UNDERSTANDING C4 EVOLUTION AND FUNCTION

EDITED BY: Sarah Covshoff, Martha Ludwig, Florian A. Busch and Roxana Khoshravesh

PUBLISHED IN: Frontiers in Plant Science

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-864-1
DOI 10.3389/978-2-88971-864-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

UNDERSTANDING C4 EVOLUTION AND FUNCTION

Topic Editors:
Sarah Covshoff, Independent researcher, United States
Martha Ludwig, University of Western Australia, Australia
Florian A. Busch, University of Birmingham, United Kingdom
Roxana Khoshravesh, University of New Mexico, United States

Citation: Covshoff, S., Ludwig, M., Busch, F. A., Khoshravesh, R., eds. (2021). Understanding C4 Evolution and Function. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-864-1

Table of Contents

frontiers in Plant Science

EDITORIAL
published: 22 October 2021
doi: 10.3389/fpls.2021.774818

Editorial: Understanding C_4 Evolution and Function

Martha Ludwig[1], *Florian A. Busch*[2,3], *Roxana Khoshravesh*[4] *and Sarah Covshoff*[5]*

[1] School of Molecular Sciences, University of Western Australia, Crawley, WA, Australia, [2] School of Biosciences and Birmingham Institute of Forest Research, University of Birmingham, Birmingham, United Kingdom, [3] Research School of Biology, Australian National University, Canberra, ACT, Australia, [4] Department of Biology, University of New Mexico, Albuquerque, MN, United States, [5] Independent Researcher, Las Vegas, NV, United States

Keywords: C_4 photosynthesis, convergent evolution, comparative biology, biodiversity, regulatory mechanisms

Editorial on the Research Topic

Understanding C_4 Evolution and Function

C_4 photosynthesis is a remarkable example of convergent evolution, having independently evolved at least 62 times over the last 60 million years (Sage et al., 2011). In C_4 species, Rubisco operates close to its maximal carboxylation rate through suppression of the oxygenation reaction. This activity is accomplished via the establishment of a molecular CO_2 pump that delivers carbon in the form of C_4 acid intermediates to a spatially sequestered Rubisco. This carbon pump can be set up using a diverse array of complex biochemical and morphological modifications relative to the ancestral C_3 photosynthetic state.

The large number of independent origins of a C_4 syndrome suggests that evolution from ancestral C_3 photosynthesis to a derived C_4 type is flexible at the molecular level and relatively easy in genetic terms (Gowik et al., 2004; Williams et al., 2013; Heckmann, 2016). With a large pool of biodiversity to exploit, such as in Southwest Asia, reviewed here by Rudov et al., natural variation in diverse phylogenetic lineages can be used to better understand the molecular changes enabling evolution of a functional C_4 syndrome. The papers presented in this Research Topic make use of this biodiversity to expand our knowledge of C_4 evolution and function.

Despite C_4 photosynthesis being highly convergent, little work has been done to understand which C_4 traits have arisen through convergence and could be considered essential for a C_4 syndrome. Here, Khoshravesh et al. use gas exchange, leaf ultrastructure and biochemistry and carbon isotope ratios to characterize the carbon assimilation pathways used by species in the eudicot family Nyctaginaceae, and in the case of the C_4 members, to determine the subtype of C_4 photosynthesis. Combining these data with those from other C_4 clades, they compiled a hierarchical list of convergent and divergent traits.

Gene duplication has been proposed as one of the early steps in the recruitment of genes during evolution of a C_4 pathway (Monson, 2003). Tronconi et al. describe a complex evolutionary history responsible for present-day C_4-associated NAD-malic enzyme (NAD-ME) in the Brassicales that involves ancestral gene duplication followed by degeneration, complementation subfunctionalization, and neofunctionalization. Gene duplication and co-option also appear to be responsible for the evolution of the C_4-associated PEP transporter, PPT1. Lyu et al. identify differences in coding and non-coding regions between C_3 and C_4 orthologs of PPT1 associated with increased expression of the transporter in C_4 mesophyll cells (MC). They find that gene duplication and neo-functionalization led to recruitment of a PPT1 paralog found in roots to a role in C_4 function.

Most C_4 species operate a carbon pump with the help of Kranz anatomy, wherein MC surround highly specialized bundle sheath cells (BSC) that are concentrically arranged around the vasculature

OPEN ACCESS

Edited and reviewed by:
Carl J. Rothfels, University of California, Berkeley, United States

***Correspondence:**
Sarah Covshoff
sarahcovshoff@gmail.com

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *13 September 2021*
Accepted: *29 September 2021*
Published: *22 October 2021*

Citation:
Ludwig M, Busch FA, Khoshravesh R and Covshoff S (2021) Editorial: Understanding C_4 Evolution and Function. Front. Plant Sci. 12:774818. doi: 10.3389/fpls.2021.774818

(Sage et al., 2014). In a small number of species, special organellar arrangements within a single cell are used to achieve high CO_2 concentrations around Rubisco (Sharpe and Offermann, 2014). In work on *Suaeda aralocaspica*, a single-cell C_4 species, Cao et al. identify paralogs encoding the C_4-associated phosphoenolpyruvate carboxylase (PEPC), which catalyzes the first step in the C_4 pathway, a housekeeping isoform, and a bacterial-type PEPC.

Given the apparent flexibility of gene recruitment during evolution of C_4 syndromes, identification of regulatory components controlling the spatial expression of C_4-associated enzymes is important for understanding C_4 function. Here, Afamefule and Raines use C_3 and C_4 grasses to screen upstream regions of genes encoding four enzymes in the Calvin-Benson-Bassham (CBB) cycle for conserved nucleotide sequences that might enable cell-preferential expression. They identify *cis*-regulatory elements putatively involved in BSC-enriched expression of genes encoding CBB enzymes as well as candidate transcription factors potentially binding to those sites. In addition, Górska et al. identify three *trans*-acting factors that bind the upstream region of the C_4-associated PEPC homolog of maize. Characterization of these factors highlights the complexity of cell-preferential expression in a C_4 leaf and the role of repression in establishing some C_4-type expression patterns.

Of course, evolution is ongoing. As suggested by the results of Moody et al. in a study on PEPCs from older and younger C_4 lineages, optimization of the enzyme continues after a C_4 syndrome is realized. Similar comparative studies of other enzymes in the C_4 acid cycle may also contribute to our understanding of how a C_4 syndrome evolves at the molecular level.

A better understanding of the molecular events underpinning evolution of a C_4 syndrome could enable a C_3 plant to be engineered for C_4 features. This is highly desirable because C_4 crops have higher yields and increased nitrogen and water use efficiency relative to C_3 crops. Replicating the C_4 process in C_3 crops such as rice would therefore help feed a growing world population. Support for introducing a C_4 pathway into rice is provided by Lin et al. Genes encoding four of the major enzymes in the maize NADP-ME-type C_4 pathway, PEPC, NADP-malate dehydrogenase (NADP-MDH), NADP-ME and pyruvate phosphate dikinase (PPDK), were inserted into the rice genome. Subsequent measurements with $^{13}CO_2$ demonstrate that production of ^{13}C-labeled malate was high in the transformants, suggesting that a partial C_4 pathway is functioning in these plants.

Studies on C_4 physiology and metabolism are also important to improve breeding programs of C_4 crops. In particular, light harvesting and nutrient availability and uptake are key determinants for crop productivity. Collison et al. explore relationships between leaf age and light availability with the phenomenon of shade maladaptation exhibited by the NADP-ME-type C_4 crops maize, sorghum and sugarcane. Leaf age had little influence on the quantum yield of CO_2 assimilation. Instead, optimization of the leaf light environment mitigates the negative effects on productivity associated with this maladaptive response. These results can inform breeding strategies related to canopy structure and agricultural practices such as planting densities to increase crop yield.

Jobe et al. highlight the need to consider the nutritional value of C_4 crops in addition to yield. They review nutrient assimilation pathways in C_4 plants and how they differ from C_3 plants as well as discuss gaps in our knowledge of how nutrient uptake and levels are controlled in C_4 plants. They also consider the effects of increasing atmospheric CO_2 on C_3 and C_4 crop micronutrient assimilation and content in light of micronutrient-related malnutrition (i.e., hidden hunger). Such considerations are important for producing future C_4 crops that will effectively address global food needs.

In summary, this collection of articles expands our understanding of C_4 evolution and function. This new knowledge will inform future work in evolutionary biology, C_4 metabolism, and crop improvement strategies.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

Gowik, U., Burscheidt, J., Akyildiz, M., Schlue, U., Koczor, M., Streubel, M., et al. (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. doi: 10.1105/tpc.019729

Heckmann, D. (2016). C_4 photosynthesis evolution: the conditional Mt. Fuji. *Curr Opin Plant Biol*. 31, 149–154. doi: 10.1016/j.pbi.2016.04.008

Monson, R. (2003). Gene duplication, neofunctionalization, and the evolution of C_4 photosynthesis. *Int. J. Plant Sci*. 164, S43–S54. doi: 10.1086/368400

Sage, R. F., Christin, P. A., and Edwards, E. J. (2011). The C_4 plant lineages of planet Earth. *J. Exp. Bot*. 62, 3155–3169. doi: 10.1093/jxb/err048

Sage, R. F., Khoshravesh, R., and Sage, T. L. (2014). From proto-Kranz to C_4 Kranz: building the bridge to C_4 photosynthesis. *J. Exp. Bot*. 65, 3341–3356. doi: 10.1093/jxb/eru180

Sharpe, R. M., and Offermann, S. (2014). One decade after the discovery of single-cell C_4 species in terrestrial plants: what did we learn about the minimal requirements of C_4 photosynthesis? *Photosynth. Res*. 119, 169–180. doi: 10.1007/s11120-013-9810-9

Williams, B. P., Johnston, I. G., Covshoff, S., and Hibberd, J. M. (2013). Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis. *Elife* 2:e00961. doi: 10.7554/eLife.00961

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 24 June 2020
doi: 10.3389/fpls.2020.00783

Light, Not Age, Underlies the Maladaptation of Maize and Miscanthus Photosynthesis to Self-Shading

*Robert F. Collison[1], Emma C. Raven[1], Charles P. Pignon[2,3,4] and Stephen P. Long[2,3,4,5]**

[1] Department of Plant Sciences, University of Oxford, Oxford, United Kingdom, [2] Carl R. Woese Institute for Genomic Biology, University of Illinois, Urbana, IL, United States, [3] Department of Crop Sciences, University of Illinois, Urbana, IL, United States, [4] Department of Plant Biology, University of Illinois, Urbana, IL, United States, [5] Lancaster Environment Centre, Lancaster University, Lancaster, United Kingdom

Zea mays and *Miscanthus* × *giganteus* use NADP-ME subtype C_4 photosynthesis and are important food and biomass crops, respectively. Both crops are grown in dense stands where shaded leaves can contribute a significant proportion of overall canopy productivity. This is because shaded leaves, despite intercepting little light, typically process light energy very efficiently for photosynthesis, when compared to light-saturated leaves at the top of the canopy. However, an apparently maladaptive loss in photosynthetic light-use efficiency as leaves become shaded has been shown to reduce productivity in these two species. It is unclear whether this is due to leaf aging or progressive shading from leaves forming above. This was resolved here by analysing photosynthesis in leaves of the same chronological age in the centre and exposed southern edge of field plots of these crops. Photosynthetic light-response curves were used to assess maximum quantum yield of photosynthesis; the key measure of photosynthetic capacity of a leaf in shade. Compared to the upper canopy, maximum quantum yield of photosynthesis of lower canopy leaves was significantly reduced in the plot centre; but increased slightly at the plot edge. This indicates loss of efficiency of shaded leaves is due not to aging, but to the altered light environment of the lower canopy, i.e., reduced light intensity and/or altered spectral composition. This work expands knowledge of the cause of this maladaptive shade response, which limits productivity of some of the world's most important crops.

Keywords: C_4 photosynthesis, canopy, bioenergy, food security, quantum yield, shade acclimation, photosynthetic light-use efficiency, leaf aging

OPEN ACCESS

Edited by:
Florian A. Busch, Australian National University, Australia

Reviewed by:
Paulo Eduardo Ribeiro Marchiori, Universidade Federal de Lavras, Brazil
Elias Kaiser, Wageningen University & Research, Netherlands

***Correspondence:**
Stephen P. Long
slong@illinois.edu

Specialty section:
This article was submitted to Plant Physiology, a section of the journal Frontiers in Plant Science

Received: *19 March 2020*
Accepted: *18 May 2020*
Published: *24 June 2020*

Citation:
Collison RF, Raven EC, Pignon CP and Long SP (2020) Light, Not Age, Underlies the Maladaptation of Maize and Miscanthus Photosynthesis to Self-Shading. Front. Plant Sci. 11:783. doi: 10.3389/fpls.2020.00783

INTRODUCTION

C_4 grasses of the Andropogoneae represent some of the most important cultivated plants on the planet, making up a significant proportion of our food and fibre production, as well as providing major bioenergy crops. All members of this monophyletic tribe use the NADP-ME subtype of C_4 photosynthesis, with some species using substantial PCK activity. This tribe includes crops such as *Saccharum officinarum* L. (sugarcane), the greatest producer of harvested biomass globally, and *Zea mays* L. (maize), the single largest source of grain

globally (Christin et al., 2009; Welker et al., 2014; FAOSTAT, 2017). Other C_4 NADP-ME crops of this tribe are highly productive in the face of extreme climatic conditions, and thus vital to food production in drought prone environments. *Sorghum bicolor* (Lu.) Moench (sorghum), for instance, is the second most extensively cultivated crop plant in Africa behind *Z. mays* thanks to its high drought tolerance (FAOSTAT, 2017; Hadebe et al., 2017). The tribe also includes the most productive temperate biomass crop known, *Miscanthus* × *giganteus* Greef et Deu. (Heaton et al., 2008; LeBauer et al., 2018).

The theoretical maximum efficiency of conversion of solar energy to biomass is 6% for C_4 compared to 4.6% for C_3 photosynthesis at 30°C and 380 ppm atmospheric CO_2: this improved photosynthetic light-use efficiency contributes to higher yields in C_4 crops (Zhu et al., 2008). The key metric for photosynthetic light-use efficiency is the quantum yield of CO_2 assimilation, i.e., the mol CO_2 assimilated per mol photons of light. In a typical light-response curve, the quantum yield of CO_2 assimilation is greatest when light is limiting, and declines at high light as photosynthesis becomes light-saturated. The maximum quantum yield of CO_2 assimilation ($\phi CO_{2\ max,app}$), achieved under limiting light, is therefore paramount for the productivity of shade leaves. Shade leaves are estimated to contribute around 50% of total canopy carbon gain in field crops and may represent >80% of leaves in a dense crop stand (Baker et al., 1988; Long, 1993; Hikosaka et al., 2016). Accordingly, leaves of most plants respond to increasing shade by maintaining or increasing $\phi CO_{2\ max,app}$ so that they can make maximum use of the limited light. However, in *Z. mays* and *M.* × *giganteus* a significant decrease in $\phi CO_{2\ max,app}$ has been observed in leaves as they become progressively shaded by new leaves forming above them, with a projected cost of up to 10% of potential canopy CO_2 assimilation (Pignon et al., 2017). With the continued trend of increasing planting density this loss will likely increase into the future (Lobell et al., 2014).

Shade acclimation in C_4 species has been studied primarily by comparing plants grown in high vs. low light (Tazoe et al., 2008; Sales et al., 2018; Sonawane et al., 2018). On this basis, it has been observed that C_4 species have relatively poor acclimation to shade relative to C_3 species (Sage and McKown, 2006), but C_4 grasses which use the NADP-ME subtype, such as *Z. mays*, acclimate to shade more readily than those using NAD-ME or PEP-CK subtypes (Sonawane et al., 2018). However, in these studies the shaded leaves grow while the entire plant is shaded, such that their entire development occurs in the shade. In crop fields, leaves form in full sunlight, but then become progressively shaded after they have completed development as new leaves form above them (Yabiku and Ueno, 2019). Less is known about acclimation in this situation, which is particularly relevant to crop productivity. Plasticity to shade in this context is more limited, since leaves are already fully formed and acclimated to high light before becoming shaded. In grasses, plasticity of key physiological traits, such as leaf nitrogen (N) content, declines with increasing leaf age (Niinemets, 2016a). In addition, shade in the lower canopy is not simply reduced light quantity, but also altered spectral light composition, with relative depletion of red and blue, and enrichment of green and near infrared, plus an increased incidence of light fluctuations due to sun flecks (Pearcy, 1990). Leaves of NADP-ME C_4 grasses lose photosynthetic efficiency under these conditions (Kromdijk et al., 2008; Kubasek et al., 2013; Pignon et al., 2017).

The two major distinctions between a sun and shade leaf in a C_4 grass canopy are leaf age and light environment. Understanding whether the decline of photosynthetic efficiency in shade leaves results from age, light environment, or both, is an important first step in devising strategies to overcome this costly maladaptation in these key crops. For instance, efforts to optimize canopy architecture have involved producing crops with more erect (Perez et al., 2018; San et al., 2018) or more transparent (Slattery et al., 2016; Walker et al., 2018) leaves that increase light availability at the bottom of the canopy to increase canopy photosynthesis (Zhu et al., 2010). This strategy may not be as effective in C_4 grass canopies if the leaves at the bottom of the canopy have lost efficiency in low light due to age, and so have limited ability to utilize the increased levels of *PPFD* enabled by these canopy alterations.

Classically, leaf shade adaptation involves maintaining maximum quantum yields on an absorbed light basis ($\phi CO_{2\ max,abs}$), and increasing leaf light absorbance (α) through increased chlorophyll concentration, to deliver increased photosynthesis in the shade. However, prior evidence has shown the reverse to occur in *Z. mays* and *M. x giganteus,* with a decrease in $\phi CO_{2\ max,abs}$ and significant cost to canopy photosynthesis (Pignon et al., 2017). Here, we tested the following hypothesis: chronological age is responsible for the loss of maximum quantum yields of photosynthesis in field plots of the C_4 NADP-ME grasses *Z. mays* and *M. x giganteus*. Leaves were collected from the top and bottom of the canopy at the south exposed edge and at the centre of field plots of these crops, such that lower canopy leaves from both plot positions were of the same chronological age, but only those at the plot centre were shaded. This enabled separation of the effects of environment and chronological age on differences in photosynthetic efficiency between sun and shade leaves in a field production setting. The maximum quantum yield of CO_2 assimilation, and its underlying physiological drivers, were determined from leaf gas exchange, modulated chlorophyll fluorescence and light absorbance measurements.

MATERIALS AND METHODS

Plant Material

Measurements were taken on *Zea mays* and *Miscanthus* × *giganteus*. Leaves were collected from >1 ha plots of a high-yielding modern *Z. mays* hybrid as described previously (Pignon et al., 2017) on the University of Illinois South Farms (40°02′N, 88°14′W, 216 m above sea level), and leaves of *M.* × *giganteus* ("Illinois" clone) from 4 ha plots on University of Illinois Energy Farm (40°07′N, 5 88°21′W, 228 m above sea level) as described previously (Joo et al., 2017). Soils at these sites are deep Drummer/Flanagan series (a fine silty, mixed, mesic Typic Endoaquoll) with high organic matter typical of the central Illinois region of the Corn Belt (Smith et al., 2013). Both

plots were rainfed. The *M.* × *giganteus* plots were 9 years old, with a stem density of about 100 tillers m^{-2}; these plots were unfertilized. *Z. mays* was sown in early May at a density of 75,000 seeds ha^{-1}. Prior to planting, 140 kg [N] ha^{-1} was applied, in line with regional production practice.

Measurements were taken between July 26 and August 06 of 2018. Leaves were cut pre-dawn at the base, then the base was submerged in water and re-cut to prevent air blockage in the xylem as described in Pignon et al. (2017). Removing leaves from plants in this way has been shown not to bias photosynthetic measurements (Leakey et al., 2006). Leaves were then brought back to the laboratory, where they remained in low light until measurement. This procedure avoided any photoinhibition or transient water stress that could develop differentially in shade and sun leaves over a day.

Leaves were sampled from two canopy positions (upper and lower) and two plot positions, centre and the south edge. For each plant sampled, two leaves were collected; an upper canopy leaf, defined as the youngest fully expanded leaf, indicated by a fully emerged ligule, and a lower canopy leaf; the seventh counting down from the first fully emerged leaf. This ensured that within a species and canopy position, leaves from the plot centre and edge were of the same age. The lower canopy leaves in the plot centre were strongly shaded, whereas lower canopy leaves at the plot's edge were not. The south edge of the plot was chosen since on clear sky days these leaves were exposed to sunlight for 12 h per day.

Measurement of Photosynthesis

Portable photosynthetic gas exchange systems (LI 6400 and LI 6400-40 modulated fluorescence chamber head; LI-COR, Inc., Lincoln, NE, United States) were used to measure CO_2 and water vapor exchange on a 2 cm^2 area of each leaf, along with modulated chlorophyll fluorescence, in the system's controlled environment leaf cuvette. Air temperature was controlled at a constant 25.0°C, chamber [CO_2] at 400 ppm, and water vapour pressure deficit at 1.6–2.4 kPa.

The measurement sequence began with estimation of maximum dark-adapted quantum yield of PSII photochemistry (F_v/F_m). A photosynthetic light response curve was generated as follows: integrated LEDs emitted uniform light consisting of 10% blue (465 nm wavelength) and 90% red (635 nm wavelength) across the leaf surface. In order to limit photoinhibition caused by sudden exposure to saturating light on enclosure in the cuvette, leaves were first subjected to a photosynthetic photon flux density (*PPFD*) of 100 μmol m^{-2} s^{-1} for 5 min, and subsequently exposed to 2000 μmol m^{-2} s^{-1} for 30–60 min until *A* reached a steady-state. *PPFD* was then decreased from 2000 in steps to 1500, 1000, 500, 200, 180, 160, 140, 120, 100, 80, 60, 40, 20, and 0 μmol m^{-2} s^{-1}. Each *PPFD* step lasted 5–10 min to allow *A* to reach a steady state before measuring. Steady-state gas-exchange was recorded at each level of *PPFD* and used to calculate *A* (von Caemmerer and Farquhar, 1981). Modulated fluorescence measurements were made at each level of *PPFD* to determine the operating quantum yield of PSII (ϕ_{PSII}) using a multiphase flash protocol (Loriaux et al., 2013). In turn, ϕ_{PSII} was used to calculate the rate of linear electron flux through *PSII* (*J*), using measured values for leaf fractional absorptance of photosynthetically active photon flux (α, described below) and assuming a photon partitioning factor of 0.4 for PSII vs. PSI, i.e., accounting for increased photon partitioning to PSI to produce ATP through cyclic electron flux (Yin and Struik, 2012; Ver Sagun et al., 2019). Each *A-PPFD* response curve was fit to a four-parameter non-rectangular hyperbola using PROC NLIN (SAS v9.4, SAS Institute, Cary, NC, United States), which produced an asymptote, taken to represent light-saturated *A* (A_{sat}), and a Y-intercept, taken to represent dark respiration (R_d). The third parameter described light-limited *A* and the fourth parameter described the inflexion between light-limited and light-saturated *A* with increasing *PPFD*.

After gas-exchange measurements were completed, absorptance (α) was measured using an integrating sphere and associated spectrometer (Jaz-Spectroclip-TR, Ocean Optics, Largo, FL, United States) and operating software (Spectrasuite, Ocean Optics). α was weighted for 10% blue (465 nm wavelength) and 90% red (635 nm wavelength) incident light to match illumination in the gas-exchange chamber.

The maximum quantum yield of CO_2 assimilation on an incident light basis ($\phi CO_{2\ max,app}$) was calculated from the slope of the linear regression of *A* against *PPFD* from 40 to 140 μmol m^{-2} s^{-1} using PROC GLM (SAS v9.4) (Yin et al., 2014; Pignon et al., 2017). This interval was chosen to account for the Kok effect where respiration increases at very low light levels ($PPFD < 40$ μmol m^{-2} s^{-1}), and to avoid high light levels ($PPFD > 140$ μmol m^{-2} s^{-1}) where *A* is no longer strictly light-limited causing deviations from the linear relationship of *A* and *PPFD*. The maximum quantum yield of CO_2 assimilation on an absorbed light basis ($\phi CO_{2\ max,abs}$) was given by $\phi CO_{2\ max,app}/\alpha$. Finally, the maximum quantum yield of CO_2 assimilation on an absorbed light basis and corrected for concurrent changes in $\phi PSII$ ($\phi CO_{2\ max,abs\ PSII}$) was calculated as in Yin et al. (2014). To test for alternative electron sinks to photosynthetic carbon metabolism, the slope of *A* vs. *J* was calculated for *PPFD* between 40 and 140 μmol m^{-2} s^{-1} using linear regression (SAS v9.4). The slope of this relationship gives the mol CO_2 assimilated per mol electrons in linear electron flux ($1/k$) (Baker, 2008). Here *k* is the mol electrons through linear electron flux required for photosynthesis to fix one mol CO_2. $1/k$ is an indicator of alternative energy sinks, where any reduction in $1/k$ is assumed to result from alternative energy sinks, including utilization of ATP and NADPH in processes other than photosynthetic carbon metabolism.

Statistical Analysis

Data were analysed by ANOVA using PROC GLM (SAS v9.4), testing for the fixed effect of species (S effect: *Z. mays* vs. *M.* x *giganteus*), the fixed effect of canopy position (C effect: upper vs. lower canopy), and the fixed effect of plot position (P effect: centre vs. edge), along with all two-way interactions (S x P, S x C, P x C). This model was used to test for significant ($p = 0.05$ threshold) and marginally significant ($p = 0.1$ threshold) differences in the following traits: $\phi CO_{2\ max,app}$, $\phi CO_{2\ max,abs}$, $\phi CO_{2\ max,abs\ PSII}$, $1/k$, α, A_{sat}, R_d, and F_v/F_m. Homogeneity of group variances was tested by Levene's at $p = 0.05$ threshold

in PROC GLM (SAS v9.4). Normality of Studentized residual distribution was tested by Shapiro–Wilk at $p = 0.01$ threshold in PROC UNIVARIATE (SAS v9.4). Replication was $n = 8–16$ in different traits.

RESULTS

In this study, the three key measures of photosynthetic efficiency ($\phi CO_{2\ max,app}$, $\phi CO_{2\ max,abs}$ and $\phi CO_{2\ max,abs\ PSII}$) all derive from the linear slope, at low *PPFD*, of the *A-PPFD* response curve (**Figure 1**). There was a significant interaction ($p < 0.05$) between canopy position and plot position for all three of these metrics (**Figure 2A**: P x C interaction, **Supplementary Table S1**: P x C interaction). This was because photosynthetic efficiency was greater at the top than the bottom of the canopy at the plot centre, while the opposite was seen at the plot edge where photosynthetic efficiency was slightly lower at the top than at the bottom of the canopy. Indeed, at the plot centre lower canopy leaves of both *Z. mays* and *M.* x *giganteus* showed a 2–18% reduction across all measures of photosynthetic efficiency compared to the upper canopy leaves ($\phi CO_{2\ max,app}$, $\phi CO_{2\ max,abs}$, $\phi CO_{2\ max,abs\ PSII}$, **Figure 2A** and **Supplementary Table S1**). In contrast, at the edge of the plots, the lower canopy leaves for both *Z. mays* and *M.* x *giganteus* showed 2–9% greater efficiency than the upper canopy leaves for the same measurements. In addition, *Z. mays* recorded significantly ($p < 0.0001$) and up to 43% greater values than *M.* x *giganteus* for $\phi CO_{2\ max,app}$, $\phi CO_{2\ max,abs}$, and $\phi CO_{2\ max,abs\ PSII}$ (**Figure 2A**: S effect and **Supplementary Table S1**: S effect). Finally, $\phi CO_{2\ max,abs}$ at the plot edge was

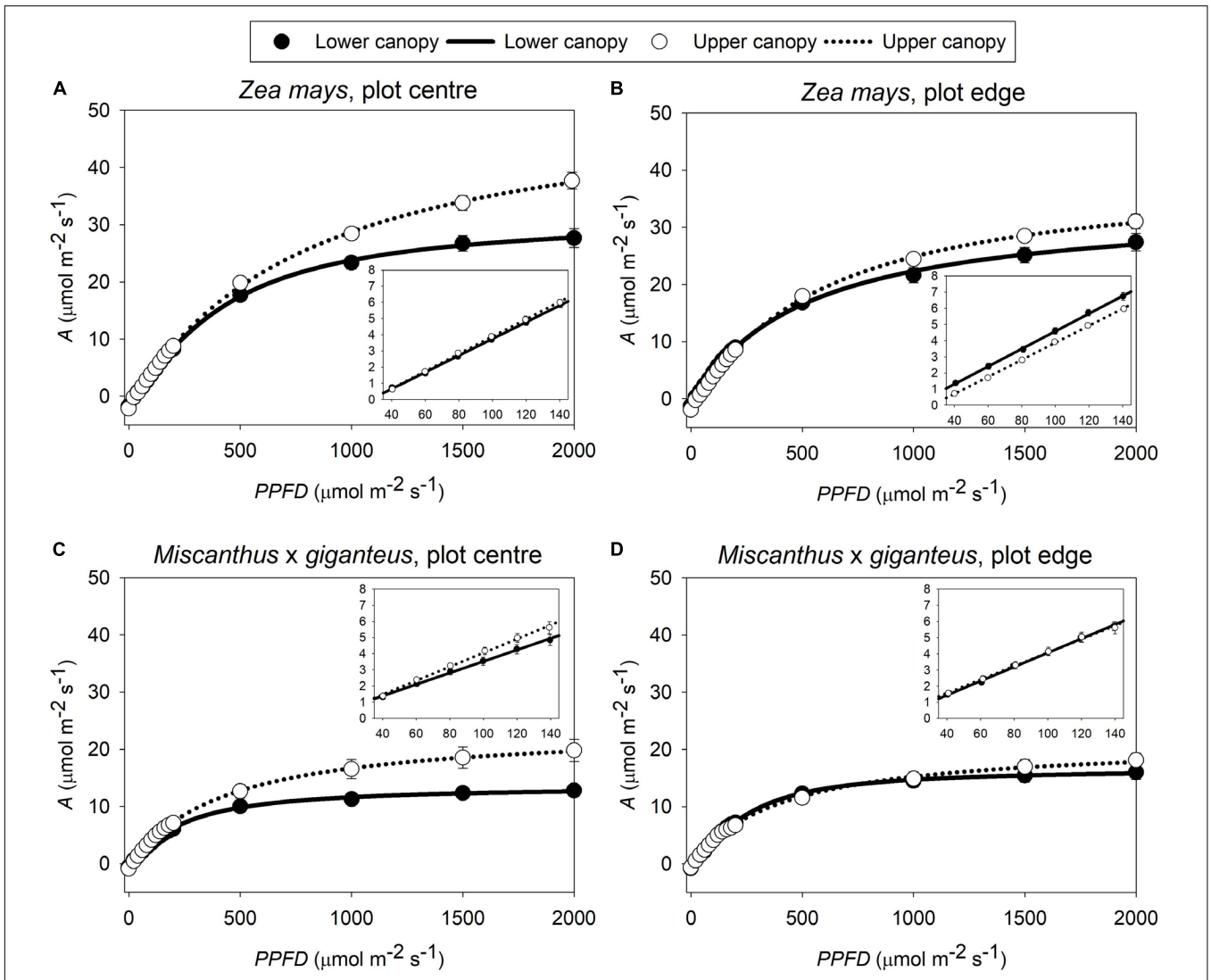

FIGURE 1 | Average response curves of *A* to *PPFD* in both canopy positions (upper, lower) of **(A)** *Z. mays* in plot centre and **(B)** plot edge; and **(C)** *M.* x *giganteus* in plot centre and **(D)** plot edge ($n = 8–16$). Symbols give the mean *A* ± s.e. at each level of *PPFD*. Lines give the non-rectangular hyperbolae fit to these measurements. In each panel, the inset shows the light-limited section of the response curve (*PPFD* from 40 to 140 μmol m^{-2} s^{-1}), used to estimate maximum quantum yields by linear regression: the slope of this regression gives the trait $\phi CO_{2\ max,app}$.

