 66:453-462.

Chinthapalli B, Murmu J and Raghavendra AS (2003) Dramatic difference in the responses of phosphoenolpyruvate carboxylase to temperature in leaves of C_3 and C_4 plants. J Exp Bot 54:707-714.

Chollet R, Vidal J and O'Leary MH (1996) Phosphoenolpyruvate carboxylase: A ubiquitous, highly regulated enzyme in plants. Annu Rev Plant Physiol Plant Mol Biol 47:273-298.

Christin P and Besnard G (2009) Two independent C_4 origins in Aristidoideae (Poaceae) revealed by the recruitment of distinct phosphoenolpyruvate carboxylase genes. Am J Bot 96:2234-2239.

Christin PA, Salamin N, Savolainen V, Duvall MR and Besnard G (2007) C_4 Photosynthesis evolved in grasses via parallel adaptive genetic changes. Curr Biol 17:1241-1247.

Crafts-Brandner SJ and Salvucci ME (2002) Sensitivity of photosynthesis in a C_4 plant, maize, to heat stress. Plant Physiol 129:1773-1780.

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

Dengler NG, Dengler RE, Donnelly PM and Filosa MF (1995) Expression of the C_4 pattern of photosynthetic enzyme accumulation during leaf development in *Atriplex rosea* (Chenopodiaceae). Am J Bot 82:318-327.

Engelmann S, Bläsing OE, Gowik U, Svensson P and Westhoff P (2003) Molecular evolution of C_4 phospho enol pyruvate carboxylase in the genus Flaveria—a gradual increase from C_3 to C_4 characteristics. Planta 217:717-725.

Engelmann S, Zogel C, Koczor M, Schlue U, Streubel M and Westhoff P (2008) Evolution of the C_4 phosphoenolpyruvate carboxylase promoter of the C_4 species *Flaveria trinervia*: The role of the proximal promoter region. BMC Plant Biol 8:4.

Foyer CH, Kingston-Smith A, Pastori G and Harbinson J (1998) Photosynthesis and antioxidant metabolism in maize leaves subjected to low temperatures. In: Garab G (ed) Photosynthesis: Mechanisms and effects. Springer, Dordrecht, pp 2425-2431.

Ghannoum O (2009) C_4 photosynthesis and water stress. Ann Bot 103:635-644.

Górska AM, Gouveia P, Borba AR, Zimmermann A, Serra TS, Lourenço TF, Margarida Oliveira M, Peterhänsel C and

Saibo NJM (2019) ZmbHLH80 and ZmbHLH90 transcription factors act antagonistically and contribute to regulate *PEPC1* cell-specific gene expression in maize. Plant J 99:270-285.

Górska AM, Gouveia P, Borba AR, Zimmermann A, Serra TS, Carvalho P, Lourenço TF, Oliveira MM, Peterhänsel C and Saibo NJM (2021) ZmOrphan94 Transcription factor downregulates *ZmPEPC1* gene expression in maize bundle sheath cells. Front Plant Sci 12:559967.

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.

Gupta S das, Levey M, Schulze S, Karki S, Emmerling J, Streubel M, Gowik U, Paul Quick W and Westhoff P (2020) The *C_4Ppc* promoters of many C_4 grass species share a common regulatory mechanism for gene expression in the mesophyll cell. Plant J 101:204-216.

Hatzig S, Kumar A, Neubert A and Schubert S (2010) PEP-carboxylase activity: A comparison of its role in a C_4 and a C_3 species under salt stress. J Agron Crop Sci 196:185-192.

Heimann L, Horst I, Perduns R, Dreesen B, Offermann S and Peterhansel C (2013) A common histone modification code on C_4 genes in maize and its conservation in sorghum and *Setaria italica.* Plant Physiol 162:456-69.

Hendron R-W and Kelly S (2020) Subdivision of light signaling networks contributes to partitioning of C_4 photosynthesis. Plant Physiol 182:1297-1309.

Horst I, Offermann S, Dreesen B, Niessen M and Peterhansel C (2009) Core promoter acetylation is not required for high transcription from the phosphoenolpyruvate carboxylase promoter in maize. Epigenetics Chromatin 2:17.

Jeanneau M, Gerentes D, Foueillassar X, Zivy M, Vidal J, Toppan A and Perez P (2002) Improvement of drought tolerance in maize: Towards the functional validation of the *Zm-Asr1* gene and increase of water use efficiency by over-expressing C_4-PEPC. Biochimie 84:1127-1135.

Kalamajka R, Hahnen S, Cavalar M, Töpsch S, Weier D and Peterhänsel C (2003) Restriction accessibility in isolated nuclei reveals light-induced chromatin reorganization at the *PEPC* promoter in maize. Plant Mol Biol 52:669-78.

Kano-Murakami Y, Suzuki I, Sugiyama T and Matsuoka M (1991) Sequence-specific interactions of a maize factor with a GC-rich repeat in the phosphoenolpyruvate carboxylase gene. Mol Gen Genet 225:203-208.