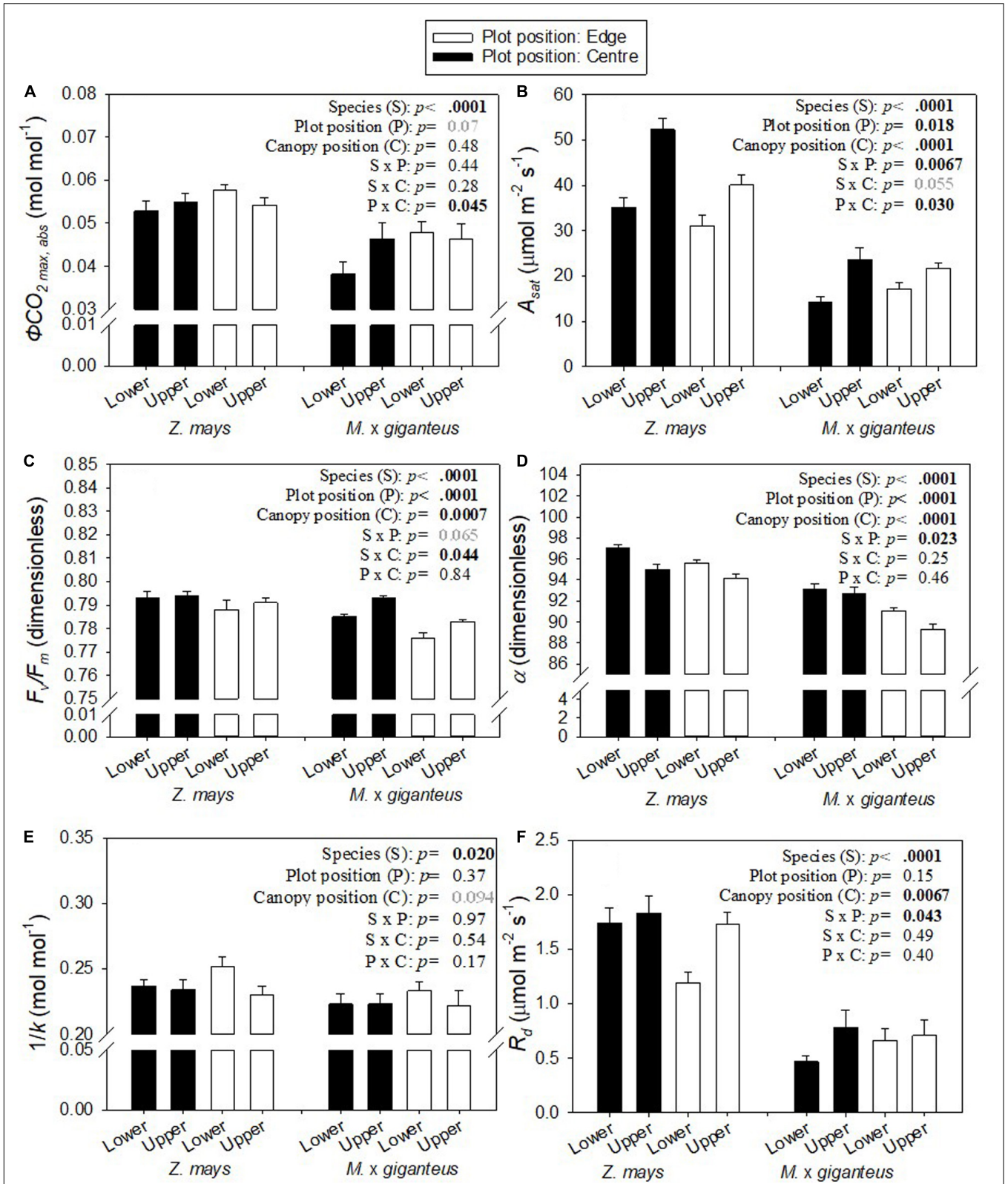

FIGURE 2 | Mean ± s.e. for **(A)** $\phi CO_{2\ max,abs}$, **(B)** A_{sat}, **(C)** F_v/F_m, **(D)** α, **(E)** 1/*k*, and **(F)** R_d for *Z. mays* and *M.* x *giganteus* for upper and lower canopy leaves in both plot positions (centre, edge) (n = 8–16). p-values are from ANOVA testing the fixed effects of species, plot position, canopy position, and all two-way interactions. Significant p-values (< 0.05) are in bold black. Marginally significant p-values (< 0.1) are in bold grey.

marginally significantly ($p = 0.07$) and up to 25% greater than at the plot centre (**Figure 2A**: P effect).

The only other measure that showed a statistically significant interaction of plot and canopy position was A_{sat}. A_{sat} was significantly ($p < 0.0001$) and up to 2-fold greater in *Z. mays* than *M.* x *giganteus* (**Figure 2B**, S effect), significantly ($p = 0.018$) greater at the centre than at the edge of the plot (**Figure 2B**, P effect), and significantly ($p < 0.0001$) greater at the top than at the bottom of the canopy (**Figure 2B**, C effect). The difference in A_{sat} between canopy levels was more pronounced at the centre than at the edge of plots, leading to a significant interaction ($p = 0.03$) of canopy position and plot position (**Figure 2B**: P x C interaction). Relative to the upper canopy, A_{sat} was decreased in the lower canopy by 30 and 40% in *Z. mays* and *M.* x *giganteus*, respectively, in the centre of the plots and by 23 and 21% in *Z. mays* and *M.* x *giganteus*, respectively, at the edge of the plots. A_{sat} showed significant interaction ($p = 0.0067$) of species and plot position (**Figure 2B**: S x P interaction), and a marginally significant interaction ($p = 0.055$) of species and canopy position (**Figure 2B**: S x C interaction). This was because differences in A_{sat} between canopy positions and between plot positions were more pronounced in *Z. mays* than in *M.* x *giganteus*.

There were statistically significant ($p = 0.0007$) decreases in F_v/F_m in the lower canopy relative to the upper canopy, but in absolute terms this was a minor difference at less than 1% (**Figure 2C**: C effect). There were similarly small, but significant ($p < 0.0001$), decreases in F_v/F_m at the edge relative to the centre (**Figure 2C**: P effect), and in *M.* x *giganteus* relative to *Z. mays* (**Figure 2C**: S effect). Differences in F_v/F_m between canopy positions and between plot positions were slightly more pronounced in *M.* x *giganteus* than in *Z. mays*, resulting in a significant interaction ($p = 0.044$) of species and canopy position (**Figure 2C**: S x C interaction), and a marginally significant interaction ($p = 0.065$) of species and plot position (**Figure 2C**: S x P interaction).

Lower canopy leaves had significantly ($p < 0.0001$) and up to 2% greater α than upper canopy leaves (**Figure 2D**: C effect). α was also significantly ($p < 0.0001$) lower in *M.* x *giganteus* in comparison to *Z. mays* (**Figure 2D**: S effect) and significantly ($p < 0.0001$) greater at the plot centre than at the edge (**Figure 2D**: P effect). There was a significant interaction ($p = 0.023$) of species with plot position (**Figure 2F**: S x P interaction) because the difference in α between species was 5% at the edge of the plots and only 3% in the centre of the plots.

1/k, i.e., the ratio of *A* to *J* (**Figure 3**), was marginally significantly ($p = 0.094$), and up to 9% greater in lower canopy leaves than upper canopy leaves (**Figure 2E**: C effect). *1/k* was also significantly ($p = 0.02$) and 4–8% greater in *Z. mays* than in *M.* x *giganteus* (**Figure 2E**: S effect).

R_d was significantly ($p = 0.0067$) and 5–66% greater in upper canopy leaves than lower canopy leaves (**Figure 2F**: C effect). In line with the higher A_{sat}, R_d was also significantly ($p < 0.0001$) greater in *Z. mays* than in *M.* x *giganteus* (**Figure 2F**: S effect). The difference in R_d between species was less pronounced at the edge than at the centre of the plots, resulting in a significant ($p = 0.043$) interaction between species and plot position (**Figure 2F**: S x P interaction). There was 144 and 80% difference between species for upper and lower canopy leaves at the plot edge, respectively, compared to 135 and 270% difference between species for upper and lower canopy leaves at the plot centre.

DISCUSSION

Reduced Maximum Quantum Yield of CO_2 Assimilation Is Not Caused by Increased Leaf Age

In a self-shading crop canopy, the optimal response to shade would be to maintain or increase $\phi CO_{2\ max,abs}$ and increase α in order to maximize photosynthesis in light limited conditions. This would increase the linear slope of the response of *A* to *PPFD* at low light. This response is observed in shade adapted C_3 plants and in C_3 cereal crops (Givnish, 1988; Beyschlag et al., 1990; Hoyaux et al., 2008). However, the two C_4 crops *Z. mays* and *M.* x *giganteus* studied here, show decreased $\phi CO_{2\ max,abs}$ in the lower canopy at the plot centre, but not at the plot edge. This suggests that loss of $\phi CO_{2\ max,abs}$ in shade leaves was not due to leaf age, since leaf age was equivalent across plot positions for each species and canopy position. Understanding the basis for this maladaptive response in photosynthetic efficiency is important, as it costs an estimated 10% of potential canopy CO_2 assimilation in the field (Pignon et al., 2017).

If not age, then some environmental factor must trigger the decline in $\phi CO_{2\ max,abs}$ of these shaded leaves. The most obvious environmental change between the top and bottom of the canopy is the light environment, with lower leaves receiving less light and an altered spectral distribution, depleted of red and blue and enriched in far-red wavelengths (Sattin et al., 1994). The hypothesis that self-shading is the primary cause for the loss of $\phi CO_{2\ max,abs}$ in shade leaves of these C_4 NADP-ME crops is supported by the following observations: (1) when comparing both studied species, the loss of $\phi CO_{2\ max,abs}$ in shade leaves at the plot centre was more pronounced in *M.* x *giganteus*, which produces a denser canopy with considerably more self-shading than *Z. mays*. Profiles of canopy light interception in field stands of both species show that the lowest photosynthetically active leaves of *Z. mays* receive as much as twice the incident *PPFD* compared to equivalent leaves in *M.* x *giganteus* (Pignon et al., 2017). (2) In a previous study comparing two field-grown sugarcane varieties with high and low self-shading, photosynthetic light response curves measured at the top and bottom of the canopy produced contrasting results in the response of *A* to *PPFD* at low *PPFD* (<500 μmol m^{-2} s^{-1}) (Marchiori et al., 2014). In the low self-shading variety, *A* at low *PPFD* was slightly greater at the bottom than at the top of the canopy, while the opposite was seen in the high self-shading variety. These studies implemented shade acclimation under realistic field conditions, which produce different results than artificial shading including altered spectral light composition and increased incidence of sun and shade flecks (Pearcy, 1990; Bellasio and Griffiths, 2014; Yabiku and Ueno, 2019).

These findings are important in light of recent efforts to develop crops with a more even vertical light distribution,

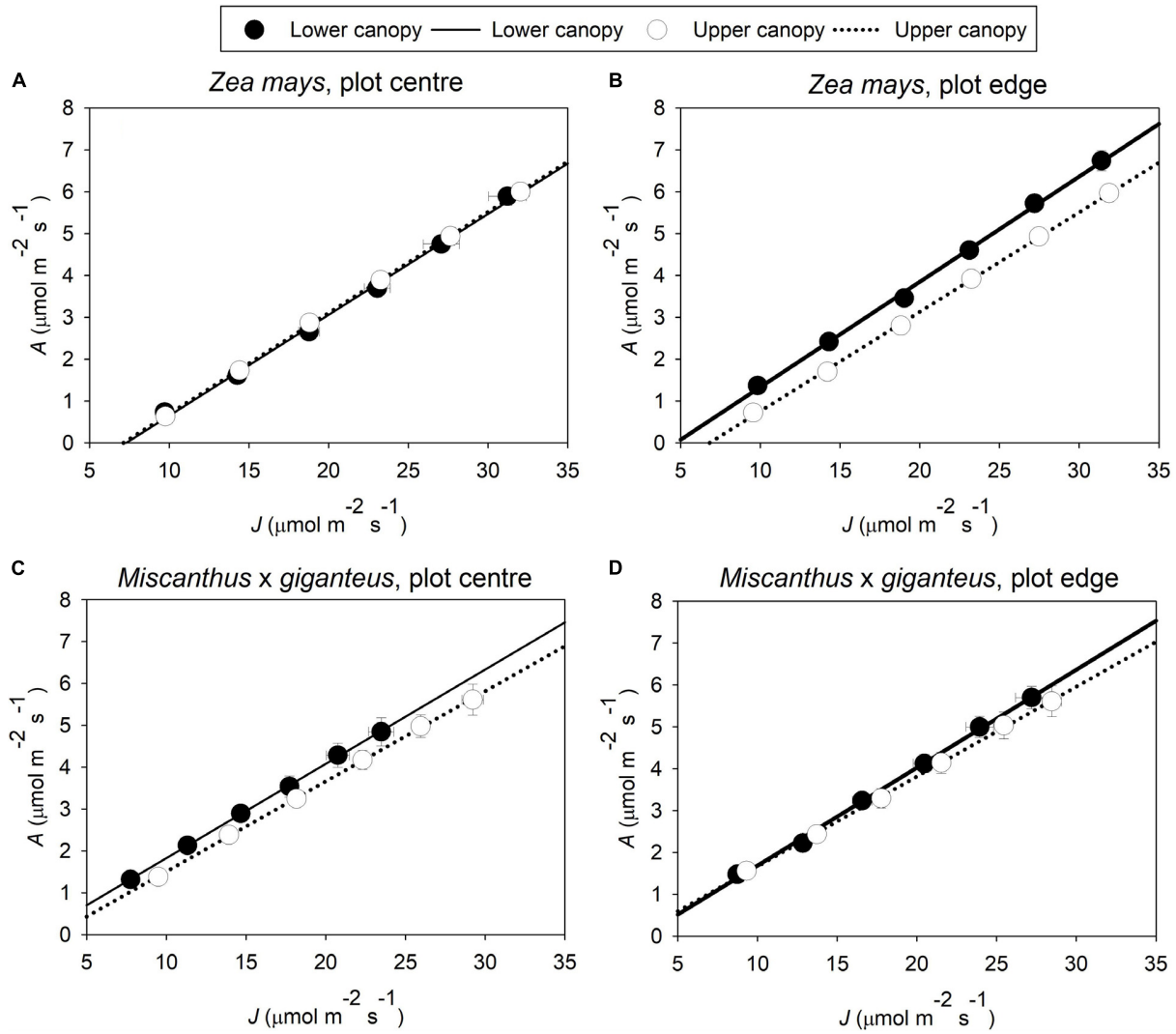

FIGURE 3 | Average linear responses of *A* to *J* in **(A)** *Z. mays* in plot centre and **(B)** plot edge; and **(C)** *M.* x *giganteus* in plot centre and **(D)** plot edge (n = 8–16). Data is from light-limited measurements (*PPFD* from 40 to 140 μmol m^{-2} s^{-1}). Symbols give the mean *A* ± s.e. and mean *J* ± s.e. at each level of *PPFD*. Lines give the best-fit linear regression; the slope of this regression gives the trait 1/*k*.

where either more erect (Perez et al., 2018; San et al., 2018) or more transparent (Slattery et al., 2016; Walker et al., 2018) leaves allow more light to filter to the bottom of the canopy, ultimately increasing canopy photosynthesis (Zhu et al., 2010). The benefits of this type of canopy manipulation could be 2-fold in NADP-ME C_4 crops such as *Z. mays*, *M.* x *giganteus*, sugarcane or sorghum, providing both increased light to drive more photosynthesis and minimizing the loss of photosynthetic efficiency in lower canopy leaves.

Apart from light, temperature is one other important change in microclimate between canopy and plot positions, as shaded leaves can be expected to be cooler. However, temperature is a less likely candidate than light to explain the lost photosynthetic efficiency of shaded leaves seen at the plot centre in the present study. In C_3 plants, $\phi CO_{2\ max,abs}$ is highly temperature-sensitive, primarily due to increased photorespiration at high temperatures (Ehleringer and Bjorkman, 1977; Long and Spence, 2013). In contrast, due to the C_4 cycle's suppression of photorespiration, $\phi CO_{2\ max,abs}$ has been found to be constant with temperature from 15 to 40°C in C_4 species such as *Atriplex rosea* (Ehleringer and Bjorkman, 1977) and *Alloteropsis semialata* (Osborne et al., 2008). Although loss of $\phi CO_{2\ max,abs}$ has been observed in NADP-ME C_4 grasses such as *Z. mays* due to photodamage during long-term exposure to a combination of high light and cool temperatures (<15°C) (Long and Spence, 2013), this is unlikely to have occurred in the warm summer months during which the present study took place, with maximum daily air temperatures ranging from 19.5 to 33°C at the time measurements were taken. Indeed, since the lower canopy leaves on the exposed southern edge of the stands were exposed to higher light intensities than the shaded lower leaves in the centre of the stands, the expectation would be of a lower $\phi CO_{2\ max,abs}$

due to photodamage in the exposed lower canopy leaves, yet the opposite was found.

Physiological Traits Underpinning Maximum Quantum Yield of CO_2 Assimilation

Under limiting light, reduced α in lower canopy leaves would limit the amount of incident light made available for use within the leaf, and would result in reduced maximum quantum yield on an incident light basis (i.e., $\phi CO_{2\ max,app}$). The fact that α increased in lower canopy leaves shows that in fact their light absorption was improved, not impaired. This pattern in α, along with R_d and A_{sat}, matches established mechanisms of acclimation to low light, as shade leaves: (1) reduce R_d, (2) remobilize N away from photosynthetic enzymes and toward chlorophyll to improve α under limiting light, and (3) translocate N to the upper canopy so sun leaves can increase photosynthetic enzyme content and improve A_{sat} (Boardman, 1977; Chen et al., 2014; Niinemets, 2016b; D'Odorico et al., 2018).

Because of the difference in light availability between sun and shade leaves, shade leaves benefit from partitioning relatively more N toward chlorophyll, compared to sun leaves that partition much more N toward photosynthetic enzymes. Therefore, while shade leaves typically reallocate the N stored in their photosynthetic enzymes and decrease total N content, this primarily results in a loss of A_{sat}, while the apparent maximum quantum yield ($\phi CO_{2\ max,app}$) rises due to increased chlorophyll and, in turn, increased α. The unusual feature in this study is that $\phi CO_{2\ max,app}$ falls despite an increase in α – hence our use of the term maladaptive to describe the response of studied shade leaves to low light. Also, as $\phi CO_{2\ max,abs}$ is measured on an absorbed light basis and derived from the initial linear slope of the light response curve, it is by definition where A is strictly light-limited, ruling out any limitation by N or protein amounts which primarily affect A_{sat} (Hikosaka and Terashima, 1995). In fact, the maximum quantum yield of CO_2 assimilation corrected for chlorophyll content was equivalent in N-stressed and control maize plants (Lu and Zhang, 2000).

Efficient energy transfer at PSII is essential to power photosynthesis under limiting light. F_v/F_m is an effective probe to determine whether damage to PSII has occurred (Baker, 2008). However, the <1% loss of F_v/F_m observed here in lower canopy leaves cannot explain the much more substantial losses in $\phi CO_{2\ max,abs}$.

$1/k$, i.e., the ratio of A to the rate of linear electron transport through PSII (J) at low light, is decreased when the energetic compounds NADPH, reduced ferredoxin, and ATP, produced through linear electron flux, are diverted away from photosynthetic carbon metabolism and into other energy-consuming processes (e.g., nitrogen metabolism, Mehler reaction) (Delatorre et al., 1991; Baker, 2008). This is observed as a reduced slope of the linear relationship of A to J. Under limiting light, this will cause a decline in $\phi CO_{2\ max,abs}$. However, in lower canopy leaves, $1/k$ was greater than at the top of the canopy, implying leaves at the bottom of the canopy actually had fewer, not more, alternative energy sinks to photosynthetic carbon metabolism. In fact, alternative energy sinks overall were minimal: $1/k$ was always close to the theoretical maximum of 0.25 mol mol^{-1}, i.e., for each mol CO_2 assimilated, a theoretical minimum of $k = 4$ mol electron equivalents must be produced through linear electron flux when there are no alternative energy sinks (Baker, 2008).

One possible explanation for loss of $\phi CO_{2\ max,abs}$ without reduced $1/k$ is that lower canopy leaves in the plot centre did have increased alternative sinks, but these were not detected by the leaf fluorescence measurements. One caveat of PSII fluorescence is that the signal is primarily obtained from PSII closest to the leaf surface, with less contribution from PSII deeper in the leaf. Therefore $1/k$ is obtained from A throughout the entire leaf cross-section, and J obtained from PSII fluorescence at the leaf surface. If alternative energy sinks diverted NADPH and ATP from deeper PSII, this could result in a decrease of $\phi CO_{2\ max,abs}$ without an apparent effect to $1/k$. Additionally, $1/k$ only measures the partitioning toward A of NADPH and ATP produced through linear electron flux. ATP can also be produced through cyclic electron flux around PSI, a process which bypasses PSII and produces only ATP (von Caemmerer, 2000). Alternative energy sinks for the ATP produced through cyclic electron flux would not be reflected in $1/k$, since $1/k$ is based on the photochemical efficiency of PSII and not PSI. For instance, shaded leaves of field-grown *M.* x *giganteus* show signs of increased leakage of CO_2 from bundle-sheath cells, which should incur additional ATP consumption to power C_4 overcycling of CO_2 (Kromdijk et al., 2008). However, C_4 NADP-ME grasses including *Z. mays* showed increased photon partitioning to PSI, but no significant change in cyclic electron flux, when grown in the shade (Ver Sagun et al., 2019).

$\phi CO_{2\ max,abs}$ measured in non-stressed conditions is typically well conserved across various species (Long et al., 1993). Surprisingly, here *Z. mays* showed $\phi CO_{2\ max,abs}$ 23% greater than *M.* x *giganteus*. This may be explained in part by the greater F_v/F_m and $1/k$ in *Z. mays* relative to *M.* x *giganteus*. In previous measurements on nearby plots of the same species, $\phi CO_{2\ max,abs}$ of *M.* x *giganteus* and *Z. mays* were within just 9% of one another (Pignon et al., 2017), suggesting the greater inter-species difference observed here may be an effect of different location or growing season.

Potential Effects of Breeding and Management on Maximum Quantum Yield of CO_2 Assimilation

These results raise the question of why such productive crops show a maladaptive acclimation to shade. *Zea mays* in particular is being grown at ever greater densities (Lobell et al., 2014), resulting in increased leaf area indices and self-shading, but these high densities are a recent construct of cultivation. The ancestors of cultivated *Z. mays* grew largely as isolated plants in semi-arid and nutrient limited environments, such that they would have evolved as plants in which most or all leaves were exposed to full sunlight and shading was rare. Similarly, *Miscanthus* spp. often occur as single tall clumps, standing above surrounding plants

and so too would experience relatively little shading, compared to field production stands. Having evolved as sun plants, there may have been insufficient time for them to adapt to the recent production in dense stands.

Although both species are part of the same C_4 evolutionary clade, modern *Z. mays* hybrids have been subject to centuries of selection for productivity, which has been particularly intense in the last 50 years, while *M.* x *giganteus* is only just emerging as a crop. This may suggest that there is variability that could be selected to overcome this significant Achilles heel in this important group of crops. *Z. mays* is considered to have diverged in the evolution of the Andropogoneae before divergence of the genera *Saccharum, Sorghum* and *Miscanthus* (Kim et al., 2014; Singh et al., 2019). The occurrence of this maladaptation in both *Z. mays* and *M.* x *giganteus* suggests that the major crops sorghum and sugarcane are likely similarly affected. Given that *Z. mays* accounts for more cereal grain than any other crop globally, overcoming this maladaptation to shade would contribute very significantly toward meeting the 60% increase in food demand anticipated for mid-century (Long et al., 2015; FAO, 2017).

DATA AVAILABILITY STATEMENT

All original data is freely available without restrictions from the Illinois Data Bank, doi: 10.13012/B2IDB-4821336_V1.

AUTHOR CONTRIBUTIONS

RC and ER collected the physiological data and wrote the manuscript. CP supervised the experiment, performed the statistical analysis, and assisted in manuscript writing. SL conceived the experiment and assisted in manuscript writing.

FUNDING

This work was supported by the project Realizing Increased Photosynthetic Efficiency that is funded by the Bill & Melinda Gates Foundation, the Foundation for Food and Agriculture Research, and the UK Department for International Development (UK Aid) under grant number OPP1172157.

ACKNOWLEDGMENTS

We thank Prof. Andrew Leakey for his support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00783/full#supplementary-material

REFERENCES

Baker, N. R. (2008). Chlorophyll fluorescence: a probe of photosynthesis in vivo. *Annu. Rev. Plant Biol.* 59, 89–113. doi: 10.1146/annurev.arplant.59.032607.092759

Baker, N. R., Long, S. P., and Ort, D. R. (1988). Photosynthesis and temperature, with particular reference to effects on quantum yield. *Symp. Soc. Exp. Biol.* 42, 347–375.

Bellasio, C., and Griffiths, H. (2014). Acclimation of C4 metabolism to low light in mature maize leaves could limit energetic losses during progressive shading in a crop canopy. *J. Exp. Bot.* 65, 3725–3736. doi: 10.1093/jxb/eru052

Beyschlag, W., Barnes, P. W., Ryel, R., Caldwell, M. M., and Flint, S. D. (1990). Plant competition for light analyzed with a multispecies canopy model.2. Influence of photosynthetic characteristics on mixtures of wheat and wild oat. *Oecologia* 82, 374–380. doi: 10.1007/bf00317486

Boardman, N. K. (1977). Comparative photosynthesis of sun and shade plants. *Annu. Rev. Plant Physiol.* 28, 355–377. doi: 10.1146/annurev.pp.28.060177.002035

Chen, A., Lichstein, J. W., Osnas, J. L. D., and Pacala, S. W. (2014). Species-independent down-regulation of leaf photosynthesis and respiration in response to shading: evidence from six temperate tree species. *PLoS One* 9:e91798. doi: 10.1371/journal.pone.0091798

Christin, P. A., Samaritani, E., Petitpierre, B., Salamin, N., and Besnard, G. (2009). Evolutionary Insights on C_4 Photosynthetic Subtypes in Grasses from Genomics and Phylogenetics. *Genome Biol. Evol.* 1, 221–230. doi: 10.1093/gbe/evp020

Delatorre, A., Delgado, B., and Lara, C. (1991). nitrate-dependent O_2 evolution in intact leaves. *Plant Physiol.* 96, 898–901. doi: 10.1104/pp.96.3.898

D'Odorico, P., Emmel, C., Revill, A., Liebisch, F., Eugster, W., and Buchmann, N. (2018). Vertical patterns of photosynthesis and related leaf traits in two contrasting agricultural crops. *Funct. Plant Biol.* 46, 213–227.

Ehleringer, J., and Bjorkman, O. (1977). Quantum yields for CO_2 uptake in C3 and C4 plants - dependence on temperature, CO_2, and O_2 concentration. *Plant Physiol.* 59, 86–90. doi: 10.1104/pp.59.1.86

FAO (2017). *The Future of Food and Agriculture Trends and Challenges*. Rome: Food and Agriculture Organisation.

FAOSTAT (2017). *Database Collection of the Food and Agriculture Organization of the United Nations*. Rome: FAO.

Givnish, T. J. (1988). Adaptation to sun and shade - a whole-plant perspective. *Austr. J. Plant Physiol.* 15, 63–92.

Hadebe, S. T., Modi, A. T., and Mabhaudhi, T. (2017). Drought tolerance and water use of cereal crops: a focus on sorghum as a food security crop in Sub-Saharan Africa. *J. Agron. Crop Sci.* 203, 177–191. doi: 10.1111/jac.12191

Heaton, E. A., Dohleman, F. G., and Long, S. P. (2008). Meeting US biofuel goals with less land: the potential of Miscanthus. *Glob. Chang. Biol.* 14, 2000–2014. doi: 10.1111/j.1365-2486.2008.01662.x

Hikosaka, K., Noguchi, K., and Terashima, I. (2016). "Modeling leaf gas exchange," in *Canopy Photosynthesis: From Basics to Applications*, Vol. 42, eds K. Hikosaka, U. Niinemets, and N. P. R. Anten (Dordrecht: Springer), 60–99.

Hikosaka, K., and Terashima, I. (1995). A model of the acclimation of photosynthesis in the leaves of C3 plants to sun and shade with respect to nitrogen use. *Plant Cell Environ.* 18, 605–618. doi: 10.1111/j.1365-3040.1995.tb00562.x

Hoyaux, J., Moureaux, C., Tourneur, D., Bodson, B., and Aubinet, M. (2008). Extrapolating gross primary productivity from leaf to canopy scale in a winter wheat crop. *Agric. Forest Meteorol.* 148, 668–679. doi: 10.1016/j.agrformet.2007.11.010

Joo, E., Zeri, M., Hussain, M. Z., Delucia, E. H., and Bernacchi, C. J. (2017). Enhanced evapotranspiration was observed during extreme drought from Miscanthus, opposite of other crops. *Glob. Chang. Biol. Bioenergy* 9, 1306–1319. doi: 10.1111/gcbb.12448

Kim, C., Wang, X. Y., Lee, T. H., Jakob, K., Lee, G. J., and Paterson, A. H. (2014). Comparative analysis of Miscanthus and Saccharum reveals a shared whole-genome duplication but different evolutionary fates. *Plant Cell* 26, 2420–2429. doi: 10.1105/tpc.114.125583

Kromdijk, J., Schepers, H. E., Albanito, F., Fitton, N., Carroll, F., Jones, M. B., et al. (2008). Bundle sheath leakiness and light limitation during C4 leaf and canopy CO_2 uptake. *Plant Physiol.* 148, 2144–2155. doi: 10.1104/pp.108.129890

Kubasek, J., Urban, O., and Santrucke, J. (2013). C4 plants use fluctuating light less efficiently than do C3 plants: a study of growth, photosynthesis and carbon isotope discrimination. *Physiol. Plant* 149, 528–539. doi: 10.1111/ppl.12057

Leakey, A. D. B., Uribelarrea, M., Ainsworth, E. A., Naidu, S. L., Rogers, A., Ort, D. R., et al. (2006). Photosynthesis, productivity, and yield of maize are not affected by open-air elevation of CO_2 concentration in the absence of drought. *Plant Physiol.* 140, 779–790. doi: 10.1104/pp.105.073957

LeBauer, D., Kooper, R., Mulrooney, P., Rohde, S., Wang, D., Long, S. P., et al. (2018). BETYdb: a yield, trait, and ecosystem service database applied to second-generation bioenergy feedstock production. *Glob. Chang. Biol. Bioenergy* 10, 61–71. doi: 10.1111/gcbb.12420

Lobell, D. B., Roberts, M. J., Schlenker, W., Braun, N., Little, B. B., Rejesus, R. M., et al. (2014). Greater sensitivity to drought accompanies maize yield increase in the US midwest. *Science* 344, 516–519. doi: 10.1126/science.1251423

Long, S. P. (1993). *The Significance of Light-limiting Photosynthesis to Crop Canopy Carbon Gain and Productivity—A Theoretical Analysis*. Dordrecht: Springer.

Long, S. P., Marshall-Colon, A., and Zhu, X. G. (2015). Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell* 161, 56–66. doi: 10.1016/j.cell.2015.03.019

Long, S. P., Postl, W. F., and Bolharnordenkampf, H. R. (1993). Quantum yields for uptake of carbon-dioxide in C3 vascular plants of contrasting habitats and taxonomic groupings. *Planta* 189, 226–234.

Long, S. P., and Spence, A. K. (2013). Toward cool C4 crops. *Annu. Rev. Plant Biol.* 64, 701–722.

Loriaux, S. D., Avenson, T. J., Welles, J. M., McDermitt, D. K., Eckles, R. D., Riensche, B., et al. (2013). Closing in on maximum yield of chlorophyll fluorescence using a single multiphase flash of sub-saturating intensity. *Plant Cell Environ.* 36, 1755–1770. doi: 10.1111/pce.12115

Lu, C. M., and Zhang, J. H. (2000). Photosynthetic CO_2 assimilation, chlorophyll fluorescence and photoinhibition as affected by nitrogen deficiency in maize plants. *Plant Sci.* 151, 135–143. doi: 10.1016/s0168-9452(99)00207-1

Marchiori, P. E. R., Machado, E. C., and Ribeiro, R. V. (2014). Photosynthetic limitations imposed by self-shading in field-grown sugarcane varieties. *Field Crops Res.* 155, 30–37. doi: 10.1016/j.fcr.2013.09.025

Niinemets, U. (2016a). Leaf age dependent changes in within-canopy variation in leaf functional traits: a meta-analysis. *J. Plant Res.* 129, 313–338. doi: 10.1007/s10265-016-0815-2

Niinemets, U. (2016b). "Within-canopy variations in functional leaf traits: structural, chemical and ecological controls and diversity of responses," in *Canopy Photosynthesis: From Basics to Applications*, Vol. 42, eds K. Hikosaka, U. Niinemets, and N. P. R. Anten (Dordrecht: Springer), 101–141. doi: 10.1007/978-94-017-7291-4_4

Osborne, C. P., Wythe, E. J., Ibrahim, D. G., Gilbert, M. E., and Ripley, B. S. (2008). Low temperature effects on leaf physiology and survivorship in the C3 and C4 subspecies of *Alloteropsis semialata*. *J. Exp. Bot.* 59, 1743–1754. doi: 10.1093/jxb/ern062

Pearcy, R. W. (1990). Sunflecks and photosynthesis in plant canopies. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 41, 421–453. doi: 10.1146/annurev.pp.41.060190.002225

Perez, R. P. A., Dauzat, J., Pallas, B., Lamour, J., Verley, P., Caliman, J. P., et al. (2018). Designing oil palm architectural ideotypes for optimal light interception and carbon assimilation through a sensitivity analysis of leaf traits. *Ann. Bot.* 121, 909–926. doi: 10.1093/aob/mcx161

Pignon, C. P., Jaiswal, D., McGrath, J. M., and Long, S. P. (2017). Loss of photosynthetic efficiency in the shade. An achilles heel for the dense modern stands of our most productive C4 crops? *J. Exper. Bot.* 68, 335–345. doi: 10.1093/jxb/erw456

Sage, R. F., and McKown, A. D. (2006). Is C4 photosynthesis less phenotypically plastic than C3 photosynthesis? *J. Exper. Bot.* 57, 303–317. doi: 10.1093/jxb/erj040

Sales, C. R. G., Ribeiro, R. V., Hayashi, A. H., Marchiori, P. E. R., Silva, K. I., Martins, M. O., et al. (2018). Flexibility of C4 decarboxylation and photosynthetic plasticity in sugarcane plants under shading. *Environ. Exp. Bot.* 149, 34–42. doi: 10.1016/j.envexpbot.2017.10.027

San, N. S., Ootsuki, Y., Adachi, S., Yamamoto, T., Ueda, T., Tanabata, T., et al. (2018). A near-isogenic rice line carrying a QTL for larger leaf inclination angle yields heavier biomass and grain. *Field Crops Res.* 219, 131–138. doi: 10.1016/j.fcr.2018.01.025

Sattin, M., Zuin, M. C., and Sartorato, I. (1994). Light quality beneath field-grown maize, soybean and wheat canopies - red-far red variations. *Physiol. Plant* 91, 322–328. doi: 10.1034/j.1399-3054.1994.910230.x

Singh, R. B., Singh, B., and Singh, R. K. (2019). Cross-taxon transferability of sugarcane expressed sequence tags derived microsatellite (EST-SSR) markers across the related cereal grasses. *J. Plant Biochem. Biotechnol.* 28, 176–188. doi: 10.1007/s13562-019-00502-6

Slattery, R. A., Grennan, A. K., Sivaguru, M., Sozzani, R., and Ort, D. R. (2016). Light sheet microscopy reveals more gradual light attenuation in light-green versus dark-green soybean leaves. *J. Exp. Bot.* 67, 4697–4709. doi: 10.1093/jxb/erw246

Smith, C. M., David, M. B., Mitchell, C. A., Masters, M. D., Anderson-Teixeira, K. J., Bernacchi, C. J., et al. (2013). Reduced nitrogen losses after conversion of row crop agriculture to perennial biofuel crops. *J. Environ. Qual.* 42, 219–228. doi: 10.2134/jeq2012.0210

Sonawane, B. V., Sharwood, R. E., Whitney, S., and Ghannoum, O. (2018). Shade compromises the photosynthetic efficiency of NADP-ME less than that of PEP-CK and NAD-ME C4 grasses. *J. Exp. Bot.* 69, 3053–3068. doi: 10.1093/jxb/ery129

Tazoe, Y., Hanba, Y. T., Furumoto, T., Noguchi, K., and Terashima, I. (2008). Relationships between quantum yield for CO_2 assimilation, activity of key enzymes and CO_2 leakiness in *Amaranthus cruentus*, a C4 dicot, grown in high or low light. *Plant Cell Physiol.* 49, 19–29. doi: 10.1093/pcp/pcm160

Ver Sagun, J., Badger, M. R., Chow, W. S., and Ghannoum, O. (2019). Cyclic electron flow and light partitioning between the two photosystems in leaves of plants with different functional types. *Photosynth. Res.* 142, 321–334.

von Caemmerer, S. (2000). "Modelling C4 photosynthesis," in *Biochemical Models of Leaf Photosynthesis*, ed. S. Von Caemmerer (Collingwood: CSIRO Publishing), 91–122.

von Caemmerer, S., and Farquhar, G. D. (1981). Some relationships between the biochemistry of photosynthesis and the gas exchange of leaves. *Planta* 153, 376–387. doi: 10.1007/bf00384257

Walker, B. J., Drewry, D. T., Slattery, R. A., VanLoocke, A., Cho, Y. B., and Ort, D. R. (2018). Chlorophyll can be reduced in crop canopies with little penalty to photosynthesis. *Plant Physiol.* 176, 1215–1232. doi: 10.1104/pp.17.01401

Welker, C. A. D., Kellogg, E. A., and Prado, J. (2014). *Andropogoneae* versus *Sacchareae* (Poaceae: Panicoideae): the end of a great controversy. *Taxon* 63, 643–646. doi: 10.12705/633.5

Yabiku, T., and Ueno, O. (2019). Structural and photosynthetic re-acclimation to low light in C4 maize leaves that developed under high light. *Ann. Bot.* 124, 437–445. doi: 10.1093/aob/mcz092

Yin, X., Belay, D., van der Putten, P., and Struik, P. (2014). Accounting for the decrease of photosystem photochemical efficiency with increasing irradiance to estimate quantum yield of leaf photosynthesis. *Photosynth. Res.* 122, 323–335. doi: 10.1007/s11120-014-0030-8

Yin, X. Y., and Struik, P. C. (2012). Mathematical review of the energy transduction stoichiometries of C4 leaf photosynthesis under limiting light. *Plant Cell Environ.* 35, 1299–1312. doi: 10.1111/j.1365-3040.2012.02490.x

Zhu, X.-G., Long, S. P., and Ort, D. R. (2008). What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Curr. Opin. Biotechnol.* 19, 153–159. doi: 10.1016/j.copbio.2008.02.004

Zhu, X.-G., Long, S. P., and Ort, D. R. (2010). Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol.* 61, 235–261. doi: 10.1146/annurev-arplant-042809-112206

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 30 June 2020
doi: 10.3389/fpls.2020.00935

What Matters for C_4 Transporters: Evolutionary Changes of Phospho*enol*pyruvate Transporter for C_4 Photosynthesis

Ming-Ju Amy Lyu[1], *Yaling Wang*[1], *Jianjun Jiang*[2], *Xinyu Liu*[1], *Genyun Chen*[1] *and Xin-Guang Zhu*[1]*

[1] National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence In Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China, [2] Wisconsin Institute for Discovery & Laboratory of Genetics, University of Wisconsin, Madison, WI, United States

OPEN ACCESS

Edited by:
Roxana Khoshravesh Astaneh, University of New Mexico, United States

Reviewed by:
Stanislav Kopriva, University of Cologne, Germany
Liangsheng Zhang, Fujian Agriculture and Forestry University, China
Sylvain Aubry, University of Zurich, Switzerland

***Correspondence:**
Xin-Guang Zhu
zhuxg@cemps.ac.cn

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: 17 March 2020
Accepted: 08 June 2020
Published: 30 June 2020

Citation:
Lyu M-JA, Wang Y, Jiang J, Liu X, Chen G and Zhu X-G (2020) What Matters for C_4 Transporters: Evolutionary Changes of Phosphoenolpyruvate Transporter for C_4 Photosynthesis. Front. Plant Sci. 11:935. doi: 10.3389/fpls.2020.00935

C_4 photosynthesis is a complex trait that evolved from its ancestral C_3 photosynthesis by recruiting pre-existing genes. These co-opted genes were changed in many aspects compared to their counterparts in C_3 species. Most of the evolutionary changes of the C_4 shuttle enzymes are well characterized, however, evolutionary changes for the recruited metabolite transporters are less studied. Here we analyzed the evolutionary changes of the shuttle enzyme phospho*enol*pyruvate (PEP) transporter (PPT) during its recruitment from C_3 to C_4 photosynthesis. Our analysis showed that among the two PPT paralogs PPT1 and PPT2, PPT1 was the copy recruited for C_4 photosynthesis in multiple C_4 lineages. During C_4 evolution, PPT1 gained increased transcript abundance, shifted its expression from predominantly in root to in leaf and from bundle sheath cell to mesophyll cell, and gained more rapid and long-lasting responsiveness to light. Modifications occurred in both regulatory and coding regions in C_4 PPT1 as compared to C_3 PPT1, however, the PEP transporting function of PPT1 remained. We found that PPT1 of a *Flaveria* C_4 species recruited a MEM1 B submodule in the promoter region, which might be related to the increased transcript abundance of PPT1 in C_4 mesophyll cells. The case study of PPT further suggested that high transcript abundance in a proper location is of high priority for PPT to support C_4 function.

Keywords: C_4 photosynthesis, evolution, *Flaveria*, phospho*enol*pyruvate transporter

HIGHLIGHTS

During the evolution of C_4 photosynthesis, one of the paralogs of PPTs, i.e., PPT1, which shows lower transcript abundance in leaf but higher transcript abundance in root was recruited in multiple C_4 lineages. Compared to its counterpart in C_3 species, PPT1 in C_4 species shows altered expression location, enhanced transcript abundance, increased light responsiveness, which might be related to a newly recruited MEM1 B submodule in its promoter.

INTRODUCTION

Compared to C_3 photosynthesis, C_4 photosynthesis has higher light, nitrogen, and water using efficiencies (Sage and Zhu, 2011). It achieves these superior properties through a CO_2 concentrating mechanism operating in a specialized leaf anatomical feature termed "Kranz anatomy" (Hatch, 1987). The CO_2 concentrating mechanism involves many enzymes and metabolite transporters, which together pump CO_2 from mesophyll cells (MC) to bundle sheath cells (BSC), creating a localized high CO_2 environment in the BSC around ribulose bisphosphate carboxylase/oxygenase (Rubisco). All genes required for the operation of C_4 photosynthesis are pre-existing in C_3 ancestors and play housekeeping functions (Aubry et al., 2011). The evolution of C_4 photosynthesis is therefore a process of recruiting and re-organizing pre-existing genes to fulfill new functions in C_4 photosynthesis (West-Eberhard et al., 2011; Burgess et al., 2016).

The specific modifications that have occurred on the genes recruited for C_4 photosynthesis are unknown. Comparisons of genes involved in C_4 photosynthesis in C_4 species and their counterparts in C_3 species showed that these genes were modified in different aspects (Gowik and Westhoff, 2011), e.g., increasing transcript abundance [see review in (Hibberd and Covshoff, 2010)]; acquiring cell specific expression (Hatch and Osmond, 1976; Aubry et al., 2014); gaining modifications in protein coding regions resulting in suitability for C_4 photosynthesis (Blasing et al., 2000; Paulus et al., 2013); obtaining new *cis*-elements (Gowik et al., 2004; Williams et al., 2016; Gupta et al., 2020); and having more copies (Bianconi et al., 2018).

Most current evolutionary studies on genes involved in C_4 photosynthesis focus on enzymes directly related to the carbon shuttle enzymes, such as phospho*enol*pyvuvate (PEP) carboxylase (PEPC), phospho*enol*pyvuvate carboxykinase (PEP-CK), nicotinamide adenine dinucleotide phosphate (NADP) malic enzyme (NADP-ME), pyruvate phosphate kinase (PPDK), and malate dehydrogenase (MDH) (Westhoff and Gowik, 2004; Christin and Besnard, 2009; Christin et al., 2013; Moreno-Villena et al., 2018). However, to establish an efficient CO_2 concentrating mechanism, C_4 plants recruited these carbon shuttle enzymes and required metabolite transporters, but also recruited a number of proteins responsible for transport of metabolites required for C_4 photosynthesis in both MC and BSC. In fact, compared to C_3 photosynthesis, the extensive usage of transporters is a major feature of C_4 photosynthesis. Quantitatively, to produce one molecule of triose phosphate for the synthesis of sucrose, only one transporter is needed in C_3 photosynthesis, while at least 30 metabolite transport steps are involved in the NADP-ME type C_4 photosynthesis (Weber and Von Caemmerer, 2010).

Furthermore, the flux through the transporters is much higher in C_4 (Wang et al., 2014). In C_3 plants, the end-product of photosynthesis, *i.e.*, triose phosphate (TP), is exported as one unit, therefore the flux through the triose phosphate transporter is 1/3 of the photosynthetic CO_2 uptake rate. In C_4 photosynthesis, however, the flux of metabolite transport between different compartments is higher than the photosynthetic CO_2 uptake rate due to the leakage of CO_2 from BSC to MC. Furthermore, C_4 plants usually have a higher leaf photosynthetic CO_2 uptake rate compared to C_3 plants. Therefore, C_4 photosynthesis demands a much higher capacity for metabolite transport (Hatch and Osmond, 1976; Von Caemmerer and Furbank, 2003). Indeed, a number of transporters on the chloroplast envelope, including PEP transporter (PPT), pyruvate transporter (BASS2), and malate transporter in MC (DIT1), all show higher transcript abundance in C_4 species than in C_3 species (Emms et al., 2016; Lyu et al., 2018; Moreno-Villena et al., 2018). Identifying the C_4 paralogs of individual metabolite transporters, understanding their evolutionary modifications and the molecular mechanisms behind the increased abundance or capacity of these transporters can help better understand the emergence of C_4 metabolism and guide the engineering of C_4 metabolism into a C_3 metabolic background.

In this study, we aim to characterize the evolutionary changes of a C_4 metabolite transporter PPT, which transports PEP (Knappe et al., 2003), a substrate for the first step of C_4 acid formation in C_4 photosynthesis. In fact, PEP is involved in a number of metabolic pathways in higher plants. **Figure 1** shows the reactions for which PEP is either a substrate or a product in a typical NADP-ME type C_4 leaf. Specifically, PEP is the substrate of PEPC, and its carboxylation represents the first step of CO_2 fixation in C_4 photosynthesis. PEP is also involved in the shikimate pathway in chloroplasts, which generates aromatic amino acids and secondary metabolites (Fischer et al., 1997; Herrmann and Weaver, 1999). Moreover, PEP is a substrate of the citric acid cycle in mitochondria (Krebs, 1940). Recent studies show that PEP is involved in nitrogen recycling from xylem (Bailey and Leegood, 2016) and in nitrogen mobilization from aging leaves (Taylor et al., 2010).

Considering that PEP functions in multiple metabolic pathways, it is safe to infer that the PEP transporting process is crucial in plants. Here, we conducted a systematic comparison of different properties of PPT between C_3 and C_4 plants. Specifically, we first constructed a phylogeny of PPT in Viridiplantae, which includes 23 species spanning chlorophytes to angiosperms to infer the orthologous relationships and copy number of PPT. Then, we compared a number of properties of PPT between C_3 and C_4 species, including PPT gene expression, amino acid sequences, and physiological functions. Our results showed that the paralog with relatively low transcript abundance in leaf of C_3 species was constantly recruited for C_4 photosynthesis in multiple C_4 lineages. In an example study in *Flaveria*, we found that PPT1 from a C_4 species gained a MEM1 B submodule, which might contribute to the changes in transcriptional properties of PPT1 in C_4 species. Comparing PPT1 between C_4 and C_3 species showed that PPT1 has dramatic modifications in the coding region, however, its metabolic function remained the same. The evolutionary changes of PPT suggest that high transcript abundance in the proper location is the key feature of transporters for C_4 photosynthesis.

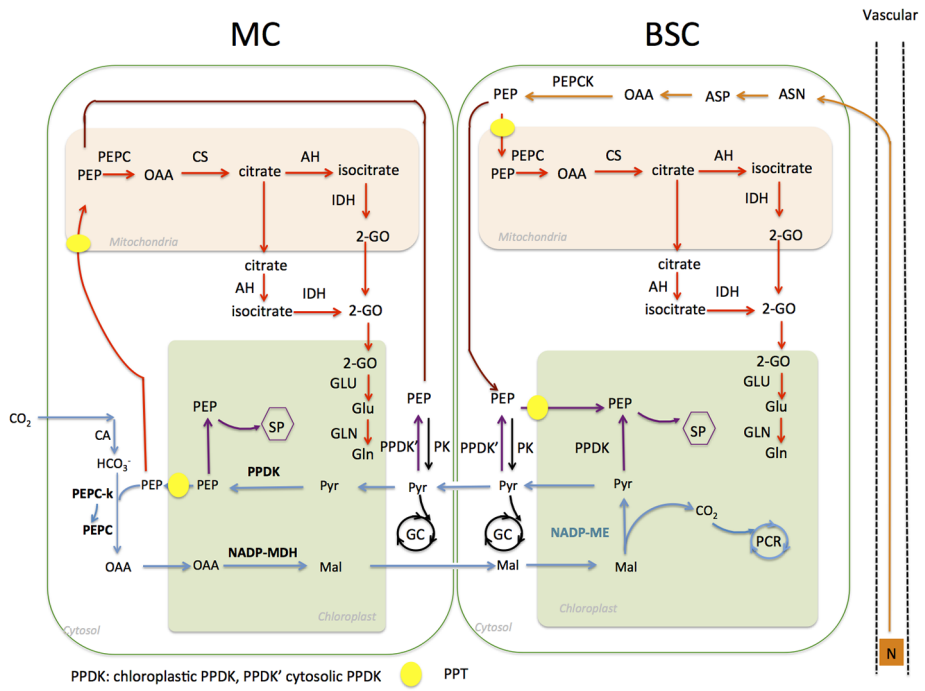

FIGURE 1 | Schematic representation of phospho*enol*pyruvate (PEP) related metabolic pathways in nicotinamide adenine dinucleotide phosphate malic enzyme (NADP-ME) type C_4 *Flaveria* species In MC, PEP is used as a substrate for PEPC catalyzed carboxylation (blue lines). It is also a substrate of the shikimate pathway (SP) in chloroplasts, which is expected to exist in both MC and BSC (purple lines); moreover, phospho*enol*pyruvate transporter (PEP) is a substrate for the citrate pathway in mitochondria (red lines) and glycolysis in cytosol (black lines). PEP is also involved in nitrogen recycling from xylem (orange lines). (Abbreviations: MC, mesophyll cell; BSC, bundle sheath cell; SP, shikimate pathway; PCR, photosynthetic carbon reduction; GC, glycolysis cycle.)

MATERIALS AND METHODS

Construction of the Phospho*enol*pyruvate Transporter Phylogenetic Tree

To construct the phylogenetic tree of PPT, we used protein sequences from 23 species with genome sequences available in phytozome (http://phytozome.jgi.doe.gov/). These included representative species along the phylogeny of Viridiplantae, spanning from basal species belonging to chlorophytes (*Micromonas pusilla* and *Chlamydomonas reinhardtii*), embryophytes (*Marchantia polymorpha*), tracheophytes (*Selaginella moellendorffii*), and to higher angiosperm plants (*Amborella trichopoda*). Among these species, ten are eudicots and eight are monocots (**Figure 2**).

The genome-wide protein sequences of these 23 species were downloaded from Phytozome. We used OrthoFinder (V2.2.7) (Emms and Kelly, 2019) with default parameters to predict the orthologous groups. We found one orthologous group containing both PPT1 (AT5G33320) and PPT2 (AT3G01550) of *Arabidopsis thaliana*, therefore, proteins from this orthologous group were regarded as members of the PPT gene family and used to construct the gene tree of PPTs. We included triose-phosphate/phosphate translocator (TPT) from *A. thaliana* (AT5G46110) as an outgroup. All orthologous proteins of the PPT gene family together with the outgroup protein were aligned using MUSCLE (Edgar, 2004) with default parameters. The gene tree was constructed with RAXML software (Stamatakis, 2006) based on protein sequence alignment with the PROTGAMMAILG model. The robustness of the tree topology was evaluated by bootstrap scores, which were calculated from 1,000 independently constructed gene trees.

Procedures Used to Survey Transcript Abundance of Phospho*enol*pyruvate Transporters From Published RNA-Seq Data

High transcript abundance is suggested as a major feature of genes recruited to support C_4 functions (Moreno-Villena et al., 2018), so we tested whether this applies to PPT. Specifically, we compared the transcript abundance of PPT1 and PPT2 in leaf among species with different photosynthetic types; we also compared the expression patterns of PPT1 and PPT2 in different tissues and cell types. We surveyed RNA-seq data from four independent C_4 lineages, namely: *Heliotropium*, *Mollugo*, *Neurachne*, and *Flaveria* available from 1 KP (http://www.onekp.com/blast.html). Except for *Neurachne*, which had RNA-seq data from shoot, RNA-seq data of other three genera were from mature leaves. RNA isolation, quality control, library preparation, and sequencing procedures are summarized in Johnson et al. (2012). Data collection information is available

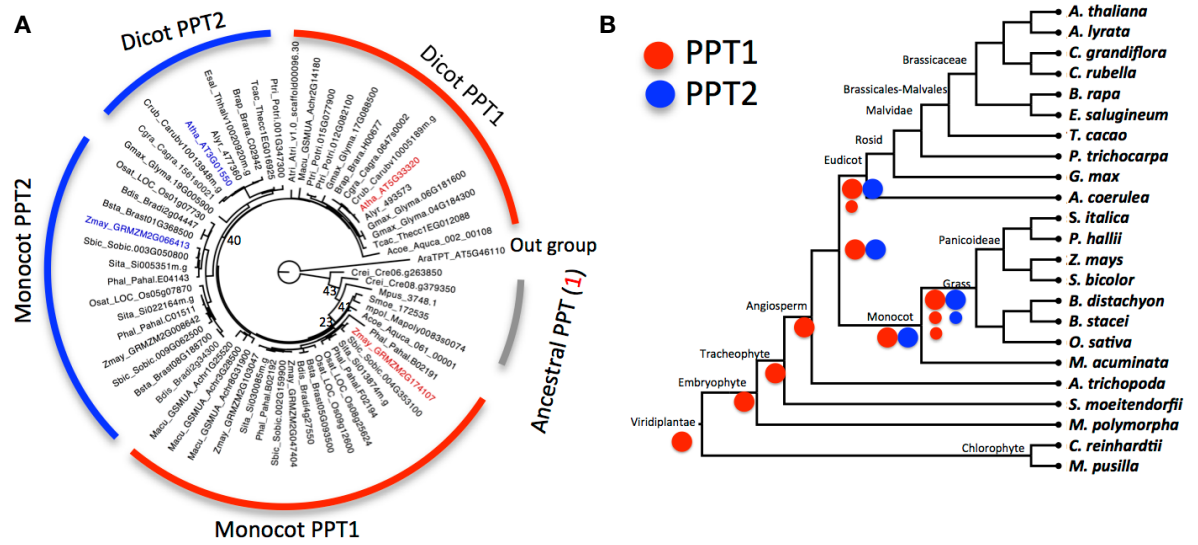

FIGURE 2 | The evolution of phospho*enol*pyruvate transporter (PPT) in Viridiplantae **(A)** Gene tree of PPT family from 23 representative species of Viridiplantae. The tree was inferred from the alignment of protein sequences of PPT using the maximum likelihood method. Numbers beside each node are the bootstrap scores from 1,000 simulated trees; bootstrap scores lower than 60 in the major branch are shown (full bootstrap scores are in **Figure S1**). PPT1 of *Arabidopsis thaliana* and *Zea mays* are highlighted in red and PPT2 in blue. Triose phosphate/phosphate translocator (TPT) of *A. thaliana* is used as outgroup. **(B)** Schematic representation of the evolution of PPT1 and PPT2 based on phylogenetic relationship of species. Ancestral species have one copy of PPT that is more similar to PPT1 than to PPT2 of higher species. PPT1 has one or two copies in eudicot species and two or three copies in monocot species. Red circles represent PPT1 and blue circles represent PPT2, large circles stand for original copies, and small circles for duplicated copies after the division of monocots and dicots. The phylogenetic relationship of species is inferred from the Phytozome website.

on the 1 KP website (http://www.onekp.com/samples/list.php?set). The RNA-seq source and quantification process for the *Flaveria* species were described in (Lyu et al., 2018). The RNA-seq analysis process for *Heliotropium*, *Mollugo*, and *Neurachne* followed the procedures used for *Flaveria*. Briefly, RNA-Seq data were generated using Illumina with a paired-end sequencing strategy with a read length of 90 bp. Transcripts were assembled using Trinity (version 2.02) (Grabherr et al., 2011) with default parameters except that the minimal length of transcript was restricted to be 300 nt.