Kausch AP, Page Owen J, Zachwieja SJ, Flynn AR and Sheen J (2001) Mesophylspecific, light and metabolic regulation of the C_4 P*PCZm1* promoter in transgenic maize. Plant Mol Biol 45:1-15.

Khan S, Rowe SC and Harmon FG (2010) Coordination of the maize transcriptome by a conserved circadian clock. BMC Plant Biol 10:126.

Koteyeva NK, Voznesenskaya EV, Cousins AB and Edwards GE (2014) Differentiation of C_4 photosynthesis along a leaf developmental gradient in two Cleome species having different forms of Kranz anatomy. J Exp Bot 65:3525-3541.

Langdale JA, Rothermel BA and Nelson T (1988) Cellular pattern of photosynthetic gene expression in developing maize leaves. Genes Dev 2:106-115.

Langdale JA, Taylor WC and Nelson T (1991) Cell-specific accumulation of maize phosphoenolpyruvate carboxylase is correlated with demethylation at a specific site > 3 kb upstream of the gene. Mol Gen Genet 225:49-55.

Leitao L, Bethenod O and Biolley JP (2007a) The impact of ozone on juvenile maize (*Zea mays* L.) plant photosynthesis: Effects on vegetative biomass, pigmentation, and carboxylases (PEPc and Rubisco). Plant Biol (Stuttg) 9:478-488.

Leitao L, Maoret JJ and Biolley JP (2007b) Changes in PEP carboxylase, Rubisco and Rubisco activase mRNA levels from maize (*Zea mays*) exposed to a chronic ozone stress. Biol Res 40:137-153.

Long SP (1983) C_4 photosynthesis at low temperatures. Plant Cell Environ 6:345-363.

Majeran W, Friso G, Ponnal L, Connolly B, Huang M, Reidel E, Zhang C, Asakura Y, Bhuiyan NH, Sun Q *et al.* (2010) Structural and metabolic transitions of C_4 leaf development and differentiation defined by microscopy and quantitative proteomics in maize. Plant Cell 22:3509-3542.

Martineau B and Taylor WC (1985) Photosynthetic gene expression and cellular differentiation in developing maize leaves. Plant Physiol 78:399-404.

Mayfield SP and Taylor WC (1984) The appearance of photosynthetic proteins in developing maize leaves. Planta 161:481-486.

Nagel DH, Pruneda-Paz JL and Kay SA (2014) FBH1 affects warm temperature responses in the Arabidopsis circadian clock. Proc Natl Acad Sci U S A 111:14595-14600.

Nelson T and Langdale JA (1989) Patterns of leaf development in C_4 plants. Plant Cell 1:3-13.

Nelson T, Harpster MH, Mayfield SP and Taylor WC (1984) Light-regulated gene expression during maize leaf development. J Cell Biol 98:558-564.

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

Offermann S, Danker T, Dreymüller D, Kalamajka R, Töpsch S, Weyand K and Peterhänsel C (2006) Illumination is necessary and sufficient to induce histone acetylation independent of transcriptional activity at the C_4-specific phosphoenolpyruvate carboxylase promoter in maize. Plant Physiol 141:1078-1088.

O'Leary MH (1982) Phosphoenolpyruvate carboxylase: An enzymologist' s view. Ann Rev Plant Physiol 33:297-315.

O'Leary B, Park J and Plaxton WC (2011) The remarkable diversity of plant PEPC (phosphoenolpyruvate carboxylase): Recent insights into the physiological functions and post-translational controls of non-photosynthetic PEPCs. Biochem J 436:15-34.

Pelleschi S, Rocher JP and Prioul JL (1997) Effect of water restriction on carbohydrate metabolism and photosynthesis in mature maize leaves. Plant Cell Environ 20:493-503.

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

Ramsperger VC, Summers RG and Berry JO (1996) Photosynthetic gene expression in meristems and during initial leaf development in a C_4 dicotyledonous plant. Plant Physiol 111:999-1010.

Sage RF (2004) The evolution of C_4 photosynthesis. New Phytol 161:341-370.

Sage RF, Christin PA and Edwards EJ (2011) The C_4 plant lineages of planet Earth. J Exp Bot 62:3155-3169.

Sánchez R and Cejudo FJ (2003) Identification and expression analysis of a gene encoding a bacterial-type phosphoenolpyruvate carboxylase from Arabidopsis and rice. Plant Physiol 132:949-957.

Schaffner AR and Sheen J (1992) Maize C_4 photosynthesis involves differential regulation of phosphoenolpyruvate carboxylase genes. Plant J 2:221-232.

Schlüter U, Mascher M, Colmsee C, Scholz U, Bräutigam A, Fahnenstich H and Sonnewald U (2012) Maize source leaf adaptation to nitrogen deficiency affects not only nitrogen and carbon metabolism but also control of phosphate homeostasis. Plant Phys 160:1384-1406.