Transcript abundance was analyzed by mapping short reads to assembled contigs of corresponding species and then normalizing the transcript abundance to the Fragments Per Kilobase of transcript per Million mapped reads (FPKM) using the RSEM package (version 1.2.10) (Li and Dewey, 2011). Functional annotations of transcripts from dicot species, namely, *Heliotropium*, *Mollugo*, and *Flaveria*, were determined by searching for the best hit in the protein dataset of *A. thaliana* in TAIR 10 (http://www.arabidopsis.org) by using BLAST in protein space with an E-value threshold of 1E−5. We annotated *Neurachne* transcripts by searching for the best hit in the protein dataset of *Zea mays* (**Table S6**). The protein sequences of *Z. mays* were downloaded from Phytozome 10.3 (http://phytozome.jgi.doe.gov/pz/portal.html). PPTs in the four genera were determined as the orthologs of PPTs of *A. thaliana* or *Z. mays*.

When comparing transcript abundance of PPTs in roots and leaves from C_3 and C_4 species, we surveyed processed RNA-seq data and identified species that have RNA-Seq data from both roots and leaves, which include two *Flaveria* species e.g., *Flaveria robusta* and *Flaveria trinervia* (Lyu et al., 2018), two Brassicaceae species, i.e., *Gynandropsis gynandra* and *Tarenaya hassleriana* (Kulahoglu et al., 2014), and 21 species in the grass family (Moreno-Villena et al., 2018). We also compared the transcript abundance of PPTs between BSC and whole leaf in C_3 species, and between BSC and MC in C_4 species based on processed RNA-seq data. Specifically, gene expression data of PPTs in BSC and whole leaf of *A. thaliana* were from (Aubry et al., 2014); gene expression data in BSC and MC of maize were from (Tausta et al., 2014); data of *G. gynandra* were from (Chang et al., 2012); data of *Setaria viridis* were from (John et al., 2014); and data of *Panicum virgatum* were from (Rao et al., 2016). The photosynthetic type and abbreviations of species are listed in **Table S1**.

Quantification of Changes in Phospho*enol*pyruvate Transporter Transcript Levels Under Light Treatments in *Flaveria* Species Using Real-Time Quantitative-PCR

Given that the genus *Flaveria* includes species at different evolutionary stages of C_4 photosynthesis, we further used this genus as a model to examine how the transcript abundance of PPT1 and PPT2 evolved along with the evolution of C_4 photosynthesis. Specifically, we studied this in five species representing four different photosynthetic types, i.e., *F. robusta* (C_3), *Flaveria sonorensis* (C_3–C_4), *Flaveria ramosissima* (C_3–C_4), *F. trinervia* (C_4), and *Flaveria australasica* (C_4). For the *Flaveria*

species used in real-time quantitative (qRT)-PCR and the subsequent genomic study, *F. robusta* and *F. ramosissima*, *Flaveria palmeri*, and *Flaveria bidentis* were provided by Prof. Peter Westhoff (Heinrich-Heine-University); *F. sonorensis*, *F. australasica*, *F. trinervia*, *Flaveria kochiana*, and *Flaveria vaginata* were provided by Prof. Rowan F. Sage (University of Toronto). *Flaveria* plants were grown in soil in growth rooms with air temperature controlled to be 25°C, relative humidity 60%, photoperiod 16/8 h day/night, and photosynthetic photon flux density (PPFD) 500 μmol m^{-2} s^{-1}. The *Flaveria* plants were watered twice a week and fertilized weekly. To study the gene expression differences of PPT1 and PPT2 in response to illumination, 1-month old plants were put into darkness at 6 pm. The dark-adapted plants were illuminated at 9:30 am the next day. Fully expanded leaves, usually the 2nd or 3rd leaf pair counted from the top, were cut after the leaves were illuminated for different time periods, *i.e.*, 0, 0.5, 2, and 4 h, and then flash frozen with into liquid nitrogen. Leaf samples were stored at −80° C before processing.

RNA was extracted following the protocol of the PureLink™ RNA kit (Thermo Fisher Scientific, USA). For qRT-PCR, 0.2–0.5 μg RNA was incubated with Superscript II Reverse Transcriptase (TransGen Biotech, Beijing) to obtain complementary DNA (cDNA). qRT-PCR was conducted following the manufacturer's instructions of the UNICON™ qPCR SYBR Green Master Mix kit (YEASEM, Shanghai). cDNA, buffer, and primers were pipetted to the Hard-Shell PCR 96-well Plates (Bio-Rad, USA), and covered by MicroSeal 'B' Seal (Bio-Rad, USA). qRT-PCR was run in the BIO-RAD CFX connect system (Bio-Rad, USA). Relative transcript abundance was calculated by comparing to ACTIN7 and data were processed using the BIO-RAD CFX Maestro software (Bio-Rad, USA). For each gene, three technical and three biological replicates were performed. The primers used here are listed in **Table S2**.

Prediction of Gene Structure and *cis*-Elements of PPT1 and PPT2 From *Flaveria* Species

The promoter sequences of PPT1 and PPT2 from four *Flaveria* species, namely, *F. robusta*, *F. sonorensis*, *F. ramosissima*, and *F. trinervia*, were obtained from the draft genome sequences of the four species. In order to detect the genomic loci of PPT1 and PPT2, we performed a BLAST search against the genome sequence by using the coding sequences (CDS) of PPT1 and PPT2 in each species as a query sequence and applying BLAST+ (v2.2.31) (Camacho et al., 2009) with E-value < 1E−5. A candidate locus of a gene is manually selected if it reports a series of gapped mapping regions with identity higher than 95%, where mapping regions represent exons and gaps represent introns. The protein sequences of PPTs from the four *Flaveria* species and *A. thaliana* were aligned with MUSCLE (Edgar, 2004), based on which the gene tree was inferred with RAXML software (Stamatakis, 2006) using the same procedure described above.

In order to quantify the transcript abundance of PPT orthologs, we generated RNA-seq data for these four species. The growth conditions of *Flaveria* plants and RNA isolation procedures were the same as those used for the qRT-PCR experiment described above. The cDNA library was constructed with NEBNext Ultra II RNA Kit (New England Biolabs, USA). RNA-seq was performed with the Illumina NovaSeq 6000 platform in the paired-end mode with a read length of 150 bp. The data were submitted to gene expression omnibus (GEO) with the accession number GSE143469. We mapped the RNA-seq reads to genome sequence of each species using STAR (V2.7) (Dobin et al., 2013) and calculated the gene expression in Transcripts per kilobase Per Million mapped reads (TPM) using RSEM (V1.3.3) (Li and Dewey, 2011). We verified the promoter sequences of the copies of PTP1 and PPT2 that showed relatively high transcript abundance in each species by PCR and sequencing. The primers used here are listed in **Table S2**.

The draft genome sequences of the four species were submitted to the National Center for Biotechnology Information (NCBI) with accession number SAMN14943594 for *F. robusta*, SAMN14943595 for *F. sonorensis*, SAMN14943596 for *F. ramosissima* and SAMN14943598 for *F. trinervia*.

Comparison of the Amino Acid Sequences of PPT1 and PPT2

The amino acid sequences of PPT1 and PPT2 of different *Flaveria* species were predicted based on *de novo* assembled transcripts as described in (Lyu et al., 2018). Protein sequences of orthologs were aligned with MUSCLE (Edgar, 2004). We further identified consistent amino acid modifications between C_3 and C_4 species, which were defined as sites that showed differences between C_3 and C_4 species, but that were conserved within C_3 species and also conserved within C_4 species. These consistently identified modifications were mapped to the phylogenetic tree of *Flaveria* (Lyu et al., 2015) to identify the evolutionary stage of their appearance during C_4 evolution in *Flaveria*. With the protein sequence information, we predicted the 3D protein structures of PPT1 of *Flaveria* species using the Iterative Threading ASSEmbly Refinement (I-TASSER) online server (Yang and Zhang, 2015).

We further tested whether PPT1 and PPT2 experienced positive selection in C_4 species using C_3 species and intermediate species as background. First, amino acid sequences of orthologous genes were aligned with the software MUSCLE (Edgar, 2004). Aligned protein sequences were then used to guide the codon-wise alignment of CDS with PAL2NAL (Suyama et al., 2006). After gaps and stop codons were removed, the aligned sequences were input into the PAML package (V4.8) (Yang, 2007) for positive selection tests. Phylogeny of the *Flaveria* species was inferred from our previous work (Lyu et al., 2015). Considering that the phylogeny of *Flaveria* contains two clades, we conducted the positive selection in two independent ways: either including species of both clade A and clade B, or excluding species from clade B which lacks a true C_4 species. In this study, the positive selection test was conducted using a branch-site model (model=2, NSsites=2) under an equal nucleotide substitution condition (CodonFreq=0, all frequencies

were fixed to be 1/61). The likelihood of the null hypothesis was calculated under this branch-site model with fixed dN/dS ratio (ω=1, neutral). The maximum likelihood of the alternative hypothesis was calculated under this branch-site model with flexible dN/dS ratio (ω > 1, positive selection). Then, the likelihood ratio test (LRT) was conducted between the null hypothesis and the alternative hypothesis under the chi-square distribution to accept or reject the alternative hypothesis with the "chi2" function in the PAML package. A threshold *p*-value of 0.05 [Benjamini Hochberg (BH) adjusted] was used to determine positive selection in C_4 species.

To investigate the copy number of 13-aa elements in different *Flaveria* species, DNA was extracted from the 2nd or 3rd pair of leaves counted from the top following the protocol of TIANquick Midi Purification kit (TIANGEN Biotech, Beijing). The primers are listed in **Table S2**.

Determining the Subcellular Localization of *Flaveria* PPT1 and PPT2

We further tested whether the subcellular localization of PPT1 and PPT2 are conserved between C_3 and C_4 species in the *Flaveria* genus. To determine the subcellular localizations of PPT1 and PPT2 from *Flaveria* C_3 and C_4 species, we generated fluorescence fusion proteins by tagging a green fluorescent protein (GFP) in the C-terminal end of PPTs and transiently expressed them in *Nicotiana benthamiana* (tobacco) leaves. Specifically, the CDS of PPT1 and PPT2 were amplified from cDNAs reverse transcribed from RNAs for *F. bidentis* (C_4), and from *de novo* synthesized DNA for *F. robusta* (C_3) by Shanghai Personalbio LLC by PCR. A CDS with the 52-amino acid (52 aa) insertion deleted, i.e., ΔFbid-PPT1, was generated *via* overlapping PCR. All primers are listed in **Table S2**. All the PCR fragments of PPT1 and PPT2 were integrated into the binary vector pCAMBIA1302 *via* homologous recombination-based in-fusion cloning (GBClonart). The promoter used was a CaMV 35S promoter. The final plasmids were verified by Sanger-sequencing (Sangon Biotech, Shanghai). The verified vectors were transformed into *Agrobacterium tumefaciens* (Agrobacterium) strain GV3101 competent cells (TransGen Biotech, Beijing). The Agrobacterium cells were cultured in liquid Luria-Bertani (LB) medium containing rifamycin and kanamycin and re-suspended in infiltration buffer [10 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH5.7, 10 mM $MgCl_2$, 200 μM acetosyringone] to OD_{600} ~1.0. The Agrobacterium cells were infiltrated into tobacco leaves with a syringe. After 36 to 48 h, the fluorescence signals from leaf pavement cells were examined using a confocal fluorescence microscope (Zeiss LSM880, Germany). The autofluorescence signal from chlorophyll was used as a marker for chloroplast thylakoids, with an excitation wavelength of 488 nm and an emission wavelenth of 507 nm.

Testing the Functional Conservation of PPT1 Between C_3 and C_4 Species

Finally, we tested whether the PEP transporting function of PPT1 was conserved between C_3 and C_4 species. We asked whether a C_4 PPT1 can rescue the phenotype of a C_3 PPT1 mutant. The *A. thaliana* PPT1 mutant *cue1-5*, which is an ethyl methanesulfonate (EMS) mutant harboring the R81C mutation in PPT1, was ordered from North American Services Center (NASC) (stock number N3156). Then, we introduced different *Flaveria* PPT1-GFP driven by the *35S* promoter into *cue1-5* mutant *via* an Agrobacterium-mediated floral dipping method. Agrobacterium cells transformed with the binary plasmids were cultured in Luria-Bertani media at 28°C for 48 h. Agrobacterium cells were pelleted and re-suspended in transformation buffer (50 g sucrose, 2.2 g Murashige and Skoog powder, 200 μl Silwet L-77, and 10 μl 6-BA for 1L, pH 5.8 to 6.0) to OD_{600} ~1.0. The *A. thaliana* flowers were dipped in buffer containing Agrobacterium cells and kept for 5 min, and then the plants were put under dark overnight. The floral dipping process was repeated 1 week later. After maturation, the seeds were collected and screened on ½ Murashige and Skoog (MS) agar plates containing hygromycin at a concentration of 35 mg/L. The positive T_1 transformants were transferred to soil. The T_2 lines were used to examine morphological phenotypes. The plants were grown in a growth chamber with a long-day condition (16 light/8 dark), a PPFD of ~100 μmol m^{-2} s^{-1}, and temperature cycle of 23°C during the day and 21°C at night.

RESULTS

The Evolution of Phospho*enol*pyruvate Transporter in the Viridiplantae

To investigate the evolution of the PPT, we first constructed a phylogenetic tree for PPT orthologs from 23 species in Viridiplantae (**Figure 2**). These species were selected to capture the major events in Viridiplantae evolution with one species representing a major evolutionary stage of the Viridiplantae phylogeny. Furthermore, we included 10 eudicot species with six species from the Brassicaceae family and eight monocot species with seven species from the grass family (**Figure 2**). These two families contain the most sequenced genomes; hence they can be used to study how PPTs evolved within families.

The gene tree showed that PPTs were present in all selected species. PPT had one copy in species that evolved before angiosperms, including the two chlorophyte species, *M. pusilla* and *C. reinhardtii*, as well as *M. polymorpha* and *S. moellendorffii*. The angiosperm species *A. trichopoda* also has one copy. In contrast, there were two copies in other angiosperms with one being the ortholog of *A. thaliana* PPT1 and another being the ortholog of *A. thaliana* PPT2 (**Figure 2A** and **Figure S1**). PPTs from lower species showed higher similarity with *A. thaliana* PPT1 than *A. thaliana* PPT2. Furthermore the single copy of PPT in *A. trichopoda* was in the PPT1 lineage of dicots, suggesting that PPT1 was the ancestral copy and PPT2 was a derived copy that originated after the split of *A. trichopoda* from other angiosperm species (**Figure 2B**). There were one or two copies of PPT1 and a single copy of PPT2 in dicot species, whereas there were two or three copies of PPT1 and one or two copies of PPT2 in grass species, consistent with an extra whole

genome duplication (WGD) event in monocot species (Jiao et al., 2014). We observed more rapid evolution of PPT2 compared to PPT1; furthermore, PPT2 in dicots, especially in Brassicaceae, showed faster evolution than in monocots. The physiological significance and underlying mechanisms behind these different evolutionary speeds are unknown.

The Evolution of Phospho*enol*pyruvate Transporters in Transcript Abundance and Tissue Specificity of Expression Along the Emergence of C_4 Species

We further compared the expression abundance between PPT1 and PPT2. First, we examined the transcript abundances of PPT1 and PPT2 in a few sets of species that are evolutionarily closely related but have different photosynthetic types. These species are from four genera with each representing an independent C_4 lineage. Among these four genera, three are dicots, i.e., *Flaveria*, *Heliotropium*, and *Mollugo*, and one is a monocot, i.e., *Neurachne* (**Figure 3A**). The RNA-seq data for the *Flaveria* species are from the 1,000 plants project (1 KP) (Matasci et al., 2014) and (Mallmann et al., 2014), with data from leaf samples, and have been demonstrated to be comparable in FPKM (Lyu et al., 2018). RNA-seq data for the other three genera are also from 1 KP (Matasci et al., 2014). In the analysis, data from mature leaves were used. The comparison showed that in C_3 species, PPT2 displayed higher transcript abundance than PPT1 except in the two C_3 species in *Heliotropium*, namely, *Heliotropium calcicola* (Hcal) and *Heliotropium karwinsky* (Hkar) (**Figure 3A**). The higher expression of PPT2 over PPT1 was also shown in C_3–C_4 species of *Mollugo* and *Neurachne*, as

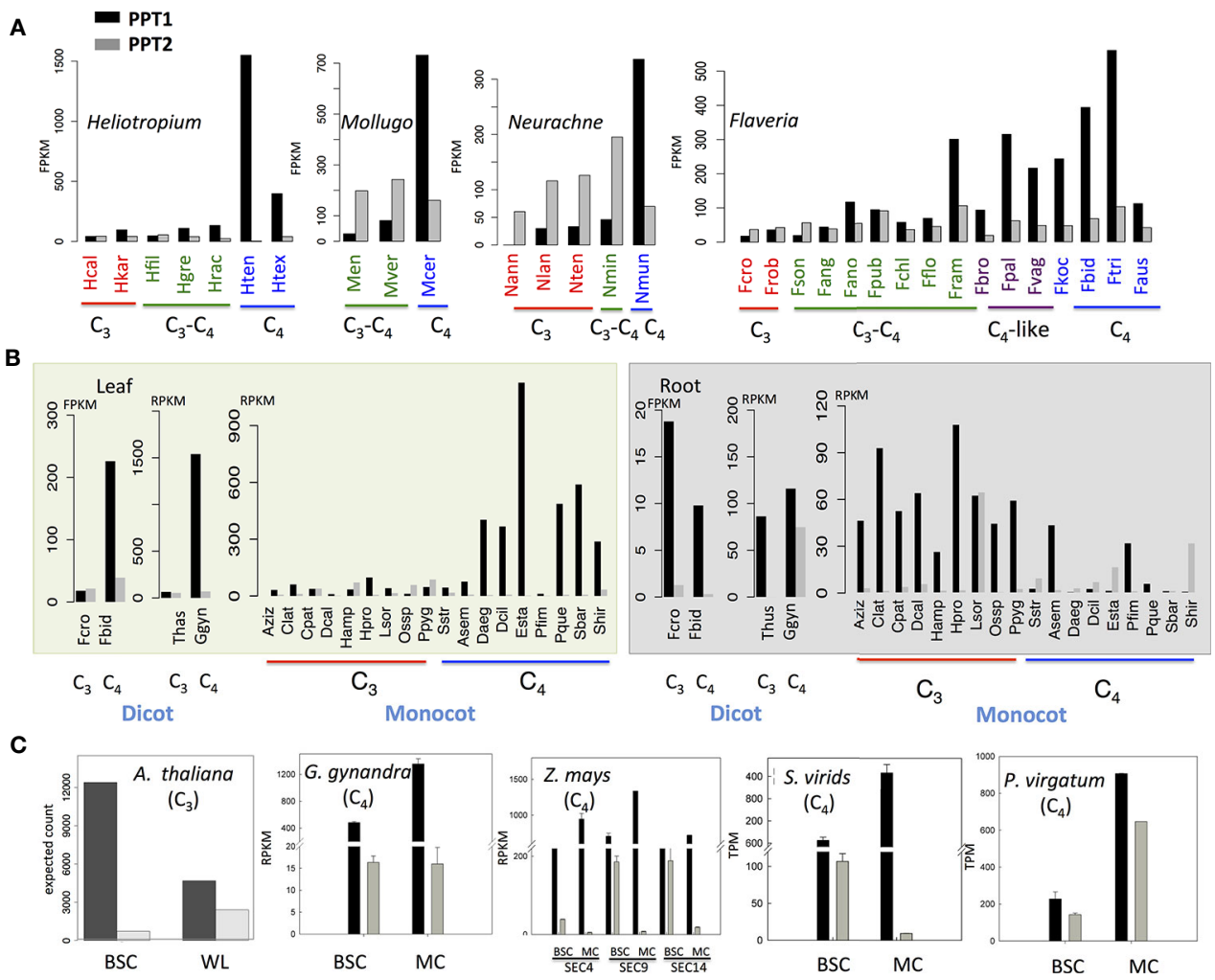

FIGURE 3 | The evolution of transcript abundance of PPT1 and PPT2 from C_3 to C_4 species **(A)** The transcript abundance of PPT1 and PPT2 along the evolution of C_4 photosynthesis in four genera. Photosynthetic types are marked with different colors, red: C_3; green: C_3–C_4; purple: C_4-like; blue: C_4. **(B)** The transcript abundance of PPT1 and PPT2 in leaf and root. **(C)** The transcript abundances of PPT1 and PPT2 in BSC and whole leaf or MC in one C_3 species (*Arabidopsis thaliana*) and four C_4 species. All the data are from published RNA-seq data; the data source is detailed in the *Materials and Methods* section. (Abbreviations: BSC, bundle sheath cell; WL, whole leaf; MC, mesophyll cell. Species abbreviations are listed in **Table S1**.)

well as in C_3–C_4 species in *Heliotropium,* i.e., *Heliotropium filiforme* (Hfil) and *Flaveria* i.e., *F. sonorensis* (Fson) (**Figure 3A**). During the transition from C_3 to C_4 photosynthesis, however, we observed an increase of transcript abundance in PPT1 while the transcript abundance of PPT2 remained similar, resulting in a higher expression abundance of PPT1 compared to PPT2 in C_4 species.

Though PPT1 in leaves of C_3 species did not show higher transcript abundance than PPT2, for the dicot C_3 species, PPT1 showed higher transcript abundance than PPT2 in root (**Figure 3B**); the same pattern was also found in most monocot C_3 species with *Lasiacis sorghoidea* (Lsor) as an exception (**Figure 3B**). In C_4 monocot species, the transcript abundance of PPT1 was not always higher than that of PPT2 in root; furthermore, the PPT expression levels in root were generally lower in C_4 as compared to C_3 species (**Figure 3B**). Therefore, PPT1, the copy recruited to support C_4 photosynthesis, did not have higher expression levels than PPT2 in the leaf tissue of C_3 plants; however, the gene had higher transcript abundance than PPT2 in root. During the evolution of C_4 photosynthesis, the transcript abundance of PPT1 was decreased in root and increased in leaf, implying a major shift in tissue specificity.

Considering that C_4 photosynthesis occurs in two cell types, which is a major evolutionary innovation, we further examined the changes in cellular specificity of PPT expression during C_4 evolution. For this purpose, we compared the transcript abundance of PPT1 and PPT2 in BSC and whole leaf in one C_3 species and that of BSC and MC in four C_4 species (**Figure 3C**). RNA-seq data from transcript residency on ribosomes (Aubry et al., 2014) shows that PPT1 had a higher expression level in BSC than in the whole leaf in *A. thaliana*, whereas PPT2 displayed the opposite pattern, which is consistent with earlier histochemical localization of the PPT promoter (Knappe et al., 2003): PPT1 localized in BSC and root, especially in root tip, while PPT2 localized in MC. In all C_4 species examined in this study, the transcript abundance of PPT1 was consistently higher in MC than in BSC (**Figure 3C**). In contrast, PPT2 showed no clear cell type specificity between the two cell types (**Figure 3C**). Therefore, during the evolution of C_4 photosynthesis, PPT1 shifted its cellular specificity from dominantly BSC to dominantly MC.

We also examined the expression patterns of PPTs based on transcriptomic data available in GENEVESTIGATOR (March, 2018), which includes four C_3 species, i.e., *A. thaliana*, *Oryza sativa*, *Solanum lycopersicum*, and *Glycine max*, and two C_4 species, i.e., *Z. mays* and *Sorghum bicolor*. We investigated the expression with a focus on developmental scale, in which the average was calculated from samples at the same development stage regardless of tissue type and cell type, and on the scale of cell type. PPT1 showed higher expression than PPT2 in general based on developmental stage (**Figure S2**). In C_3 species, either PPT2 or PPT1 together with PPT2 showed high expression in leaf, whereas PPT1 was dominant in root, with an exception in rice, in which PPT1 and PPT2 had comparable transcript levels (**Figure S2**). The dominant role of PPT1 was more obvious in root tip in C_3 species. In C_4 species, the expression patterns of PPT1 and PPT2 switched between leaf and root and between MC and BSC, which is in line with the above results.

The Changes in Transcriptional Regulation of Phospho*enol*pyruvate Transporter During Evolution From C_3 to C_4 Photosynthesis

The mechanism by which PPT1 gained new expression patterns to support C_4 photosynthesis, e.g., shifting its tissue specificity from primarily in root to primarily in leaf, and shifting its cellular specificity from predominantly in BSC to predominantly in MC is unknown. Examination of the expression patterns of PPTs between BSC and MC in four segments of maize shows that PPT1 has higher transcript abundance in MC than in BSC (**Figure 3C**). Given that the leaf MC typically receives more light than BSC (Xiao et al., 2016), one possibility is that the C_4 PPT1 might have acquired light-responsive *cis*-elements, which enables PPT1 to show light-dependent transcript accumulation patterns. To test the possibility, we first examined the light responsiveness of PPT1 and PPT2 along the C_4 phylogeny. Specifically, we compared the transcript abundance of PPT1 and PPT2 in mature leaves after 0, 0.5, 2, and 4 h of illumination. We quantified the transcript abundance using qRT-PCR in five *Flaveria* species, representing different photosynthetic types, *i.e.*, C_3 photosynthesis, *F. robusta*; type I C_3–C_4 species, *F. sonorensis*; type II C_3–C_4 species, *F. ramosissima*; and C_4 species, *F. trinervia* and *F. australasica* (Sage et al., 2012) (**Figure 4**). Our results demonstrated a gradual increase in the speed of changes of PPT1 transcript abundance to light from C_3 to C_3–C_4 intermediate to C_4 species. Specifically, the transcript abundance of PPT1 did not show significant up-regulation ($P < 0.05$, *t*-test) until 4 h under illumination in the C_3 *F. robusta*, whereas significant up-regulation of PPT1 transcript abundance was observed at 2 h in C_3–C_4 species. In the C_4 species *F. australasica*, the transcript abundance of PPT1 was up-regulated at 0.5 h under illumination with marginal significance ($P=0.075$, *t*-test). Therefore, during C_4 evolution, PPT1 acquired new mechanisms enabling it to be rapidly up-regulated upon illumination.

We further examined the patterns of increase in transcript abundance of PPT upon illumination change along the evolution from C_3 to C_4 species. Type I C_3–C_4 species showed the maximal PPT1 transcript abundance at 2 h under illumination, while the transcript abundance of PPT1 in type II C_3–C_4 and C_4 species kept increasing even after 4 h under illumination (**Figure 4**). The light responsiveness of PPT2 showed an opposite pattern as compared to PPT1 along C_4 evolution. Specifically, in the C_3 *F. robusta*, PPT2 showed significantly higher transcript abundance than PPT1. An up-regulated expression level of PPT2 in *F. robusta* was observed at 0.5 h under illumination, and a further increase was observed until 2 h. Nevertheless, in both C_3–C_4 species and the C_4 species *F. trinervia*, significantly increased expression of PPT2 was not detected until 2 h under illumination. Although PPT2 was induced at 0.5 h under light in C_4 species *F. trinervia*, in another C_4 species *F. australasica*, the transcript abundance of PPT2 showed no significant up-regulation under the illumination. Therefore, during evolution

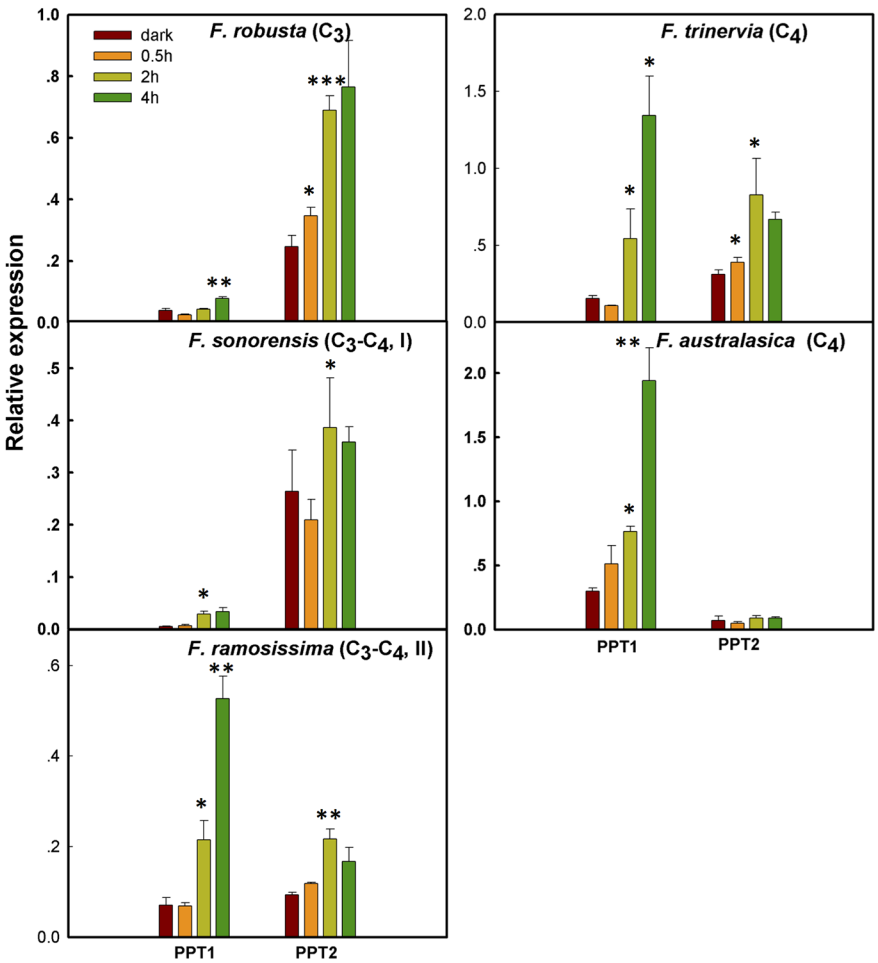

FIGURE 4 | The change in light responsiveness of PPT1 and PPT2 along the evolution of C_4 photosynthesis in the genus *Flaveria* Real-time quantitative (qRT)-PCR was used to quantify the transcript abundance of PPT1 and PPT2 in mature leaves after 0, 0.5, 2, and 4 h upon illumination. Significance levels represent the significance of the differences between the transcript abundance at a time point compared to that at the preceding time point (*t*-test; *: 0.05–0.01, **: 0.01–0.001, ***: < 0.001).

PPT1 gained not only higher transcript abundance in leaf, in particular in the MC, but also a more rapid and long-lasting response to light illumination, while PPT2 gradually lost its light responsiveness.

C_4 PPT1 Promoter Acquired MEM1 B Submodule But Not in C_3 and C_3–C_4 Species

Changes in transcriptional responses to external stimuli can be driven by changes in gene regulatory mechanisms. We tested whether C_4 PPT1 might have acquired new *cis*-elements that are responsible for the altered expression patterns. Based on the draft genome sequences of four *Flaveria* species, we found that there are two copies of PPT1 in *F. ramosisisma* (C_3–C_4, II) and *F. trinervia* (C_4), and one copy in *F. robusta* (C_3) and *F. sonorensis* (C_3–C_4, I). The promoter sequences (3 kbp upstream of the start codon) of two *F. trinervia* PPT1 are same, however, only one of these two copies was expressed. This copy also showed the highest expression level among the four species (with a TPM of 2,053); we name this copy *PPT1A* (**Figure 5A**). In *F. ramosissima*, one of the two copies of PPT1 has no intron and showed very low transcript abundance with a TPM of 1 (*PPT1B*), while the TPM of another copy was 272 (*PPT1A*). Moreover, the promoter sequences of the two PPT1s from *F. ramosissima* are not conserved with a sequence identity of only 11%. For PPT2, all species have one copy, which also show comparable transcript abundance in the four species (**Figure 5A**). In terms of genomic structure, both PPT1 and PPT2 have nine exons in most *Flaveria* species and *A. thaliana*, with the exception that in *F. ramosissima* there are eight exons in PPT1 and six exons in PPT2 (**Figure 5A**).

Further examination of the promoter structure shows that there is a highly conserved region between the proximal region of PPT1 promoter from *F. trinervia* (–2,325 to –1 bp upstream from the start codon) and that from *F. ramosissima*. The conserved region

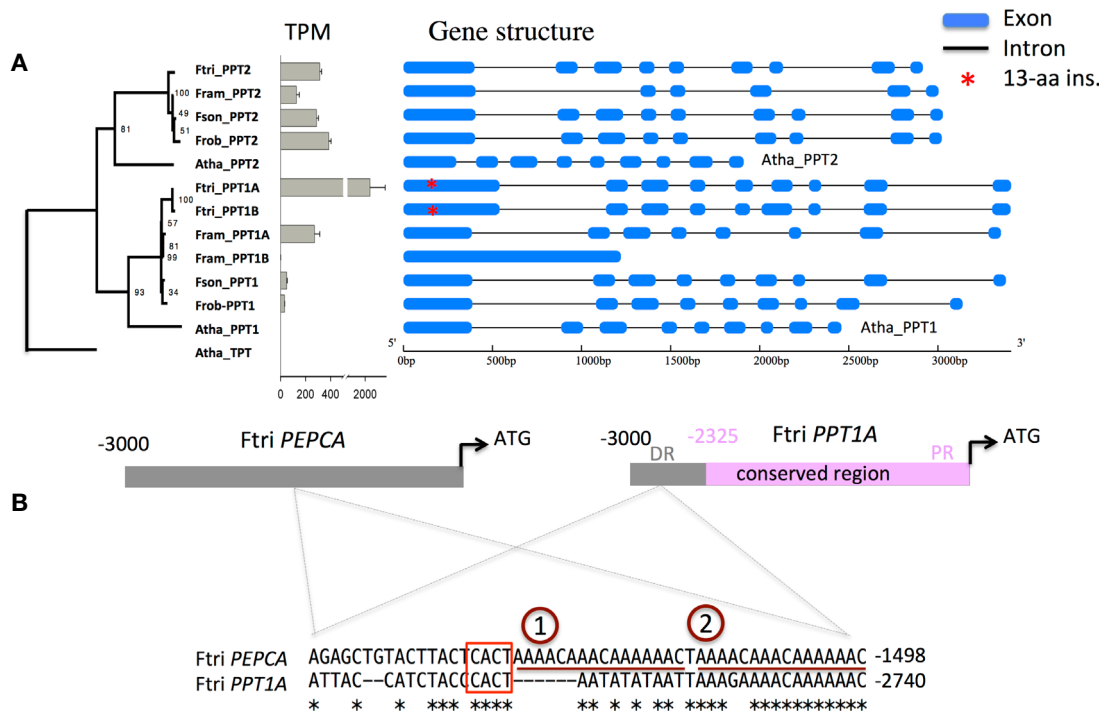

FIGURE 5 | Promoter of C_4 PPT1 has a MEM1 B submodule that might have contributed to the increased transcript abundance of PPT1 **(A)** Shows transcript abundance and gene structure of PPT1 and PPT2 in four *Flaveria* species. The gene tree of PPTs was constructed based on the alignment of protein sequences of PPT1 and PPT2 using a maximum likelihood method. Numbers on each node show bootstrap scores from 1,000 independent simulated trees. Transcript levels of *Flaveria* PPT1 and PPT2 based on RNA-seq data were shown in Transcript per kilobase Per Million mapped reads (TPM) in grey bars with error bar showing standard error estimated based on three replicates. *Flaveria trinervia* and *Flaveria ramosissima* have two copies of PPT1 but only one of the copies has TPM higher than 2. The 4*13-aa insertion occurred at the first exon of PPT1 in *F. trinervia* and is shown as red star. **(B)** The MEM1 B submodule of *PEPCA* gene promoter from *F. trinervia* is also present in the promoter of PPT1 from *F. trinervia*. The sequence alignment shows the MEM1 B submodule of the *PEPCA* from *Flaveria rinervia*, which has two copies of "AAAACAAACAAAAAC." Asterisks represent identical nucleotides. (Abbreviations: Ftri, *F. trinervia*; Fram, *F. ramosissima*; Fson, *F. sonorensis*; Frob, *F. robusta*; Atha, *A. thaliana*; DR, distal region; PR, proximal region.)

was divided into two parts by an insertion in *F. ramosissima*. Moreover, the two conserved parts were also observed in the promoters of PPT1 from *F. robusta* and *F. sonorensis* (**Supplemental File 3**). We further found a mesophyll expression module (MEM1) B submodule at the distal region (−2,783 to −2740 bp from the start codon) of the PPT1 promoter from *F. trinervia*, but this MEM1 B submodule was not present either in the counterpart in PPT1 from the other three *Flaveria* species, or the counterpart in PPT2 from any tested *Flaveria* species (**Figure 5B** and **Supplemental File 3**). The MEM1 B submodule is responsible for the increased expression level of the *PEPCA* (Akyildiz et al., 2007) and the *CA3* in *Flaveria* C_4 species, therefore, the same MEM1 B *cis*-element in the promoter of C_4 PPT1 might confer its higher expression. However, neither submodule A of MEM1, which is present in *PEPCA* (Akyildiz et al., 2007), nor submodule A of MEM1-like, which is present in *CA3* (Akyildiz et al., 2007), was observed in the C_4 *PPT1A* promoter. It is possible that other *cis*-element(s) that have same function with the A submodule may also be present in the promoter of C_4 PPT1 to realize the specific expression of PPT1 in MC.

Changes in the Physiological Functions of Phospho*enol*pyruvate Transporters During C_4 Evolution

We finally examined the functional changes of PPTs between C_3 and C_4 species. Usually the functional changes of a protein are underlined by changes in the amino acid sequence. Here we examined the changes in amino acid sequences of PPT1 and PPT2 from 16 species in the genus *Flaveria*, which covers C_3, C_3–C_4, C_4-like and C_4 species. Because of a lack of genome reference for some *Flaveria* species, the protein sequences of PPT1 and PPT2 were predicted based on *de novo* assembled transcripts for those species, and the genes that showed the highest sequence similarity with PPT1 and PPT2 from *F. robusta* were selected for comparison; therefore, only one copy of PPT1 and PPT2 from each *Flaveria* species were compared. We specifically examined the number of consistent amino acid modifications, which were defined as sites that have the same amino acid sequences in C_4 species but differ with those in C_3 species. The results show that PPT1 had more consistent amino acid modifications than PPT2 when the sequences from C_4 and C_3 species were compared.

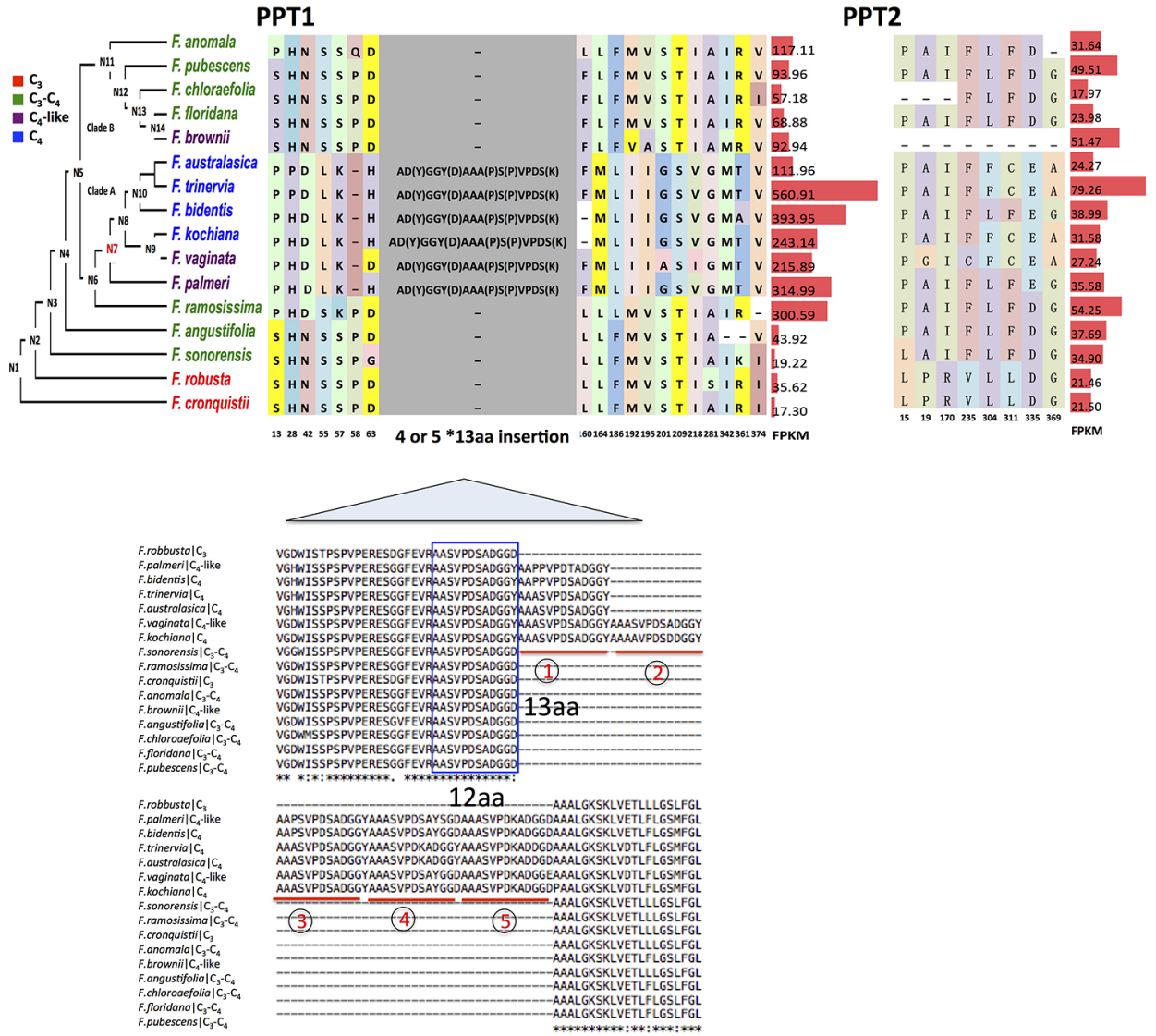

FIGURE 6 | The comparison of PPT1 and PPT2 protein sequences in the genus *Flaveria* Amino acid changes of PPT1 and PPT2 among different species in *Flaveria* genus are shown on the phylogeny. Transcript abundance calculated as Fragments Per Kilobase of transcript per Million reads mapped (FPKM) is displayed on the right with red bars. Numbers below amino acids are the aligned locations. The "-" symbols show an alignment gap. PPT1 (406 amino acids in *Flaveria cronquistii*) shows more frequent amino acid changes than PPT2 (417 amino acids in *F. cronquistii*). An insertion composed of four or five 13-aa elements occurs at the ancestral node "N7" (marked in red) on the phylogenetic tree. The sequence of the 13-aa segment is variant AAA(P)SVPDS(K)AD(Y)GGY(D) at four sites. A 12-aa element (blue frame) at the N-terminal end of the insertion is present in all species, which should be the origin of the 13-aa element.

Specifically, the amino acid sequence of PPT1 had 19 consistent amino acid modifications between C_3 and C_4 species; in contrast, PPT2 exhibited eight consistent amino acid modifications (**Figure 6**). To test whether these modifications were specific adaptations gained during evolution of C_4 photosynthesis, we performed a positive selection test in protein coding sequences of C_4 species against that of C_3 species in the genus *Flaveria*. PPT2 showed a signal of positive selection in C_4 species; however, the two predicted positive selected sites of PPT2 were neither C_4 specific nor C_4 consistent modifications (**Figure S3**). In contrast, PPT1 showed no signal of positive selection in C_4 species, suggesting that the consistent mutations observed in C_4 species may occur by chance during evolution. Though we did not identify any particular amino acid sequence under positive selection, PPT1 however showed a large insertion acquired at the common ancestor of C_4-like and C_4 species in clade A. The insertion segments had either four (*F. palmeri*, *F. bidentis*, *F. trinervia*, and *F. australasica*) or five (*F. vaginata* and *F. kochiana*) repeats with each repeat comprising 13 amino acids, i.e., a 13-aa element (**Figure S4**).

To investigate whether the 13-aa element is also present in PPT1 of other C_3 and C_4 species, we compared amino acid

sequences of PPT1 from other C_3 and C_4 species, including three C_3 species, namely, *A. thaliana*, *T. hassleriana* and *F. robusta*, and six C_4 species, *F. bidentis*, *F. trinervia*, *G. gynandra*, *S. bicolor*, *Z. mays*, and *S. italica*. The alignment shows that the 13-aa-element insertion is only present in *Flaveria* C_4 species (**Figure S5**). We determined that the insertion might have been generated by slipping mispairing during DNA synthesis as reported in *Z. mays* (Wessler et al., 1990) (**Figure S6**). We found that the 12-aa segment in *F. ramosissima* missed one alanine at the N-terminal end compared to the 13-aa segment (**Figure 6**). The DNA sequence encoding the 12-aa segment could form a stable hairpin structure (**Figure S6A**). Coincidently, there is a triplet alanine following the C-terminal end of the 12-aa segment (**Figure 6** and **Figure S6A**). In the coding sequence of the 12-aa segment and the triplet alanine (termed 15-aa segment), a 6-bp nucleotide sequence "GCGGCG" appears both at the head and the tail of the 15-aa segment. It is possible that the hairpin structure may shorten the distance between the "GCGGCG" at the 5′ end and "GCGGCG" at the 3′ end (**Figure S6B**), which facilitated "slipping mispairing" during DNA synthesis resulting in formation of the 4x13-aa insertion (**Figure S6C**).

Given these changes of amino acid sequences in PPT1 during C_4 evolution, we tested whether the function of PPT1 was conserved between C_3 and C_4 species using a genetic approach by expressing *Flaveria* PPT1 in a C_3 *A. thaliana* PPT1 loss-of-function mutant *cue1-5* (Li et al., 1995). Specifically, we expressed PPT1-GFP driven by a *35S* promoter in *cue1-5* through gene engineering (**Figure 7A**). The PPT1 used was from four different *Flaveria* species, including one C_3 species *Flaveria cronquistii*, two intermediate species *F. ramosissima* (C_3–C_4) and *F. palmeri* (C_4-like), and one C_4 species *F. bidentis* (C_4). The results showed that the PEP transporting function of PPT1 from all four species complemented the reticulate leaf phenotype and small rosette size of *cue1-5*

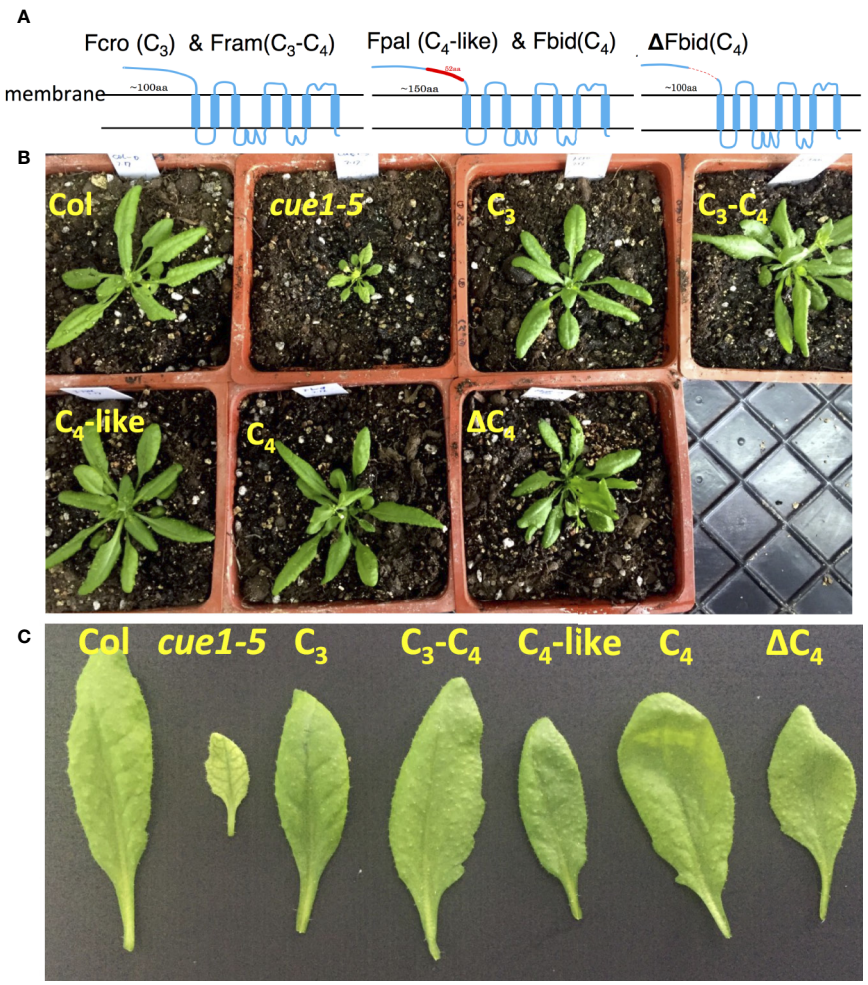

FIGURE 7 | *Flaveria* PPT1 complements the phenotype of *A. thaliana* PPT1 loss-of-function mutant *cue1-5* **(A)** The position of a 52-aa insertion in protein sequence of *Flaveria bidentis* (C_4) PPT1. The insertion is predicted to be located at the non-membrane-portion. **(B, C)** *Arabidopsis thaliana cue1-5* shows reticulate leaf phenotype and decreased rosette size; PPT1 from different photosynthetic types of *Flaveria* species rescues the phenotype of *A. thaliana cue1-5*. PPT1 from *F. bidentis* without the insertion (ΔC_4) also recovers the phenotype of the *A. thaliana cue1-5*. (Abbreviations: Fcro, *F. cronquistii*; Fram, *F. ramosissima*; Fpal, *F. palmeri*; Fbid, *F. bidentis*.)

(**Figures 7B, C**), indicating that PPT1 was functionally conserved in leaf in these different *Flaveria* species and *A. thaliana*. One caveat is that the 35S promoter might not be sensitive enough to detect potential differences in the affinities of PPT1 from different species. So, there still might be physiological significance of the altered amino acids in the PPT1 sequence, which needs detailed enzymatic studies to elucidate.

Given that there exists an insertion in the C_4 and C_4-like species in the *Flaveria* genus in the protein-coding region of PPT1, we further tested whether this insertion affected the function of PPT1. We explored this question by removing the 4x13-aa insertion in *F. bidentis* (C_4) (ΔFbidPPT1 for short) and expressing it in *cue1-5 A. thaliana* (**Figure 7A**). The transgenic plant *ΔFbidPPT1*/*cue1* showed the same phenotype as *FbidPPT1*/*cue1* (**Figures 7B, C**), suggesting that the insertion had no effect on the PEP transport function of PPT1 in C_3 leaf. It was likely, therefore, that this extra insertion had no influence on the structure of PPT1 in the thylakoid membrane. Indeed, protein structure prediction using I-TASSER showed that the insertion site lies in the outer membrane portion of FbidPPT1 (**Figure 7A**), which might not influence the functional path required for PEP transport in PPT1 in the thylakoid membrane.

Considering that coding sequences can potentially harbor *cis*-elements responsible for cell specificity, as in NAD-ME in *G. gynandra* (Brown et al., 2011), we further checked whether the extra insertion affects the subcellular location of PPT1 by using transient expression of PPT1-GFP in tobacco leaves. Transient transgenic experiments showed that both FbidPPT1 and ΔFbidPPT1 were localized in chloroplasts (**Figures S7A–C**), suggesting that the insertion had no impact on the subcellular localization of PPT1. Further experiments showed that PPT1 and PPT2 from both C_3 and C_4 *Flaveria* species were localized to chloroplast (**Figure S7**). Therefore, all the results from sequence analysis and functional tests based on transgenic experiments suggested that though there were major changes in the amino acid sequences in PPT1 during C_3 to C_4 evolution, these changes neither changed the PEP transporting function of PPT1, nor altered the localization of PPT1. The potential role of these changes may be involved in the transport efficiency of PPT1, which needs more detailed biochemical studies.

DISCUSSION

This study presented a comparative survey of PPT, one of the metabolite transporters involved in C_4 photosynthesis. The analysis showed that though PPT1 had lower transcript abundance in leaf compared to PPT2, it was recruited to support C_4 photosynthesis in multiple C_4 lineages. During C_4 evolution, PP1 switched its expression from predominantly in root to in leaf and from predominately in BSC to in MC; it also acquired increased responsiveness of expression to light induction, which might be related to a newly recruited MEM1 B submodule in the PPT1 promoter in the *Flaveria* C_4 species. PPT1 also shows major changes in amino acid sequences during C_4 evolution, though they do not change the PEP transporting function. In this section, we discuss these findings in terms of their implications for C_4 photosynthesis.

Factors Contributing to Recruitment of PPT1 Instead of PPT2 for C_4 Function

Potential of high expression: studies on the evolution of C_4 genes identified a number of properties associated with the recruited paralogs for C_4 function, which include relatively high expression levels (Moreno-Villena et al., 2018), availability of gene copies, which provide a fast route to increase gene expression (Bianconi et al., 2018), and suitable enzyme catalytic properties *via* accumulated mutations in the coding region (Christin et al., 2013). In the case of PPT, in terms of transcript abundance, though PPT1 had lower transcript abundance than PPT2 in leaves, PPT1 had very high transcript abundance in root, especially in the root tip (**Figure 3** and **Figure S2**). Based on the data from GENEVESTIGATOR, the total transcript abundance of PPT1 was higher than PPT2 regardless of tissue type (**Figure S2**), suggesting that some pre-existing regulatory mechanism can confer higher transcript abundance of PPT1 than PPT2. On the other hand, we found that PPT1 had a higher or the same copy number with PPT2 in angiosperms (without considering *A. trichopoda*), as well as in *F. trinervia*. Having more gene copies can free up one copy to acquire new regulatory or catalytic properties required for C_4 photosynthesis without jeopardizing its native role in C_3 plants. In fact, gene duplication and neofunctionalization have been recognized as major factors contributing to evolution of C_4 photosynthesis (Monson, 2003; Emms et al., 2016).

Protein properties: though PPT1 and PPT2 are functionally redundant in terms of their role of transporting PEP in *A. thaliana*, PPT1 shows lower specificity to PEP and higher permeability to 2-phosphoglycerate, i.e., there are differences in protein properties between PPT1 and PPT2 (Knappe et al., 2003). Here we observed more amino acid changes in PPT1 than PPT2 during evolution in *Flaveria* (**Figure 6**). We also found a larger insertion with either four or five repeated 13-aa elements in C_4-like and C_4 species in *Flaveria* clade A species, but this element was not observed in PPT1 from other C_4 species, such as *G. gynandra* and *Z. mays* (**Figure S5**). Accumulation of new mutations may contribute to the suitability of PPT1 for C_4 photosynthesis. The observation that PPT1 from either C_3 or C_4 *Flaveria* rescued the reticulate leaf phenotype of *cue1* showed that the PEP transport function of PPT1 was not altered between C_3 and C_4 plants (**Figure 7**). In the future, more detailed functional studies of catalytic properties of PPT1 from different C_3 and C_4 species are needed to test whether the acquired amino acid modifications in PPT1 have particular sequence variations that make it more suitable to function in a C_4 context, e.g., increased specificity or transport rate.

Mechanisms Underlying Establishment of New Expression Patterns of PPT1

Gaining new cis-elements: the function of C_4 photosynthesis requires high expression of required genes upon light induction in specific cell types (Sheen, 1999). Reports show that the expression levels of C_4 related genes are usually up-

regulated upon light induction in C_4 plants, while they are not necessarily up-regulated in C_3 plants, suggesting that the mechanisms controlling the light-induced of expression of C_4 related genes were acquired during C_4 evolution (Christin et al., 2013; Xu et al., 2016). Here we showed that, during C_4 evolution, PPT1 shifted its expression from predominantly in root to in leaf, and from predominantly in BSC to in MC (**Figures 3B, C**). We also found that PPT1 gained faster and long-lasting light induction during evolution (**Figure 4**). Moreover, we found that PPT1 gained a MEM1 B submodule (**Figure 5**) in its promoter region, which controls the high transcript abundance of *PEPCA* and *CA3* in MC in C_4 *Flaveria* species (Akyildiz et al., 2007; Gowik et al., 2017). Notably, a similar MEM1 B submodule was also identified at 1,500 bp upstream of the start codon of PPDK (**Supplemental File 3**).