Selinioti E, Karabourniotis G, Manetas Y and Gavalas NA (1985) Modulation of phosphoenolpyruvate carboxylase by 3-phosphoglycerate: Probable physiological significance for C_4-photosynthesis. J Plant Physiol 121:353-360.

Shi J, Yi K, Liu Y, Xie L, Zhou Z, Chen Y, Hu Z, Zheng T, Liu R, Chen Y *et al.* (2015) Phosphoenolpyruvate carboxylase in Arabidopsis leaves plays a crucial role in carbon and nitrogen metabolism. Plant Physiol 167:671-681.

Soros CL and Dengler NG (2001) Ontogenetic derivation and cell differentiation in photosynthetic tissues of C_3 and C_4 Cyperaceae. Am J Bot 88:992-1005.

Stockhaus J, Schlue U, Koczor M, Chitty JA, Taylor WC and Westhoff P (1997) The promoter of the gene encoding the C_4 form of phosphoenolpyruvate carboxylase directs mesophyll-specific expression in transgenic C_4 Flaveria spp. Plant Cell 9:479-489.

Sugiharto B and Sugiyama T (1992) Effects of nitrate and ammonium on gene expression of phosphoenolpyruvate carboxylase and nitrogen metabolism in maize leaf tissue during recovery from nitrogen stress. Plant Physiol 98:1403-1408.

Sugiharto B, Suzuki I, Burnell JN and Sugiyama T (1992) Glutamine induces the N-dependent accumulation of mRNAs encoding phosphoenolpyruvate carboxylase and carbonic anhydrase in detached maize leaf tissue. Plant Physiol 100:2066-2070.

Sugiharto B, Miyata K, Nakamoto H, Sasakawa H and Sugiyama T (1990) Regulation of expression of carbon-assimilating enzymes by nitrogen in maize leaf. Plant Physiol 92:963-969.

Suzuki I, Cretin C, Omata T and Sugiyama T (1994) Transcriptional and posttranscriptional regulation of nitrogen-responding expression of phosphoenolpyruvate carboxylase gene in maize. Plant Physiol 105:1223-1229.

Svensson P, Bläsing OE and Westhoff P (2003) Evolution of C_4 phosphoenolpyruvate carboxylase. Arch Biochem Biophys 414:180-188.

Taniguchi M, Izawa K, Ku MSB, Lin JH, 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 C_4 phosphoenolpyruvate carboxylase gene confers its differential transcription in mesophyll cells. Plant Mol Biol 44:543-557.

Tao S, Liu P, Shi Y, Feng Y, Gao J, Chen L, Zhang A, Cheng X, Wei H, Zhang T *et al.* (2022) Single-cell transcriptome and network analyses unveil key transcription factors regulating mesophyll cell development in maize. Genes (Basel) 13:374.

Tolley BJ, Woodfield H, Wanchana S, Bruskiewich R and Hibberd JM (2012) Light-regulated and cell-specific methylation of the maize *PEPC* promoter. J Exp Bot 63:1381-1390.

Voznesenskaya EV, Franceschi VR, Artyusheva EG, Black CC, Pyankov VI and Edwards GE (2003) Development of the C_4 photosynthetic apparatus in cotyledons and leaves of *Salsola richteri* (Chenopodiaceae). Int J Plant Sci 164:471-487.

Wakayama M, Ueno O and Ohnishi JI (2003) Photosynthetic enzyme accumulation during leaf development of *Arundinella hirta*, a C_4 grass having Kranz cells not associated with veins. Plant Cell Physiol 44:1330-1340.

Wang JL, Klessig DF and Berry JO (1992) Regulation of C_4 Gene Expression in Developing Amaranth Leaves. Plant Cell 4:173-184.

Wang JL, Long JJ, Hotchkiss T and Berry JO (1993) C_4 photosynthetic gene expression in light- and dark-grown amaranth cotyledons. Plant Physiol 102:1085-1093.

Wang H, Zhao SC, Liu RL, Zhou W and Jin JY (2009) Changes of photosynthetic activities of maize (*Zea mays* L.) seedlings in response to cadmium stress. Photosynthetica 47:277-283.

Westhoff P and Gowik U (2004) Evolution of C_4 phosphoenolpyruvate carboxylase. Genes and proteins: A case study with the genus Flaveria. Ann Bot 93:13-23.

Williams BP, Aubry S and Hibberd JM (2012) Molecular evolution of genes recruited into C_4 photosynthesis. Trends Plant Sci 17:213-220.

Xu J, Bräutigam A, Weber APM and Zhu XG (2016) Systems analysis of *cis*-regulatory motifs in C_4 photosynthesis genes using maize and rice leaf transcriptomic data during a process of de-etiolation. J Exp Bot 67:5105-5117.

Yamori W, Hikosaka K and Way DA (2014) Temperature response of photosynthesis in C_3, C_4, and CAM plants: Temperature acclimation and temperature adaptation. Photosynth Res 119:101-117.

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

Associate Editor: Loreta Brandão de Freitas