Considering that PPDK catalyzes generation of PEP, while PEPC uses PEP as its substrate, possessing a common *cis*-element in PPT1, PEPC, and PPDK enables coordinated up-regulation of expression of these proteins for efficient C_4 photosynthesis. This re-utilization of shared *cis*-regulatory elements can be realized through transposon-mediated movement of *cis*-elements between genes (Cao et al., 2016). Interestingly, in an earlier study of the origin of crassulacean acid metabolism (CAM), Zhang et al. (2016) showed that the expression levels of PEPC, PEPC kinase, and PPDK are strongly co-regulated during day and night, and thus might have gained common transcriptional regulatory mechanisms enabling them to be co-recruited to support CAM. Therefore, co-recruiting common *cis*-regulatory elements might have played critical roles during the evolution of both C_4 and CAM.

Using root regulatory mechanisms: PPT1 showed higher transcript abundance in root in C_3 species, and it is possible that the increased transcript abundance of PPT1 in leaf may be a result of using pre-presenting regulatory mechanisms in root, including both *cis*- and *trans*-elements, as illustrated in **Figure 8**. Kulahoglu et al. (2014) showed that a module in root includes genes with high transcript abundance in *T. hassleriana*, a C_3 species in the Cleomaceae family, while at the same time they show high transcript abundance in leaves of the C_4 species *G. gynandra*, which is from the same family (Kulahoglu et al., 2014). This root module was recruited to support C_4 with carbonic anhydrase and DIC1. The same scenario might underlie the increased transcript abundance of PPT1 in C_4 leaf.

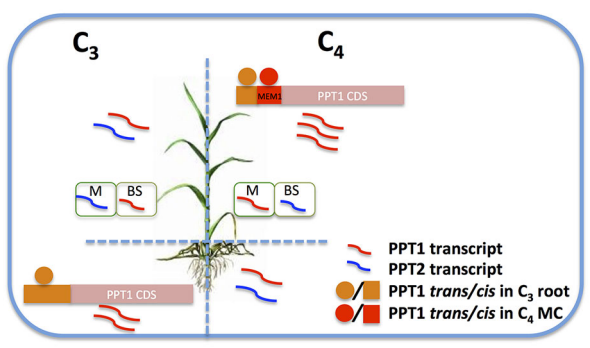

FIGURE 8 | A hypothesis of the increased transcript abundance of PPT1 in C_4 leaf MC. In C_3 species, PPT1 (red curved line) shows predominant expression in root and leaf BSC, while PPT2 (blue curved line) predominates in leaf MC. PPT1 was recruited to C_4 photosynthesis by gaining a dramatic increase of the transcript level in leaf MC, resulting in a switched transcript pattern compared to C_3 species in terms of leaf and root, MC and BSC. The enhanced transcript abundance of PPT1 in C_4 leaf MC may be a result of using *trans*-elements (red circle) in C_4 MC by recruiting new *cis*-elements (red rectangle) such as MEM1 B submodule, or/and using the original regulatory mechanisms presented in C_3 root (orange circle and rectangle). (Abbreviations, MC, mesophyll cell; BSC, bundle sheath cell.)

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: GSE54339 for Lyu et al. (2018), SRP036637 and SRP036837 for Kulahoglu et al., BioProject PRJNA395007 for Moreno-Villena et al. 2018, SRA066236 for Aubry et al. (2014), SRA012297 for Li et al. (2010), GSE54272 for Tausta et al. (2014), SRP052802 for Denton et al. (2017), SRP009063 for Chang et al. (2012), PRJEB5074 for John et al. (2014), and SRX1160366 for Rao et al. (2016).

AUTHOR CONTRIBUTIONS

X-GZ, GC and M-JL designed the project and wrote the paper. M-JL did bioinformatics analysis and qRT-PCR. YW, JJ, and XL conducted the transgenic experiments.

FUNDING

This work was sponsored by Shanghai Sailing Program (17YF421900), Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB27020105), National Science Foundation of China (31870214, 31701139, 31500988), National Research and Development Program of Ministry of Science and Technology of China (2019YFA0904600).

ACKNOWLEDGMENTS

The authors appreciate Prof. Rowan F. Sage and Prof. Peter Westhoff for sharing us *Flaveria* materials. We thank Prof. Daiyin Chao and Prof. Haiyang Hu for great discussion and suggestion; M-JL also thanks Faming Chen, Fengfeng Miao and Yongyao Zhao for help in experiment parts. This manuscript has been released as a preprint at bioRxiv (Lyu et al., 2019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00935/full#supplementary-material

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. doi: 10.1105/tpc.107.053322

Aubry, S., Brown, N. J., and Hibberd, J. M. (2011). The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway. *J. Exp. Bot.* 62, 3049–3059. doi: 10.1093/jxb/err012

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. doi: 10.1371/journal.pgen.1004365

Bailey, K. J., and Leegood, R. C. (2016). Nitrogen recycling from the xylem in rice leaves: dependence upon metabolism and associated changes in xylem hydraulics. *J. Exp. Bot.* 67, 2901–2911. doi: 10.1093/jxb/erw132

Bianconi, M. E., Dunning, L. T., Moreno-Villena, J. J., Osborne, C. P., and Christin, P. A. (2018). Gene duplication and dosage effects during the early emergence of C_4 photosynthesis in the grass genus Alloteropsis. *J. Exp. Bot.* 69, 1967–1980. doi: 10.1093/jxb/ery029

Blasing, O. E., 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. doi: 10.1074/jbc.M909832199

Brown, N. J., Newell, C. A., Stanley, S., Chen, J. E., Perrin, A. J., Kajala, K., et al. (2011). Independent and parallel recruitment of preexisting mechanisms underlying C4 photosynthesis. *Science* 331, 1436–1439. doi: 10.1126/science.1201248

Burgess, S. J., Granero-Moya, I., Grange-Guermente, M. J., Boursnell, C., Terry, M. J., and Hibberd, J. M. (2016). Ancestral light and chloroplast regulation form the foundations for C4 gene expression. *Nat. Plants* 2. doi: 10.1038/nplants.2016.161

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinf.* 10, 421. doi: 10.1186/1471-2105-10-421

Cao, C., Xu, J., Zheng, G., and Zhu, X. G. (2016). Evidence for the role of transposons in the recruitment of cis-regulatory motifs during the evolution of C4 photosynthesis. *BMC Genomics* 17, 201. doi: 10.1186/s12864-016-2519-3

Chang, Y. M., Liu, W. Y., Shih, A. C., Shen, M. N., Lu, C. H., Lu, M. Y., et al. (2012). Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. *Plant Physiol.* 160, 165–177. doi: 10.1104/pp.112.203810

Christin, P. A., 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. doi: 10.3732/ajb.0900111

Christin, P. A., Boxall, S. F., Gregory, R., Edwards, E. J., Hartwell, J., and Osborne, C. P. (2013). Parallel recruitment of multiple genes into C4 photosynthesis. *Genome Biol. Evol.* 5, 2174–2187. doi: 10.1093/gbe/evt168

Denton, A. K., Mass, J., Kulahoglu, C., Lercher, M. J., Brautigam, A., and Weber, A. P. (2017). Freeze-quenched maize mesophyll and bundle sheath separation uncovers bias in previous tissue-specific RNA-Seq data. *J. Exp. Bot.* 68, 147–160. doi: 10.1093/jxb/erw463

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238. doi: 10.1186/s13059-019-1832-y

Emms, D. M., Covshoff, S., Hibberd, J. M., and Kelly, S. (2016). Independent and parallel evolution of new genes by gene duplication in two origins of C4 photosynthesis provides new insight into the mechanism of phloem loading in C_4 species. *Mol. Biol. Evol.* 33, 1796–1806. doi: 10.1093/molbev/msw057

Fischer, K., Kammerer, B., Gutensohn, M., Arbinger, B., Weber, A., Hausler, R. E., et al. (1997). A new class of plastidic phosphate translocators: A putative link between primary and secondary metabolism by the phosphoenolpyruvate/phosphate antiporter. *Plant Cell* 9, 453–462. doi: 10.1105/tpc.9.3.453

Gowik, U., and Westhoff, P. (2011). The path from C_3 to C_4 photosynthesis. *Plant Physiol.* 155, 56–63. doi: 10.1104/pp.110.165308

Gowik, U., Burscheidt, J., Akyildiz, M., Schlue, U., Koczor, M., Streubel, M., et al. (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. doi: 10.1105/tpc.019729

Gowik, U., Schulze, S., Saladie, M., Rolland, V., Tanz, S. K., Westhoff, P., et al. (2017). A MEM1-like motif directs mesophyll cell-specific expression of the gene encoding the C4 carbonic anhydrase in Flaveria. *J. Exp. Bot.* 68, 311–320. doi: 10.1093/jxb/erw475

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Gupta, S. D., Levey, M., Schulze, S., Karki, S., Emmerling, J., Streubel, M., et al. (2020). The C_4 Ppc promoters of many C_4 grass species share a common regulatory mechanism for gene expression in the mesophyll cell. *Plant J.* 101, 204–216. doi: 10.1111/tpj.14532

Hatch, M. D., and Osmond, C. B. (1976). Compartmentation and transport in C_4 photosynthesis. *Encyclopedia Plant Physiol.* 3, 144–184. doi: 10.1007/978-3-642-66417-5_5

Hatch, M. D. (1987). C_4 photosynthesis - a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochim. Et Biophys. Acta* 895, 81–106. doi: 10.1016/S0304-4173(87)80009-5

Herrmann, K. M., and Weaver, L. M. (1999). The shikimate pathway. *Annu. Rev. Plant Biol.* 50, 473–503. doi: 10.1146/annurev.arplant.50.1.473

Hibberd, J. M., and Covshoff, S. (2010). The regulation of gene expression required for C_4 photosynthesis. *Annu. Rev. Plant Biol.* 61, 181–207. doi: 10.1146/annurev-arplant-042809-112238

Jiao, Y., Li, J., Tang, H., and Paterson, A. H. (2014). Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. *Plant Cell* 26, 2792–2802. doi: 10.1105/tpc.114.127597

John, C. R., Smith-Unna, R. D., Woodfield, H., Covshoff, S., and Hibberd, J. M. (2014). Evolutionary convergence of cell-specific gene expression in independent lineages of C_4 grasses. *Plant Physiol.* 165, 62–75. doi: 10.1104/pp.114.238667

Johnson, M. T., Carpenter, E. J., Tian, Z., Bruskiewich, R., Burris, J. N., Carrigan, C. T., et al. (2012). Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLoS One* 7, e50226. doi: 10.1371/journal.pone.0050226

Knappe, S., Lottgert, T., Schneider, A., Voll, L., Flugge, U. I., and Fischer, K. (2003). Characterization of two functional phospho*enol*pyruvate/phosphate translocator (PPT) genes in Arabidopsis–AtPPT1 may be involved in the provision of signals for correct mesophyll development. *Plant J.* 36, 411–420. doi: 10.1046/j.1365-313X.2003.01888.x

Krebs, H. A. (1940). The citric acid cycle and the szent-gyogyi cycle in pigeon breast muscle. *Biochem. J.* 34, 775–779. doi: 10.1042/bj0340775

Kulahoglu, C., Denton, A. K., Sommer, M., Mass, J., Schliesky, S., Wrobel, T. J., et al. (2014). Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C3 and C4 plant species. *Plant Cell* 26 (8), 3243–3260. doi: 10.1105/tpc.114.123752

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* 12, 323. doi: 10.1186/1471-2105-12-323

Li, H. M., Culligan, K., Dixon, R. A., and Chory, J. (1995). *Cue1* - a mesophyll cell-specific positive regulator of light-controlled gene-expression in Arabidopsis. *Plant Cell* 7, 1599–1610. doi: 10.2307/3870022

Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S. L., et al. (2010). The developmental dynamics of the maize leaf transcriptome. *Nat. Gene.* 42, 1060–1067. doi: 10.1038/ng.703

Lyu, M. J., Gowik, U., Kelly, S., Covshoff, S., Mallmann, J., Westhoff, P., et al. (2015). RNA-Seq based phylogeny recapitulates previous phylogeny of the genus Flaveria (Asteraceae) with some modifications. *BMC Evolutionary Biol.* 15, 116. doi: 10.1186/s12862-015-0399-9

Lyu, M. J., Gowik, U., Westhoff, P., Tao, Y., Kelly, S., Covshoff, S., et al. (2018). The Coordination and Jumps along C_4 Photosynthesis Evolution in the Genus Flaveria. *bioRxiv* 460287. doi: 10.1101/460287

Lyu, M.-J. A., Wang, Y., Jiang, J., Chen, G., and Zhu, X.-G. (2019). Evolutionary changes of phosphoenolpyruvate transporter (PPT) during the emergence of C4 photosynthesis. *bioRxiv*. doi: 10.1101/713537

Mallmann, J., Heckmann, D., Brautigam, A., Lercher, M. J., Weber, A. P., Westhoff, P., et al. (2014). The role of photorespiration during the evolution of C4 photosynthesis in the genus Flaveria. *Elife* 3, e02478. doi: 10.7554/eLife.02478

Matasci, N., Hung, L. H., Yan, Z., Carpenter, E. J., Wickett, N. J., Mirarab, S., et al. (2014). Data access for the 1,000 Plants (1KP) project. *Gigascience* 3, 17. doi: 10.1186/2047-217X-3-17

Monson, R. K. (2003). Gene duplication, neofunctionalization, and the evolution of C_4 photosynthesis. *Int. J. Plant Sci.* 164, S43–S54. doi: 10.1086/368400

Moreno-Villena, J. J., Dunning, L. T., Osborne, C. P., and Christin, P. A. (2018). Highly expressed genes are preferentially co-opted for C4 photosynthesis. *Mol. Biol. Evol.* 35, 94–106. doi: 10.1093/molbev/msx269

Paulus, J. K., Schlieper, D., and Groth, G. (2013). Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution. *Nat. Commun.* 4, 1518. doi: 10.1038/ncomms2504

Rao, X., Lu, N., Li, G., Nakashima, J., Tang, Y., and Dixon, R. A. (2016). Comparative cell-specific transcriptomics reveals differentiation of C4 photosynthesis pathways in switchgrass and other C4 lineages. *J. Exp. Bot.* 67, 1649–1662. doi: 10.1093/jxb/erv553

Sage, R. F., and Zhu, X. G. (2011). Exploiting the engine of C4 photosynthesis. *J. Exp. Bot.* 62, 2989–3000. doi: 10.1093/jxb/err179

Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and evolution of C_4 photosynthesis. *Annu. Rev. Plant Biolog.* 63, 19–47. doi: 10.1146/annurev-arplant-042811-105511

Sheen, J. (1999). C_4 gene expression. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50. doi: 10.1146/annurev.arplant.50.1.187

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–W612. doi: 10.1093/nar/gkl315

Tausta, S. L., Li, P., Si, Y., Gandotra, N., Liu, P., Sun, Q., et al. (2014). Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C4-related processes. *J. Exp. Bot.* 65, 3543–3555. doi: 10.1093/jxb/eru152

Taylor, L., Nunes-Nesi, A., Parsley, K., Leiss, A., Leach, G., Coates, S., et al. (2010). Cytosolic pyruvate,orthophosphate dikinase functions in nitrogen remobilization during leaf senescence and limits individual seed growth and nitrogen content. *Plant J.* 62, 641–652. doi: 10.1111/j.1365-313X.2010.04179.x

Von Caemmerer, S., and Furbank, R. T. (2003). The C4 pathway: an efficient CO2 pump. *Photosynthesis Res.* 77, 191–207. doi: 10.1023/A:1025830019591

Wang, Y., Long, S. P., and Zhu, X. G. (2014). Elements Required for an Efficient NADP-Malic Enzyme Type C4 Photosynthesis. *Plant Physiol.* 164, 2231–2246. doi: 10.1104/pp.113.230284

Weber, A. P., and Von Caemmerer, S. (2010). Plastid transport and metabolism of C3 and C4 plants–comparative analysis and possible biotechnological exploitation. *Curr. Opin. Plant Biol.* 13, 257–265. doi: 10.1016/j.pbi.2010.01.007

Wessler, S., Tarpley, A., Purugganan, M., Spell, M., and Okagaki, R. (1990). Filler DNA Is Associated with Spontaneous Deletions in Maize. *Proc. Natl. Acad. Sci. United States America* 87, 8731–8735. doi: 10.1073/pnas.87.22.8731

West-Eberhard, M. J., Smith, J. A. C., and Winter, K. (2011). Photosynthesis, reorganized. *Science* 332, 311–312. doi: 10.1126/science.1205336

Westhoff, P., and Gowik, U. (2004). Evolution of C4 phosphoenolpyruvate carboxylase. Genes and proteins: a case study with the genus Flaveria. *Ann. Bot.* 93, 13–23. doi: 10.1093/aob/mch003

Williams, B. P., Burgess, S. J., Reyna-Llorens, I., Knerova, J., Aubry, S., Stanley, S., et al. (2016). An Untranslated cis-Element Regulates the Accumulation of Multiple C4 Enzymes in Gynandropsis gynandra Mesophyll Cells. *Plant Cell* 28, 454–465. doi: 10.1105/tpc.15.00570

Xiao, Y., Tholen, D., and Zhu, X. G. (2016). The influence of leaf anatomy on the internal light environment and photosynthetic electron transport rate: exploration with a new leaf ray tracing model. *J. Exp. Bot.* 67, 6021–6035. doi: 10.1093/jxb/erw359

Xu, J. J., Brautigam, A., Weber, A. P. M., and Zhu, X. G. (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. doi: 10.1093/jxb/erw275

Yang, J., and Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* 43, W174–W181. doi: 10.1093/nar/gkv342

Yang, Z. H. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Zhang, L. S., Chen, F., Zhang, G. Q., Zhang, Y. Q., Niu, S., Xiong, J. S., et al. (2016). Origin and mechanism of crassulacean acid metabolism in orchids as implied by comparative transcriptomics and genomics of the carbon fixation pathway. *Plant J.* 86, 545–545. doi: 10.1111/tpj.13236

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 03 July 2020
doi: 10.3389/fpls.2020.01014

Kinetic Modifications of C_4 PEPC Are Qualitatively Convergent, but Larger in *Panicum* Than in *Flaveria*

Nicholas R. Moody[1], *Pascal-Antoine Christin*[2] *and James D. Reid*[1]*

[1] Department of Chemistry, University of Sheffield, Sheffield, United Kingdom, [2] Department of Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom

C_4 photosynthesis results from a set of anatomical features and biochemical components that act together to concentrate CO_2 within the leaf and boost productivity. This complex trait evolved independently many times, resulting in various realizations of the phenotype, but in all C_4 plants the primary fixation of atmospheric carbon is catalyzed by phosphoenolpyruvate carboxylase. Comparisons of C_4 and non-C_4 PEPC from a few closely related species suggested that the enzyme was modified to meet the demands of the C_4 cycle. However, very few C_4 groups have been investigated, hampering general conclusions. To test the hypothesis that distant C_4 lineages underwent convergent biochemical changes, we compare the kinetic variation between C_4 and non-C_4 PEPC from a previously assessed young lineage (*Flaveria*, Asteraceae) with those from an older lineage found within the distantly related grass family (*Panicum*). Despite the evolutionary distance, the kinetic changes between the non-C_4 and C_4 PEPC are qualitatively similar, with a decrease in sensitivity for inhibitors, an increased specificity (k_{cat}/K_m) for bicarbonate, and a decreased specificity (k_{cat}/K_m) for PEP. The differences are more pronounced in the older lineage *Panicum*, which might indicate that optimization of PEPC for the C_4 context increases with evolutionary time.

Keywords: C_4 photosynthesis, carbon fixation, enzyme evolution, feedback inhibition, kinetics, phosphoenolpyruvate carboxylase

Edited by:
Sarah Covshoff, Independent Researcher, Las Vegas, NV, United States

Reviewed by:
Robert Edward Sharwood, Australian National University, Australia
Asaph Cousins, Washington State University, United States

***Correspondence:**
James D. Reid
j.reid@sheffield.ac.uk

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *22 March 2020*
Accepted: *22 June 2020*
Published: *03 July 2020*

Citation:
Moody NR, Christin P-A and Reid JD (2020) Kinetic Modifications of C_4 PEPC Are Qualitatively Convergent, but Larger in Panicum Than in Flaveria. Front. Plant Sci. 11:1014. doi: 10.3389/fpls.2020.01014

INTRODUCTION

C_4 photosynthesis is a CO_2-concentrating mechanism that boosts productivity in tropical conditions (Atkinson et al., 2016). The higher efficiency of C_4 plants results from the increased concentration of CO_2 around ribulose-1,5-bisphosphate carboxylase oxygenase (Rubisco), the entry enzyme of the Calvin–Benson cycle (Sage et al., 2012). Rubisco has a tendency to confuse CO_2 and O_2 (Tcherkez et al., 2006). The reaction of O_2 produces compounds that need to be recycled in the energetically costly photorespiration pathway (Nisbet et al., 2007). In C_3 plants, Rubisco is in direct contact with atmospheric gases, and photorespiration can become consequential in conditions that decrease the relative concentration of CO_2, including high temperature, aridity and salinity (Ehleringer and Bjorkman, 1977; Skillman, 2007). C_4 plants tackle this problem by segregating primary carbon fixation from the enzyme Rubisco into two cell types (Hatch, 1987; Sage, 2004; Sage et al., 2012). In C_4 plants, atmospheric CO_2 in the form of bicarbonate is initially fixed by the

enzyme phosphoenolpyruvate carboxylase (PEPC) (Hatch, 1987). PEPC produces the four-carbon acid oxaloacetate, which is rapidly converted into the more stable four-carbon acids malate and/or aspartate (Bräutigam et al., 2014). These acids are shuttled to a cell isolated from the atmosphere in which Rubisco is localized, and CO_2 is released. This biochemical pumping of CO_2 leads to an increase of the relative concentration of CO_2 by a factor of 10 when compared to a non-C_4 cell, and a dramatic increase of photosynthetic efficiency at high temperature (Ehleringer and Bjorkman, 1977; von Caemmerer and Furbank, 2003; Sage, 2004; Sage et al., 2012).

The C_4 photosynthetic mechanism is a classic example of convergent evolution, having evolved more than 60 times independently in various groups of flowering plants (Sage et al., 2011). As all known C_4 enzymes exist in C_3 plants, the evolution of C_4 photosynthesis involved the co-option of genes and proteins essential for the cycle followed by adaption of their expression levels and, at least in some cases, their kinetic properties (Blasing et al., 2002; Tausta et al., 2002; Ghannoum et al., 2005; Aubry et al., 2011; Christin et al., 2013; Heckmann et al., 2013; Kulahoglu et al., 2014; Huang et al., 2017; Moreno-Villena et al., 2018; Alvarez et al., 2019; Niklaus and Kelly, 2019). In particular, the transcript level, enzyme abundance and activity of PEPC are massively increased in all C_4 lineages screened so far (Engelmann et al., 2003; Marshall et al., 2007; Bräutigam et al., 2014; Christin et al., 2015; Moreno-Villena et al., 2018). In contrast, the kinetic behavior of the PEPC enzyme has received less attention and has been investigated mainly in a few systems of eudicot plants that contain closely related C_4 and non-C_4 species, such as the *Flaveria* genus [Asteraceae, (McKown et al., 2005)]. In *Flaveria*, the C_4 PEPC has a ten-fold lower specificity for phosphoenolpyruvate (PEP), an increased sensitivity to activators such as glucose-6-phosphate, and a decreased sensitivity to feedback inhibition from malate and aspartate (Svensson et al., 1997; Engelmann et al., 2003; Svensson et al., 2003; Paulus et al., 2013a; DiMario and Cousins, 2018). Comparison of PEPCs from C_3 to C_4 intermediate species in *Flaveria* further suggested that C_4 properties of the enzyme were gradually acquired during the diversification of the genus (Engelmann et al., 2003). Investigations of PEPC in Amaranthaceae, a distantly related family of eudicots that contains multiple C_4 origins, have shown that PEP specificity evolved convergently in the two groups of C_4 eudicots (Gowik et al., 2006). In contrast, kinetics of PEPC from grasses (Poaceae), the group that contains the largest number of C_4 species, and the most productive and ecologically successful ones (Cerling et al., 1997; Osborne and Beerling, 2006; Sage et al., 2011), remain poorly known. Indeed, previous investigations of PEPC from grass species have used whole leaf preparations, which report on the behavior of mixtures of isoforms and not on well defined, pure enzymes (Huber and Edwards, 1975; Holaday and Black, 1981). PEPC isoforms are encoded by a multi-gene family, with at least six highly divergent gene lineages in most grasses (Christin et al., 2007). The kinetic behaviors have been compared among distant grass paralogs (Dong et al., 1998), but comparisons of closely related C_4 and non-C_4 orthologs in the family are missing.

According to molecular dating, the origins of C_4 photosynthesis are spread throughout the last 35 million years (Christin et al., 2008; Christin et al., 2011). The genus *Flaveria* represents one of the most recent C_4 origins, its different photosynthetic types having diverged in the last 3 million years, with the emergence of fully C_4 plants 1–2 million years ago (Christin et al., 2011). While old C_4 groups exist in eudicots, the previously investigated *Alternanthera* (Gowik et al., 2006) is only slightly older than *Flaveria*, having evolved the C_4 trait 5–10 million years ago (Christin et al., 2011). With more than 22 C_4 origins spanning a recent past up to 35 million years ago, the grass family contains the oldest and largest C_4 lineages (Christin et al., 2008; Christin et al., 2011). In terms of C_4 PEPC evolution, grasses and eudicots co-opted different genes (Christin et al., 2015). Genes encoding C_4-specific PEPC evolved under positive selection in several C_4 groups, but the identity and quantity of fixed amino acid changes varies among families (Besnard et al., 2009; Rosnow et al., 2015). In particular, more of these amino acid changes are observed among grasses than in *Flaveria* (Christin et al., 2007), which might result from the longer divergence between the photosynthetic types. Alternatively, the genes co-opted for C_4 photosynthesis in grasses might have been less fit for the C_4 function, requiring therefore more adaptive changes (Christin et al., 2010). Testing these hypotheses requires generating kinetic data for orthologous non-C_4 and C_4 PEPC genes from grasses. The PEPCs from *Flaveria* are well-studied (Svensson et al., 1997; Svensson et al., 2003; Paulus et al., 2013a; DiMario and Cousins, 2018) and make an excellent starting point for a detailed comparison with other non-characterized PEPCs.

In this work, we characterize the enzymes encoded by orthologous non-C_4 and C_4 genes from two grass species belonging to the same tribe (the C_4 *Panicum queenslandicum* and the C_3 *Panicum pygmaeum* from the tribe Paniceae) and compare them to non-C_4 and C_4 PEPC from *Flaveria* to test the hypotheses that (i) despite very different starting points, qualitatively similar changes happened in C_4 PEPC from *Flaveria* and grasses, and (ii) the kinetic changes differ more between C_4 and non-C_4 PEPC in grasses than in *Flaveria* due to an expanded period of adaptive evolution. We describe the changes in specificity for both substrates (bicarbonate and PEP) as well as the nature of inhibition by aspartate and malate. Overall, out work sheds new light on the impacts of evolutionary time and distance on the convergent evolution of enzyme kinetics.

MATERIALS AND METHODS

Unless otherwise stated, reagents and components were from Sigma, protein purification equipment was from GE Healthcare and both enzymes for cloning and *E. coli* strains were from NEB.

DNA Preparation

Genes that encode the *Flaveria trinervia* PEPC gene and the *Flaveria pringlei* PEPC gene in the pTrc 99A plasmid were provided by Peter Westhoff (Dusseldorf). The PEPC genes were sub cloned into the pET-1B His6 TEV LIC vector plasmid, provided by Scott Gradia (Berkley; Addgene plasmid #29653). Genes were sub cloned using the ligation independent cloning method with Q5 DNA polymerase and T4 DNA polymerase (NEB). Cloned plasmids were isolated using a Miniprep DNA kit (Qiagen). Plasmids were Sanger sequenced to confirm the sequence identity (GATC Biotech).

Leaf samples were collected from *P. queenslandicum* at midday in full daylight and flash frozen in liquid nitrogen. Leaf samples were homogenized with a pestle and mortar in liquid nitrogen. RNA was extracted from ground leaves using the RNeasy Kit (Qiagen). Libraries of cDNA were generated with SuperScript II Reverse Transcriptase (Thermo Fischer Scientific). The PEPC from *P. queenslandicum* was amplified using the primers PquFor1B and PquRev1B (**Supplementary Table 1**), and Q5 polymerase. The amplified gene was Sanger sequenced (GATC Biotech) with the PCR primers and with the primers Pqu_1323_Seq_For and Pqu_1752_Seq_Rev (Primers synthesized by Sigma, summarized in **Supplementary Table 1**). The gene was then cloned into the pET-1B His6 TEV LIC vector plasmid as above.

Because non-C_4 PEPC from C_4 grasses generally represent distant paralogs resulting from ancient duplications that predate the origin of the family (Christin et al., 2007), the most closely related non-C_4 PEPC are in most cases those from related C_3 species. We consequently selected a gene from a C_3 species from the same tribe as *P. queenslandicum*. The sequence for PEPC from *P. pygmeaum* has been previously obtained *via* leaf transcriptome sequencing (Dunning et al., 2017). The sequence was codon optimized for expression in *E. coli* and synthesized by GenArt Gene Synthesis in the pTRCC Plasmid. The synthesized gene was sub-cloned into the pET-1B His6 TEV LIC plasmid and verified by Sanger sequencing.

Protein Expression

For protein expression the BL21λ (DE3) strain of *E. coli* (NEB) was used. Chemically competent *E. coli* cells were transformed with each of the plasmids. Eight liters of cultures were grown in LB media at 37°C to OD_{600} 0.8. Cultures were cooled to 4°C for 1 h prior to recombinant protein induction with 0.5 mM IPTG (Fischer). Cultures were then incubated at 18°C for 18 h. Cells were harvested by centrifugation at 5,400×*g* for 25 min and stored at −80°C.

Protein Purification

Cells were suspended in IMAC buffer (25 mM Tris, 0.5 M NaCl, 0.3 M glycerol, 20 mM imidazole (Acros Scientific)), 10 ml per 2 L of culture with 50 µl of 50 $mgml^{-1}$ DNase I and 100 µl of 100 $mgml^{-1}$ Pefabloc. Cells were passed twice through a cell disruptor (Constant Systems) before centrifugation at 276,000×*g* for 40 min. The supernatant was passed through a 0.45 µm pore filter (Elkay Labs.). PEPC was separated from soluble protein with a prepacked 1 ml nickel affinity column using an ÄKTA Pure 25 L Chromatography System. The loaded column was washed with 50 column volumes of IMAC buffer, then 50 column volumes of IMAC buffer containing 150 mM imidazole. Pure PEPC was eluted with 10 column volumes of IMAC buffer containing 400 mM imidazole.

Protein eluted from IMAC purification was loaded onto a Sephadex G50 desalting column (Amersham Biosciences) and rebuffered in storage buffer (20 mM Tris, 5% v/v glycerol, 150 mM KCl, 1 mM DTT (AnaSpec. Inc.)). Protein was aliquoted and frozen at −80°C until use.

Enzyme Quantification

PEPC enzyme concentration was quantified by absorption at 280 nm. Enzyme extinction coefficients were calculated using the ExPASy protein parameter tool and corrected by determining the absorbance of the protein denatured in 6 M guanidine hydrochloride (Gill and von Hippel, 1989). The extinction coefficients for the *F. trinervia*, *F. pringlei*, *P. queenslandicum* and *P. pygmaeum* PEPC were 120,480 $M^{-1}cm^{-1}$, 117,030 $M^{-1}cm^{-1}$, 105,810 $M^{-1}cm^{-1}$ and 111,510 $M^{-1}cm^{-1}$, respectively. Gel based protein quantification was not used.

Protein samples were analyzed for purity using SDS PAGE analysis. Samples of cell lysate or pure protein (25 µg or 5 µg protein respectively; BCA assay from Pierce) were denatured in 2 × SDS PAGE loading dye (200 mM Tris HCl pH 6.8, 2% SDS, 20% Glycerol, 0.01% Bromophenol blue (BDH Laboratory Supplies) and 7% β-mercaptoethanol). Protein was loaded onto an 8% acrylamide SDS gel with 2 µl of Blue Prestained Protein Standard Broad Range (11–190 kDa) (NEB). Gels were run for 50 min at 200 V with 1 × Tris/Glycine/SDS running buffer (Geneflow). Gels were stained with InstantBlue (Expedeon) and imaged with a ChemiDoc MP (BioRad).

Enzyme Assays

PEPC activity was measured spectroscopically at 340 nm by coupling to NADH-malate dehydrogenase. Assays with a high fixed concentration of bicarbonate were observed using a FLUOstar plate reader (BMG Labtech) through a 340 nm ± 5 nm filter in absorbance mode. These assays were conducted in a reaction volume of 150 µl at 25°C. A typical reaction mixture contained 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$ (Fluka), 5 mM $KHCO_3$, 0.2 mM NADH (Fischer) and 0.1 U μl^{-1} malate dehydrogenase. Assays were initiated with the addition of PEPC enzyme. Rates were calculated with a NADH calibration curve; this method takes account of the short pathlength in microtiter plates.

Assays at a range of bicarbonate concentrations were observed with a Cary spectrophotometer (Agilent Technologies) in the same reaction buffer, in a total reaction volume of 600 µl. Initial rates were calculated using the Cary analysis software. To remove background bicarbonate, the water and tricine buffer were sparged with nitrogen for 18 h prior to use in assays. These assays were constructed under a nitrogen flow and performed in a sealed cuvette. The reaction was initiated with the addition of 50 nM PEPC, delivered with a gastight

syringe (Hamilton). Bicarbonate concentrations were controlled with the addition of freshly prepared potassium bicarbonate.

The background bicarbonate was determined using an endpoint assay with no potassium bicarbonate (30 min). This procedure determines the total concentration of dissolved and hydrated CO_2, (i.e. $CO_{2\ (aq)}$, H_2CO_3, HCO_3^- and CO_3^{2-}), at this pH over 97% is in the form of bicarbonate. Reported bicarbonate concentrations are the sum of the background and the added bicarbonate.

Data Analysis

Kinetic parameters were evaluated by non-linear regression analysis in Igor Pro (Version 7.0.8.1; Wavemetrics Inc., Lake Oswego, Oregon). In all cases, the enzyme was assumed to be fully active. Primary plots were analyzed using Equation (1).

$$\frac{v_i}{[E]_T} = \frac{k_{cat}^{app}[S]}{K_M^{app} + [S]} \quad \text{(Equation 1)}$$

Analysis of secondary plots (*i.e.* of k_{cat}^{app} or k_{cat}^{app}/k_m^{app} vs [PEP]) with Equation (2) allowed determination of the steady-state kinetic parameters, k_{cat}/K_m^{HCO3-} and k_{cat}/K_m^{PEP}, K_i^{PEP} and k_{cat}.

$$k = \frac{k^{app}[S]}{K + [S]} \quad \text{(Equation 2)}$$

Where k and k^{app} are the true and apparent values of k_{cat} or k_{cat}/K_m^{HCO3-} and K is K_m^{PEP} or K_i^{PEP}.

In the case of inhibition data, secondary plots were analyzed using Equation (3), where in the case of competitive inhibition k^{app} is k_{cat}^{app}/k_m^{app} and K_{i} is the competitive inhibition constant K_{ic} or in the case of non-competitive inhibition k^{app} is k_{cat}^{app} and K_{i} is the non-competitive inhibition constant K_{iu}.

The non-competitive inhibition constant (K_{iu}) was determined by the secondary plot of k_{cat}^{app} against inhibitor concentration. The competitive inhibition constant (K_{ic}) was determined by the secondary plot of k_{cat}^{app}/k_m^{app} against inhibitor concentration.

$$k^{app} = \frac{k}{1 + \frac{[I]}{K_i}} \quad \text{(Equation 3)}$$

All data points shown on plots of initial rate against substrate concentration are individual measurements. Standard errors are provided for every parameter estimate. In secondary plots of apparent kinetic parameters against substrate or inhibitor concentration the standard error of those parameter estimates are shown. These standard errors are provided directly by the nonlinear regression analysis routine implemented within Igor Pro.

RESULTS

DNA Cloning and Protein Purification

Four PEPC isoforms were characterized. In grasses, the C_4 and non-C_4 forms of *ppc-1P3* genes were isolated from the C_4 *P. queenslandicum* and synthesized based on the sequence of the C_3 species *P. pygmaeum*, respectively. The cloned genes were 962 and 969 codons long, respectively. They have an 86.2% identity in amino acids and a 93.2% similarity, including on the two positions that have been linked in C_4 *Flaveria* to K_m^{PEP} and decreased inhibition (positions 774 and 884, respectively; Blasing et al., 2000; DiMario and Cousins, 2018). In *Flaveria*, the C_4 and non-C_4 *ppc-1E2* genes corresponding to the C_4 *F. trinervia* and the C_3 species *F. pringlei* were analyzed [ppcA as described in Svensson et al. (1997)]. The two genes are both 967 codons long, with a 94.7% identity and a 97.5% similarity. The orthologous relationships between these pairs of genes were confirmed by phylogenetic analyses (**Supplementary Figure 1**).

All four genes were prepared in vectors for over-expression in *E. coli* with an N-terminal His tag. In all cases, expressed protein was purified to >95% purity as assessed by SDS PAGE with a single immobilized metal column (**Supplementary Figure 2**).

The Presence of an N-Terminal His_6 Tag Does Not Affect Activity

Assays at saturating bicarbonate and variable concentrations of PEP (**Supplementary Figure 3**) showed that both His tagged *Flaveria* PEPCs behaved similarly to untagged proteins previously described (Svensson et al., 1997; Blasing et al., 2000; Jacobs et al., 2008). Specifically, at pH 8.0, 10 mM $MgCl_2$, 10 mM $KHCO_3$, coupled to malate dehydrogenase, the His_6-PEPC from *F. trinervia* catalyses the formation of oxaloacetate with a K_m^{PEP} of 0.61 ± 0.05 mM and a k_{cat} of 47.99 ± 1.22 s^{-1}. Literature values are K_m^{PEP} ranging from 0.278 to 0.652 mM and V_{max} of 29 U mg^{-1}, allowing for the different protein concentration, our k_{cat} would be equivalent to a V_{max} of 25.56 U mg^{-1} (Svensson et al., 1997; Blasing et al., 2000). Under the same conditions, the His_6-PEPC from *F. pringlei* catalyses the formation of oxaloacetate with a K_m^{PEP} of 0.05 ± 0.01 mM and a k_{cat} of 52.65 ± 1.37 s^{-1}; literature values are K_m^{PEP} ranging from 0.029 to 0.061 mM and V_{max} of 27 U mg^{-1}, and allowing for the different protein concentration, our k_{cat} would be equivalent to a V_{max} of 28.02 U mg^{-1} (Svensson et al., 1997; Blasing et al., 2000). This confirms previous reports (Paulus et al., 2013a) that the presence of an N-terminal poly-histidine tag does not adversely affect the activity of these proteins.

Kinetic Analyses Demonstrate That the C_4 Enzyme Forms Show a Lower k_{cat}/K_m Towards PEP and a Higher k_{cat}/K_m to Bicarbonate Than the Related Non-C_4 Forms

The specificity for bicarbonate of all four enzymes was determined using a gas-tight assay system. Background bicarbonate was reduced to *ca.* 50 μM by sparging with nitrogen gas. Assays were performed at five PEP concentrations, while varying the concentration of bicarbonate (**Figure 1**). The analysis of secondary plots (**Supplementary Figures 4** and **5**) provided estimates of k_{cat} and the specificity constant, $k_{\mathrm{cat}}/K_{\mathrm{m}}$, for both substrates (**Table 1**).

The specificity for bicarbonate ($k_{\mathrm{cat}}/K_{\mathrm{m}}$) of the C_4 *P. queenslandicum* PEPC is 1.09×10^6 $M^{-1}s^{-1}$, almost twice as

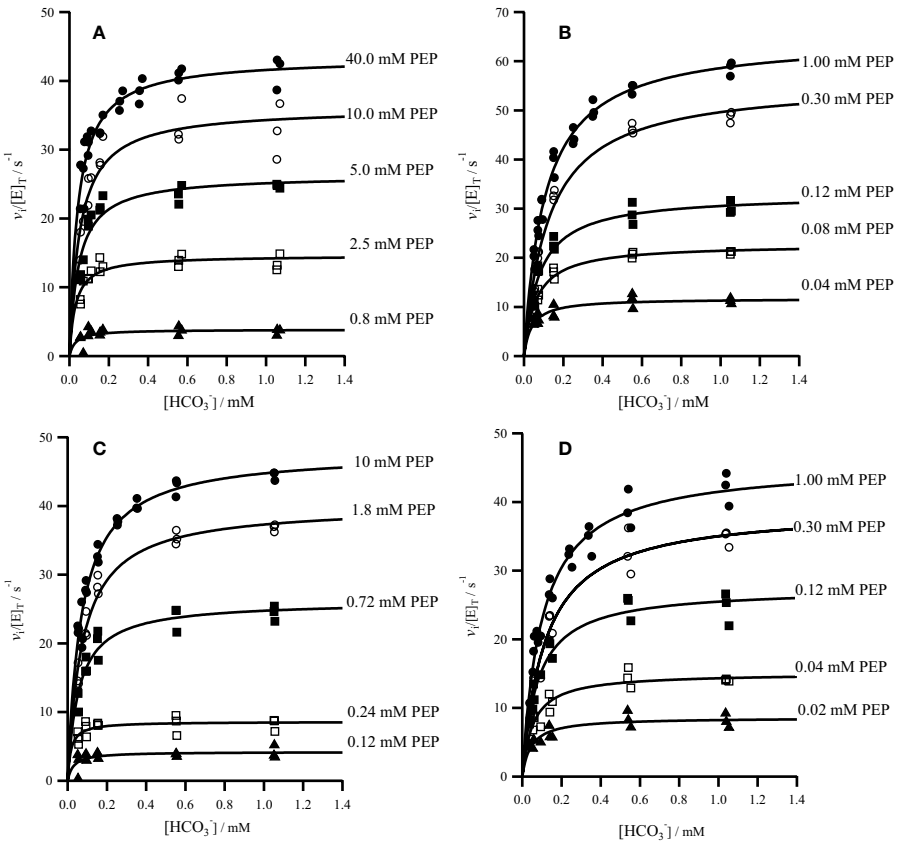

FIGURE 1 | Initial rates of oxaloacetate formation catalyzed by PEPC. Assay conditions were 50 mM Tricine.KOH pH 8.0, 10 mM $MgCl_2$, 0.2 mM NADH, 0.01 $U\mu l^{-1}$ malate dehydrogenase and 50 nM PEPC at 25°C, the concentration of PEP also varied as shown. Assays were repeated (n = 3) at each concentration. Individual data points are shown for the following PEPC **(A)** *Panicum queenslandicum* **(B)** *Panicum pygmaeum* **(C)** *Flaveria trinervia* and **(D)** *Flaveria pringlei* PEPC. Kinetic parameters are summarized in **Table 1**.

TABLE 1 | Summary of kinetic parameters of PEPC found in this study.

PEPC Species	k_{cat} (s^{-1})	K_m^{PEP} (mM)	K_i^{PEP} (mM)	k_{cat}/K_m^{PEP} ($M^{-1}s^{-1}$)	K_m^{HCO3-} (mM)	k_{cat}/K_m^{HCO3-} ($M^{-1}s^{-1}$)
Panicum queenslandicum (C_4)	46.96 ± 1.71	4.17 ± 0.30	4.39 ± 1.10	$0.01\times 10^6 \pm 0.11 \times 10^4$	0.04 ± 0.02	$1.09 \times 10^6 \pm 8.88 \times 10^4$
Panicum pygmaeum (C_3)	65.59 ± 1.74	0.17 ± 0.05	0.05 ± 0.01	$0.50 \times 10^6 \pm 2.44 \times 10^4$	0.12 ± 0.02	$0.60 \times 10^6 \pm 2.93 \times 10^4$
Flaveria trinervia (C_4)	47.99 ± 1.21	0.60 ± 0.05	0.40 ± 0.13	$0.08 \times 10^6 \pm 0.54 \times 10^4$	0.07 ± 0.01	$0.69 \times 10^6 \pm 4.17 \times 10^4$
Flaveria pringlei (C_3)	52.65 ± 1.37	0.06 ± 0.01	0.02 ± 0.01	$0.94 \times 10^6 \pm 8.49 \times 10^4$	0.10 ± 0.01	$0.44 \times 10^6 \pm 2.17 \times 10^4$

Standard errors are given, based on fitted theoretical curves.

large as that of the non-C_4 *P. pygmaeum* enzyme (**Table 1**). The specificity of this non-C_4 enzyme is comparable to that of the C_4 PEPC of *Flaveria* at 0.69×10^6 $M^{-1}s^{-1}$ (**Table 1**), which again is slightly higher than that of the *Flaveria* non-C_4 PEPC (**Table 1**). In both cases the specificity constant for PEP is smaller in the C_4 form of the enzyme (**Table 1**). In terms of bicarbonate K_m values these are within the range previously reported for C_4 and non-C_4 plant PEPC isoforms in work with reasonably careful control of background bicarbonate (O'Leary, 1982; Bauwe, 1986; Janc et al., 1992; Dong et al., 1998; DiMario and Cousins, 2018).

Both C_4 PEPC Enzymes Are Less Sensitive to the Inhibitors Malate and Aspartate at Any Concentration of PEP

For both non-C_4 and C_4 enzymes, we investigated inhibition by the two feedback inhibitors, malate (**Supplementary Figure 6**)

TABLE 2 | Summary of inhibition parameters of PEPC found in this study.

PEPC Species	K_{ic}^{Malate} (mM)	K_{iu}^{Malate} (mM)	$K_{ic}^{Aspartate}$ (mM)
Panicum queenslandicum (C_4)	7.51 ± 1.17	146.08 ± 20.40	49.44 ± 7.86
Panicum pygmaeum (C_3)	0.52 ± 0.22	31.23 ± 0.65	2.27 ± 0.02
Flaveria trinervia (C_4)	10.96 ± 1.55	40.72 ± 4.59	40.02 ± 6.49
Flaveria pringlei (C_3)	2.14 ± 0.62	4.56 ± 1.72	4.13 ± 0.60

Standard errors are given, based on fitted theoretical curves.

and aspartate (**Supplementary Figure 7**) across a range of PEP concentrations. These two structurally related inhibitors show different kinetic characteristics; unlike aspartate, malate remains an inhibitor at saturating concentration of PEP (**Table 2**).

The C_4 cycle of *Flaveria* produces both malate and aspartate (Moore and Edwards, 1986; Meister et al., 1996), while *Panicum* species are expected to produce mainly aspartate around PEPC (Rao and Dixon, 2016). The two molecules have however been shown to inhibit PEPC in a variety of C_4 species (Huber and Edwards, 1975). In our analyses, all the PEPC enzymes are inhibited by malate at both limiting and saturating concentrations of PEP, and malate is a mixed inhibitor (**Figures 2** and **3**). This mixed inhibition can be characterized by two inhibition constants; K_{ic}^{Malate} at limiting PEP and K_{iu}^{Malate} at saturating PEP. In all cases, $K_{ic}^{Malate} >> K_{iu}^{Malate}$, which means that malate can be viewed as a predominantly competitive inhibitor. The two C_4 forms of the enzyme are both less sensitive to malate than the two non-C_4 forms (**Table 2**). Unlike malate, aspartate is solely a competitive inhibitor for all of these enzymes (**Figure 4**). Increasing concentrations of aspartate do not affect k_{cat} (**Supplementary Figure 8**). Once again, the two C_4 forms of the enzyme are much less sensitive to aspartate than the two non-C_4 forms (**Table 2**). Overall, our analyses indicate that the C_4 forms are much less sensitive to both inhibitors, independently of the taxonomic group and C_4 subtype, confirming previous reports (Huber and Edwards, 1975).

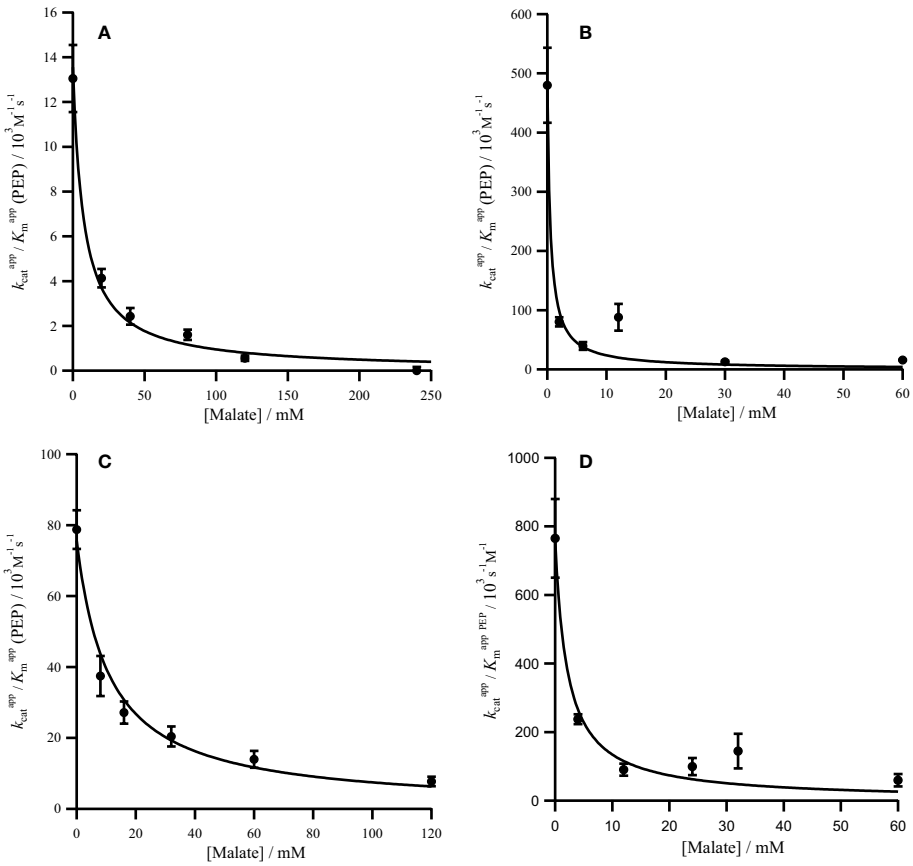

FIGURE 2 | Competitive inhibition of PEPC by malate. Markers represent $k_{cat}^{app}/K_m^{appPEP}$ from assays in the presence of malate (**Supplementary Figure 6**) and error bars represent the standard errors. $k_{cat}^{app}/K_m^{appPEP}$ against malate concentration with inhibition curves characterized by Equation (3) and a K_{ic} for the following PEPC **(A)** *Panicum queenslandicum* (K_{ic}^{Malate} = 7.51 ± 1.17 mM), **(B)** *Panicum pygmaeum* (K_{ic}^{Malate} = 0.52 ± 0.22 mM), **(C)** *Flaveria trinervia* (K_{ic}^{Malate} = 10.96 ± 1.55 mM), and **(D)** *Flaveria pringlei* (K_{ic}^{Malate} = 2.14 ± 0.62 mM). Inhibition parameters are summarized in **Table 2**.

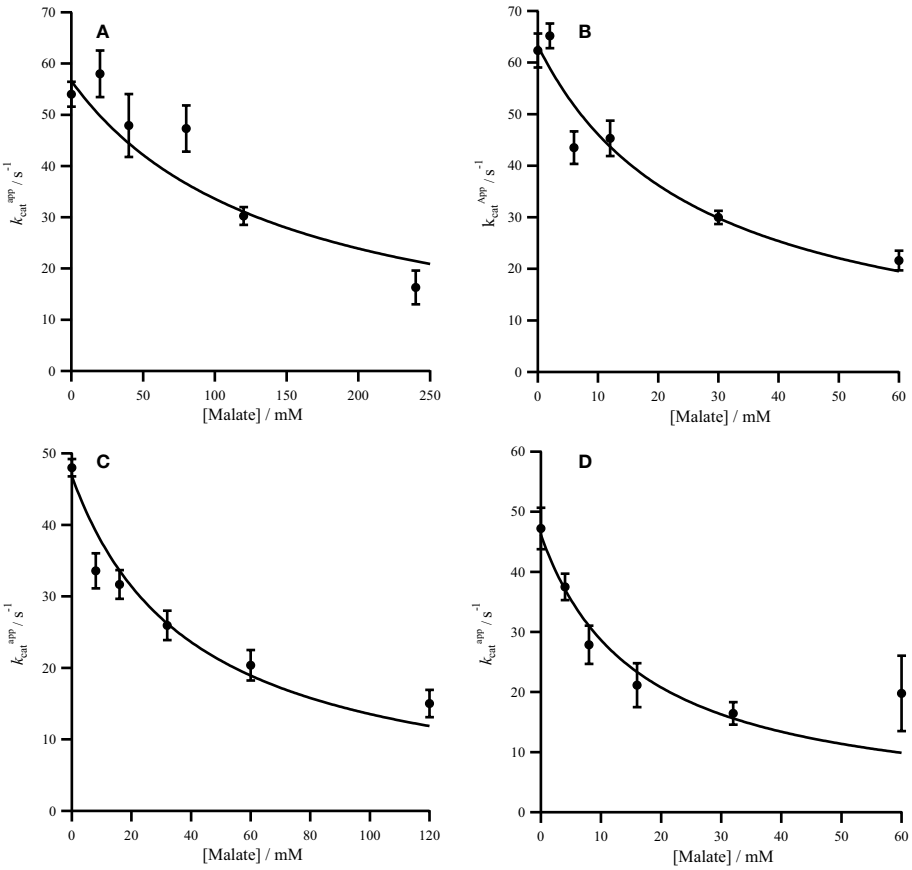

FIGURE 3 | Non-competitive inhibition of PEPC by malate. Markers represent the k_{cat}^{app} from assays in the presence of malate (**Supplementary Figure 6**) and error bars represent the standard errors. k_{cat}^{app} against malate concentration with inhibition curves characterized by Equation (3) and a K_{iu} for the following PEPC **(A)** *Panicum queenslandicum* (K_{iu}^{Malate} = 146.08 ± 20.40 mM), **(B)** *Panicum pygmaeum* (K_{iu}^{Malate} = 31.23 ± 0.65 mM), **(C)** *Flaveria trinervia* (K_{iu}^{Malate} = 40.72 ± 4.59 mM) and **(D)** *Flaveria pringlei* (K_{iu}^{Malate} = 4.56 ± 1.72 mM). Inhibition parameters are summarized in **Table 2**.

DISCUSSION

Convergent Kinetic Changes Across C_4 Flowering Plants

The non-C_4 genes encoding the PEPC enzymes of the C_3 plants *P. pygmaeum* and *F. pringlei* diverged about 150 million years ago and since then have accumulated numerous mutations and undergone multiple gene duplications (Christin et al., 2007; Christin et al., 2015). They share an 83.5% identity and a 91.2% similarity, and greater than 93% similarity with their respective C_4 proteins. While the exact function of each non-C_4 isoform is unknown, they are transcribed at similarly moderate levels (Moreno-Villena et al., 2018). Our investigation shows that the two non-C_4 enzymes characterized here exhibit functionally similar kinetic characteristics, including high sensitivity to competitive inhibition by malate and aspartate and a similar sensitivity to bicarbonate. However, the two non-C_4 isoforms differ in their K_m^{PEP} which is three-fold lower in the *F. pringlei* enzyme (**Table 1**). While systematic screens of non-C_4 PEPC are missing, those of *Alternanthera* and a distant root paralog from *Z. mays* have a similar K_m^{PEP} to the *F. pringlei* enzyme (Dong et al., 1998; Gowik et al., 2006). These data suggest that despite hundreds of million years of divergence, non-C_4 PEPC are generally associated with high sensitivity to inhibitors and low (<0.2 mM) K_m for both substrates. These properties are likely required for a tight regulation and fast response of isoforms involved in anaplerotic functions, where the concentrations of substrates and products are low.

In both *Flaveria* and *Panicum*, the C_4 PEPC shows a markedly reduced sensitivity to both malate and aspartate as compared with the non-C_4 ortholog (**Table 2**). This reduction in sensitivity, reported before in *Flaveria* (Blasing et al., 2002; Paulus et al., 2013b; DiMario and Cousins, 2018) and a variety of grasses from different C_4 subtypes (Huber and Edwards, 1975), is observed at all concentrations of PEP (**Figures 2–4**). Our observations are thus consistent with the conclusion that the same selective pressures act in C_4 eudicots and at least some grasses to decrease the sensitivity to the inhibitors malate and aspartate. In C_4 plants the concentration of malate and aspartate are high, so this reduced sensitivity prevents PEPC being

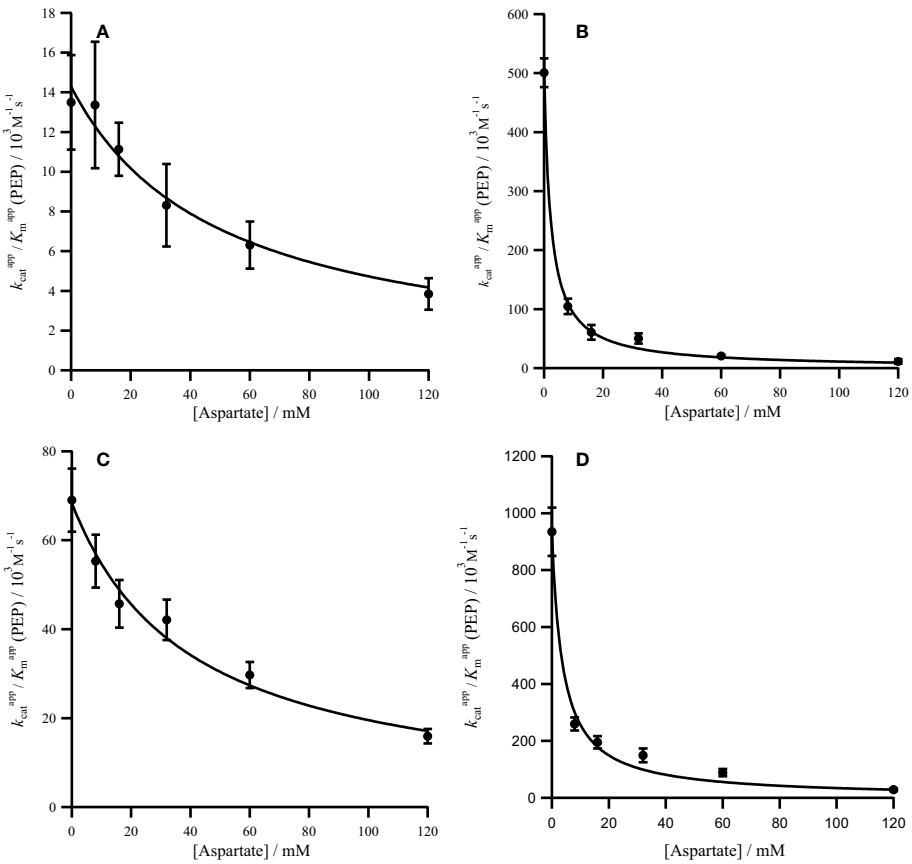

FIGURE 4 | Competitive inhibition of PRPC by aspartate. Markers represent the $k_{cat}^{app}/K_m^{appPEP}$ from assays in the presence of aspartate (**Supplementary Figure 7**) and error bars represent the standard errors. $k_{cat}^{app}/K_m^{appPEP}$ against aspartate concentration with inhibition curves characterized by Equation (3) and a K_{ic} for the following PEPC **(A)** *Panicum queenslandicum* ($K_{ic}^{Aspartate}$ = 49.44 ± 7.86 mM), **(B)** *Panicum pygmaeum* ($K_{ic}^{Aspartate}$ = 2.27 ± 0.02 mM), **(C)** *Flaveria trinervia* ($K_{ic}^{Aspartate}$ = 40.02 ± 6.49 mM) and **(D)** *Flaveria pringlei* ($K_{ic}^{Aspartate}$ = 4.31 ± 0.60 mM). Inhibition parameters are summarized in **Table 2**.

inhibited by downstream metabolites (Arrivault et al., 2017). The respective amounts of malate and aspartate vary among C_4 species (Moore and Edwards, 1986; Meister et al., 1996; Rao and Dixon, 2016), and concerted reduction of inhibition by both species is consistent with them sharing a binding site (Paulus et al., 2013a).

The adaptation of PEPC to the demands of the C_4 pathway involved qualitatively similar changes in substrate specificity between *Flaveria* and the grasses considered here (**Table 1**). In both cases the specificity for PEP decreases and the specificity for bicarbonate increases. The C_4 form of *Zea mays*, an independent C_4 origin within grasses, has an affinity for PEP that is similar to *P. queenslandicum* (Dong et al., 1998). In addition, changes of K_m for PEP in the same direction have been reported in *Alternanthera* (Gowik et al., 2006), suggesting that decreases in PEP K_m happened convergently across C_4 origins. The functional value of these changes remains speculative and might be a side-effect of adaptation of other properties of the enzyme or a direct target of selection for tighter regulation when PEP concentrations are higher (Svensson et al., 2003). The differences in K_m for bicarbonate are less marked than those of PEP (**Table 1**). The K_m for bicarbonate parameter is much higher in the non-C_4 root isoform from *Z. mays* (Dong et al., 1998), indicating it varies tremendously among non-C_4 PEPC. Data from more species are needed to determine whether the qualitative convergence observed here between *Flaveria* and *Panicum* is universal, or depends on the co-opted gene or the details of the C_4 phenotype (e.g. biochemical and anatomical subtypes). Indeed, the cellular concentration of bicarbonate depends on the action of the enzyme carbonic anhydrase, in addition to the cell pH, and it is thus possible that variation in these factors changes the adaptive value of bicarbonate affinity.

The Differences in Enzyme Behavior Are Quantitatively More Important in *Panicum* Than in *Flaveria*

While differences in substrate specificity and sensitivity to inhibitors are qualitatively convergent between *Flaveria* and the two grasses considered here, they are more marked in the latter (**Table 1**). These quantitative differences might be linked to

the contrast between the length of time spent as C_4 in each lineage, from more than 16 million years for *P. queenslandicum* to less than three for *Flaveria* (Christin et al., 2008; Christin et al., 2011). The C_4 PEPCs share a 76.5% identity and an 88.1% similarity. Indeed, the kinetic properties observed in the PEPC of extant taxa result from adaptive changes accumulated since the initial origin of C_4 photosynthesis. According to current models, an initial C_4 pathway can evolve *via* enzyme upregulation and limited modifications of the proteins (Sage et al., 2012; Heckmann et al., 2013; Dunning et al., 2019; Heyduk et al., 2019), as observed in C_3–C_4 intermediates (Svensson et al., 2003; Dunning et al., 2017). Once a C_4 pathway is in place, selection will act to improve its efficiency (Heckmann et al., 2013), and variation among members of the same C_4 lineage indicates that such process can take protracted periods of selection on novel mutations (Heyduk et al., 2019).

Because these are likely necessary for a function of PEPC in C_4 cells with high concentrations of metabolites, we suggest that relaxed sensitivity to inhibitors happens early during the evolution of C_4 PEPC. This hypothesis is supported by the fact that changes in sensitivity to inhibitors are observed in intermediates from *Flaveria* (Engelmann et al., 2003). It has moreover been shown that one single amino acid replacement is sufficient to generate a large decrease in sensitivity (Paulus et al., 2013a). The C_4-specific residue at this site is observed in multiple C_4 lineages of both grasses and eudicots (Paulus et al., 2013a; Paulus et al., 2013b), suggesting that a rapid decrease of inhibition is involved in many origins of C_4 PEPC.

Other properties of C_4-specific PEPC might represent secondary adaptations to the C_4 context, which might happen either to strengthen the early trends or in response to other changes of the plant biochemical phenotype. Over time, sustained diversifying selection on C_4 PEPC would have led to stronger differences between *P. queenslandicum* and the C_3 grasses. This view is supported by the similar kinetic parameters between the C_4 PEPC of *P. queenslandicum* and *Zea mays*, two grass lineages of similar age, as well as similar kinetic parameters observed between the C_4 PEPC in *Alternanthera* and *Flaveria*, two comparatively young lineages (Christin et al., 2011). It is however possible that secondary PEPC adaptations vary among and maybe even within old C_4 lineages, as different biochemical and anatomical C_4 subtypes evolved. Data from more lineages are needed to test the hypothesis that such diversifying PEPC secondary adaptation happened.

The molecular basis of the C_4 specific properties reported here are not well understood. Analysis of the evolution of the amino acid sequence of C_4 PEPC has shown that at least 22 sites underwent positive selection in grasses and sedges (Christin et al., 2007). Of these sites, three are also observed in C_4 *Flaveria* (Christin et al., 2007; Besnard et al., 2009). Some of these mutations have been shown to be responsible for key C_4 specific kinetic properties. Of these, a mutation for alanine to serine at position 774 (*Flavaria* numbering) has been identified as an important determinant of the low specificity for PEP of the C_4 form of the enzyme (Blasing et al., 2000); interestingly, the effect of this position on bicarbonate specificity depends on the rest of the sequence and concentrations of allosteric regulators (DiMario and Cousins, 2018). Additionally, a mutation at position 884 (*Flaveria* numbering), in the allosteric inhibitor binding site, has been shown to have a notable effect on the IC_{50} for malate. An arginine residue in this position, as seen in the non-C_4 form of the enzyme is well placed to directly interact with the inhibitor, increasing the susceptibility to inhibition (Paulus et al., 2013a). These amino acid changes are observed in the C_4 PEPC of *Panicum* but not its non-C_4 ortholog, and presumably contribute to the kinetic differences between the two. The specific role of other grass mutations has yet to be identified, a task that will be complicated by the large amount of variation among grass PEPC and possible epistasy among sites. These factors make it difficult to associate specific kinetic changes with specific amino acid replacements. Here, we compared the characteristics of PEPC from old, diverse lineages; these efforts now need to be expanded to other C_4 lineages, with the well characterized isoforms of *Flaveria* continuing to serve as a model to assess the effect of specific sites.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version. NM carried out all experimental work.

FUNDING

P-AC is supported by a Royal Society University Research Fellowships (URF\R\180022), JR is supported by the Biotechnology and Biological Sciences Research Council (BBSRC, UK; award number BB/M000265/1). NM was supported by the Grantham Centre for Sustainable Futures.

ACKNOWLEDGMENTS

We would like to thank Peter Westhoff (Heinrich Heine University, Dusseldorf) for supplying expression plasmids for the *Flaveria trinervia* PEPC and *Flaveria pringlei* PEPC, Luke Dunning for his help in constructing a maximum likelihood tree, and Scott Gradia (Berkley) for providing the expression vector pET-1B His6 TEV LIC vector plasmid (Addgene plasmid #29653).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01014/full#supplementary-material

REFERENCES

Alvarez, C. E., Bovdilova, A., Höppner, A., Wolff, C.-C., Saigo, M., Trajtenberg, F., et al. (2019). Molecular adaptations of NADP-malic enzyme for its function in C4 photosynthesis in grasses. *Nat. Plants* 5 (7), 755–765. doi: 10.1038/s41477-019-0451-7

Arrivault, S., Obata, T., Szecowka, M., Mengin, V., Guenther, M., Hoehne, M., et al. (2017). Metabolite pools and carbon flow during C_4 photosynthesis in maize: $^{13}CO_2$ labeling kinetics and cell type fractionation. *J. Exp. Bot.* 68 (2), 283–298. doi: 10.1093/jxb/erw414

Atkinson, R. R., Mockford, E. J., Bennett, C., Christin, P. A., Spriggs, E. L., Freckleton, R. P., et al. (2016). C_4 photosynthesis boosts growth by altering physiology, allocation and size. *Nat. Plants* 2 (5), 16038. doi: 10.1038/nplants.2016.38

Aubry, S., Brown, N. J., and Hibberd, J. M. (2011). The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway. *J. Exp. Bot.* 62 (9), 3049–3059. doi: 10.1093/jxb/err012

Bauwe, H. (1986). An efficient method for the determination of K_m values for HCO_3^- of phosphoenolpyruvate carboxylase. *Planta* 169 (3), 356–360. doi: 10.1007/BF00392131

Besnard, G., Muasya, A. M., Russier, F., Roalson, E. H., Salamin, N., and Christin, P. A. (2009). Phylogenomics of C_4 Photosynthesis in Sedges (Cyperaceae): Multiple Appearances and Genetic Convergence. *Mol. Biol. Evol.* 26 (8), 1909–1919. doi: 10.1093/molbev/msp103

Blasing, O., Westhoff, P., and Svensson, P. (2000). Evolution of C_4 phosphoenolpyruvate carboxylase in Flaveria, a conserved serine residue in the carboxyl-terminal part of the enzyme is a major determinant for C_4-specific characteristics. *J. Biol. Chem.* 275 (36), 27917–27923. doi: 10.1074/jbc.M909832199

Blasing, O. E., Ernst, K., Streubel, M., Westhoff, P., and Svensson, P. (2002). The non-photosynthetic phosphoenolpyruvate carboxylases of the C_4 dicot Flaveria trinervia - implications for the evolution of C_4 photosynthesis. *Planta* 215 (3), 448–456. doi: 10.1007/s00425-002-0757-x

Bräutigam, A., Schliesky, S., Külahoglu, C., Osborne, C. P., and Weber, A. P. M. (2014). Towards an integrative model of C_4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C_4 species. *J. Exp. Bot.* 65 (13), 3579–3593. doi: 10.1093/jxb/eru100

Cerling, T. E., Harris, J. M., MacFadden, B. J., Leakey, M. G., Quade, J., Eisenmann, V., et al. (1997). Global vegetation change through the Miocene/Pliocene boundary. *Nature* 389 (6647), 153–158. doi: 10.1038/38229

Christin, P. A., Salamin, N., Savolainen, V., Duvall, M. R., and Besnard, G. (2007). C_4 Photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr. Biol.* 17 (14), 1241–1247. doi: 10.1016/j.cub.2007.06.036

Christin, P. A., Besnard, G., Samaritani, E., Duvall, M. R., Hodkinson, T. R., Savolainen, V., et al. (2008). Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. *Curr. Biol.* 18 (1), 37–43. doi: 10.1016/j.cub.2007.11.058

Christin, P. A., Weinreich, D. M., and Besnard, G. (2010). Causes and evolutionary significance of genetic convergence. *Trends Genet.* 26 (9), 400–405. doi: 10.1016/j.tig.2010.06.005

Christin, P. A., Osborne, C. P., Sage, R. F., Arakaki, M., and Edwards, E. J. (2011). C_4 eudicots are not younger than C_4 monocots. *J. Exp. Bot.* 62 (9), 3171–3181. doi: 10.1093/jxb/err041

Christin, P. A., Boxall, S. F., Gregory, R., Edwards, E. J., Hartwell, J., and Osborne, C. P. (2013). Parallel recruitment of multiple genes into C_4 photosynthesis. *Genome Biol. Evol.* 5 (11), 2174–2187. doi: 10.1093/gbe/evt168

Christin, P. A., Arakaki, M., Osborne, C. P., and Edwards, E. J. (2015). Genetic enablers underlying the clustered evolutionary origins of C_4 photosynthesis in angiosperms. *Mol. Biol. Evol.* 32 (4), 846–858. doi: 10.1093/molbev/msu410

DiMario, R. J., and Cousins, A. B. (2018). A single serine to alanine substitution decreases bicarbonate affinity of phosphoenolpyruvate carboxylase in C_4 Flaveria trinervia. *J. Exp. Bot.* 70 (3), 995–1004. doi: 10.1093/jxb/ery403

Dong, L. Y., Masuda, T., Kawamura, T., Hata, S., and Izui, K. (1998). Cloning, Expression, and Characterization of a Root-Form Phosphoenolpyruvate Carboxylase from Zea mays: Comparison with the C_4-Form Enzyme. *Plant Cell Physiol.* 39 (8), 865–873. doi: 10.1093/oxfordjournals.pcp.a029446

Dunning, L. T., Lundgren, M. R., Moreno-Villena, J. J., Namaganda, M., Edwards, E. J., Nosil, P., et al. (2017). Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives. *Evolution* 71 (6), 1541–1555. doi: 10.1111/evo.13250

Dunning, L. T., Moreno-Villena, J. J., Lundgren, M. R., Dionora, J., Salazar, P., Adams, C., et al. (2019). Key changes in gene expression identified for different stages of C4 evolution in Alloteropsis semialata. *J. Exp. Bot.* 70 (12), 3255–3268. doi: 10.1093/jxb/erz149

Ehleringer, J., and Bjorkman, O. (1977). Quantum Yields for CO_2 Uptake in C_3 and C_4 Plants: Dependence on Temperature, CO_2, and O_2 Concentration. *Plant Physiol.* 59 (1), 86–90. doi: 10.1104/pp.59.1.86

Engelmann, S., Blasing, O., Gowik, U., Svensson, P., and Westhoff, P. (2003). Molecular evolution of C_4 phosphoenolpyruvate carboxylase in the genus Flaveria–a gradual increase from C_3 to C_4 characteristics. *Planta* 217 (5), 717–725. doi: 10.1007/s00425-003-1045-0

Ghannoum, O., Evans, J. R., Chow, W. S., Andrews, T. J., Conroy, J. P., and von Caemmerer, S. (2005). Faster Rubisco is the key to superior nitrogen-use efficiency in NADP-malic enzyme relative to NAD-malic enzyme C_4 grasses. *Plant Physiol.* 137 (2), 638–650. doi: 10.1104/pp.104.054759

Gill, S. C., and von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182 (2), 319–326. doi: 10.1016/0003-2697(89)90602-7

Gowik, U., Engelmann, S., Bläsing, O., Raghavendra, A., and Westhoff, P. (2006). Evolution of C_4 phosphoenolpyruvate carboxylase in the genus Alternanthera: gene families and the enzymatic characteristics of the C_4 isozyme and its orthologues in C_3 and C_3/C_4 Alternantheras. *Planta* 223 (2), 359–368. doi: 10.1007/s00425-005-0085-z

Hatch, M. D. (1987). C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochim. Biophys. Acta* 895 (2), 81–106. doi: 10.1016/s0304-4173(87)80009-5

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P., et al. (2013). Predicting C_4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell* 153 (7), 1579–1588. doi: 10.1016/j.cell.2013.04.058

Heyduk, K., Moreno-Villena, J. J., Gilman, I. S., Christin, P. A., and Edwards, E. J. (2019). The genetics of convergent evolution: insights from plant photosynthesis. *Nat. Rev. Genet.* 20 (8), 485–493. doi: 10.1038/s41576-019-0107-5

Holaday, A. S., and Black, C. C. (1981). Comparative characterization of phosphoenolpyruvate carboxylase in C_3, C_4, and C_3-C_4 intermediate Panicum species. *Plant Physiol.* 67 (2), 330–334. doi: 10.1104/pp.67.2.330

Huang, P., Studer, A. J., Schnable, J. C., Kellogg, E. A., and Brutnell, T. P. (2017). Cross species selection scans identify components of C_4 photosynthesis in the grasses. *J. Exp. Bot.* 68 (2), 127–135. doi: 10.1093/jxb/erw256

Huber, S. C., and Edwards, G. E. (1975). Inhibition of phosphoenolpyruvate carboxylase from C_4 plants by malate and aspartate. *Can. J. Bot.* 53 (17), 1925–1933. doi: 10.1139/b75-216

Jacobs, B., Engelmann, S., Westhoff, P., and Gowik, U. (2008). Evolution of C_4 phosphoenolpyruvate carboxylase in Flaveria: determinants for high tolerance towards the inhibitor L-malate. *Plant Cell Environ.* 31 (6), 793–803. doi: 10.1111/j.1365-3040.2008.01796.x

Janc, J., O'Leary, M., and Cleland, W. (1992). A kinetic investigation of phosphoenolpyruvate carboxylase from Zea mays. *Biochemistry* 31 (28), 6421–6426. doi: 10.1021/bi00143a009

Kulahoglu, C., Denton, A. K., Sommer, M., Mass, J., Schliesky, S., Wrobel, T. J., et al. (2014). Comparative Transcriptome Atlases Reveal Altered Gene Expression Modules between Two Cleomaceae C_3 and C_4 Plant Species. *Plant Cell* 26 (8), 3243–3260. doi: 10.1105/tpc.114.123752

Marshall, D. M., Muhaidat, R., Brown, N. J., Liu, Z., Stanley, S., Griffiths, H., et al. (2007). Cleome, a genus closely related to Arabidopsis, contains species spanning a developmental progression from C_3 to C_4 photosynthesis. *Plant J.* 51 (5), 886–896. doi: 10.1111/j.1365-313x.2007.03188.x

McKown, A. D., Moncalvo, J. M., and Dengler, N. G. (2005). Phylogeny of Flaveria (Asteraceae) and inference of C_4 photosynthesis evolution. *Am. J. Bot.* 92 (11), 1911–1928. doi: 10.3732/ajb.92.11.1911

Meister, M., Agostino, A., and Hatch, M. D. (1996). The roles of malate and aspartate in C4 photosynthetic metabolism of Flaveria bidentis (L.). *Planta* 199 (2), 262–269. doi: 10.1007/BF00196567

Moore, B. D., and Edwards, G. E. (1986). Photosynthetic Induction in a C_4 Dicot, Flaveria trinervia: II. Metabolism of Products of CO_2 Fixation after Different Illumination Times. *Plant Physiol.* 81 (2), 669–673. doi: 10.1104/pp.81.2.669

Moreno-Villena, J. J., Dunning, L. T., Osborne, C. P., and Christin, P. A. (2018). Highly Expressed Genes Are Preferentially Co-Opted for C_4 Photosynthesis. *Mol. Biol. Evol.* 35 (1), 94–106. doi: 10.1093/molbev/msx269

Niklaus, M., and Kelly, S. (2019). The molecular evolution of C_4 photosynthesis: opportunities for understanding and improving the world's most productive plants. *J. Exp. Bot.* 70 (3), 795–804. doi: 10.1093/jxb/ery416

Nisbet, E. G., Grassineau, N. V., Howe, C. J., Abell, P. I., Regelous, M., and Nisbet, R. E. R. (2007). The age of Rubisco: the evolution of oxygenic photosynthesis. *Geobiology* 5 (4), 311–335. doi: 10.1111/j.1472-4669.2007.00127.x

O'Leary, M. H. (1982). Phosphoenolpyruvate Carboxylase: An Enzymologist's View. *Annu. Rev. Plant Physiol.* 33 (1), 297–315. doi: 10.1146/annurev.pp.33.060182.001501

Osborne, C. P., and Beerling, D. J. (2006). Nature's green revolution: the remarkable evolutionary rise of C_4 plants. *Philos. Trans. R. Soc. B.: Biol. Sci.* 361 (1465), 173–194. doi: 10.1098/rstb.2005.1737

Paulus, J. K., Schlieper, D., and Groth, G. (2013). Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution. *Nat. Commun.* 4, 1518. doi: 10.1038/ncomms2504

Paulus, J. K., Niehus, C., and Groth, G. (2013). Evolution of C_4 Phosphoenolpyruvate Carboxylase: Enhanced Feedback Inhibitor Tolerance Is Determined by a Single Residue. *Mol. Plant* 6 (6), 1996–1999. doi: 10.1093/mp/sst078

Rao, X., and Dixon, R. A. (2016). The Differences between NAD-ME and NADP-ME Subtypes of C_4 Photosynthesis: More than Decarboxylating Enzymes. *Front. Plant Sci.* 7, 1525. doi: 10.3389/fpls.2016.01525

Rosnow, J. J., Evans, M. A., Kapralov, M. V., Cousins, A. B., Edwards, G. E., and Roalson, E. H. (2015). Kranz and single-cell forms of C_4 plants in the subfamily Suaedoideae show kinetic C_4 convergence for PEPC and Rubisco with divergent amino acid substitutions. *J. Exp. Bot.* 66 (22), 7347–7358. doi: 10.1093/jxb/erv431

Sage, R. F., Christin, P. A., and Edwards, E. J. (2011). The C_4 plant lineages of planet Earth. *J. Exp. Bot.* 62 (9), 3155–3169. doi: 10.1093/jxb/err048

Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and the Evolution of C_4 Photosynthesis. *Annu. Rev. Plant Biol.* 63 (1), 19–47. doi: 10.1146/annurev-arplant-042811-105511

Sage, R. F. (2004). The evolution of C_4 photosynthesis. *New Phytol.* 161 (2), 341–370. doi: 10.1111/j.1469-8137.2004.00974.x

Skillman, J. B. (2007). Quantum yield variation across the three pathways of photosynthesis: not yet out of the dark. *J. Exp. Bot.* 59 (7), 1647–1661. doi: 10.1093/jxb/ern029

Svensson, P., Blasing, O., and Westhoff, P. (1997). Evolution of the enzymatic characteristics of C_4 phosphoenolpyruvate carboxylase–a comparison of the orthologous PPCA phosphoenolpyruvate carboxylases of Flaveria trinervia (C_4) and Flaveria pringlei (C_3). *Eur. J. Biochem.* 246 (2), 452–460. doi: 10.1111/j.1432-1033.1997.t01-1-00452.x

Svensson, P., Blasing, O., and Westhoff, P. (2003). Evolution of C_4 phosphoenolpyruvate carboxylase. *Arch. Biochem. Biophys.* 414 (2), 180–188. doi: 10.1016/S0003-9861(03)00165-6

Tausta, S. L., Coyle, H. M., Rothermel, B., Stiefel, V., and Nelson, T. (2002). Maize C_4 and non-C_4 NADP-dependent malic enzymes are encoded by distinct genes derived from a plastid-localized ancestor. *Plant Mol. Biol.* 50 (4-5), 635–652. doi: 10.1023/A:1019998905615

Tcherkez, G., Farquhar, G., and Andrews, T. (2006). Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc. Natl. Acad. Sci.* 103 (19), 7246–7251. doi: 10.1073/pnas.0600605103

von Caemmerer, S., and Furbank, R. T. (2003). The C_4 pathway: an efficient CO_2 pump. *Photosynth. Res.* 77 (2/3), 191–207. doi: 10.1023/a:1025830019591

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

REVIEW
published: 18 August 2020
doi: 10.3389/fpls.2020.01267

Ensuring Nutritious Food Under Elevated CO_2 Conditions: A Case for Improved C_4 Crops

*Timothy O. Jobe, Parisa Rahimzadeh Karvansara†, Ivan Zenzen and Stanislav Kopriva**

Institute for Plant Sciences, Cluster of Excellence on Plant Sciences (CEPLAS), University of Cologne, Cologne, Germany

Global climate change is a challenge for efforts to ensure food security for future generations. It will affect crop yields through changes in temperature and precipitation, as well as the nutritional quality of crops. Increased atmospheric CO_2 leads to a penalty in the content of proteins and micronutrients in most staple crops, with the possible exception of C_4 crops. It is essential to understand the control of nutrient homeostasis to mitigate this penalty. However, despite the importance of mineral nutrition for plant performance, comparably less is known about the regulation of nutrient uptake and homeostasis in C_4 plants than in C_3 plants and mineral nutrition has not been a strong focus of the C_4 research. Here we review what is known about C_4 specific features of nitrogen and sulfur assimilation as well as of homeostasis of other essential elements. We identify the major knowledge gaps and urgent questions for future research. We argue that adaptations in mineral nutrition were an integral part of the evolution of C_4 photosynthesis and should be considered in the attempts to engineer C_4 photosynthetic mechanisms into C_3 crops.

Keywords: sulfur, nitrogen, phosphorus, C_4 photosynthesis, maize, *Flaveria*, hidden hunger

OPEN ACCESS

Edited by:
Martha Ludwig, University of Western Australia, Australia

Reviewed by:
Daniel Marino, University of the Basque Country, Spain
Robert Edward Sharwood, Australian National University, Australia

***Correspondence:**
Stanislav Kopriva
skopriva@uni-koeln.de

†Present address:
Parisa Rahimzadeh Karvansara, Centre Algatech, Institute of Microbiology of the Czech Academy of Sciences, Třeboň, Czechia

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *19 May 2020*
Accepted: *03 August 2020*
Published: *18 August 2020*

Citation:
Jobe TO, Rahimzadeh Karvansara P, Zenzen I and Kopriva S (2020) Ensuring Nutritious Food Under Elevated CO_2 Conditions: A Case for Improved C_4 Crops. Front. Plant Sci. 11:1267. doi: 10.3389/fpls.2020.01267

INTRODUCTION

As global population continues to increase, crop yields must increase proportionally to meet the future demand for food. However, the quantity of food is not the only threat to food security, but also the nutritional quality of the food produced (Myers et al., 2017). Indeed, micronutrient deficiencies are estimated to affect over 2 billion people worldwide (Amoroso, 2016). Thus, micronutrient deficiencies impinge on agricultural production, food security, and human health. Global climate change is another factor negatively influencing crop nutritional quality. Many crops grown under the predicted elevated atmospheric CO_2 concentration show an increase in yield, but a decrease in micronutrients (zinc, iron) and proteins (as nitrogen) (Loladze, 2014; Zhu et al., 2018; Ujiie et al., 2019). This decrease is partly due to an increased synthesis of carbohydrates at the expense of proteins, often referred to as the carbon dilution effect. However, it is also caused by the immobilization of nitrogen in vegetative tissues and soil (Luo et al., 2004) and by direct reduction in nitrate assimilation by elevated CO_2 (Bloom et al., 2010). Interestingly, at least for rice, the decreased protein and nitrogen content observed is not completely due to a general carbon dilution, but due to differential responses of the superior grains (derived from early flowers) and the inferior grains (derived from late flowers) to elevated CO_2 (Zhang et al., 2013). Nitrogen content decreases in

superior grains, but it does not change in inferior grains. However, inferior grains are frequently lost during harvest, which further decreases the total grain protein yield (Zhang et al., 2013). Decreased protein content in crops means sulfur will also be less available for human nutrition as plant proteins are the primary source of the essential sulfur-containing amino acid methionine (Parcell, 2002). Indeed, independent FACE (free-air CO_2 enrichment) experiments in wheat showed a 7% decrease in total grain sulfur and an 8% decrease in methionine and cysteine content (Hogy et al., 2009; Fernando et al., 2012). This nutrient penalty has been observed for multiple crops, with one notable exception—C_4 crops (Myers et al., 2014). Presumably, because C_4 crops profit much less from elevated CO_2 as carbon uptake in C_4 plants is saturated at ambient CO_2 levels (Von Caemmerer and Furbank, 2003), no carbon dilution effect occurs, and the elevated CO_2 does not affect protein and micronutrient levels. Thus, C_4 crops have great potential to deliver sufficient nutrients for human food and health. However, more effort is needed to understand the control of nutrient fluxes and homeostasis in C_4 plants to ensure that this will also be true in the coming decades.

Compared to C_3 crops, such as rice, wheat, or oil-seed rape, less is known about specific alterations in mineral nutrition of C_4 plants, despite substantial differences in the organization of nitrate and sulfate assimilation (Jobe et al., 2019). Therefore, in this review, we summarize what is known about C_4 specific features of nitrogen and sulfur metabolism as well as of homeostasis of other essential elements. To identify the major knowledge gaps and urgent questions for future research, we relate the current knowledge of plant mineral nutrition in C_3 vs. C_4 plants with future needs for human nutrition and health and with the predicted changes in atmospheric CO_2 levels. Finally, we discuss future directions and approaches to prevent additional declines in the nutritional quality of crops, mainly engineering C_4 photosynthetic mechanisms into C_3 crops.

C_4 PHOTOSYNTHESIS AND PLANT NUTRITION

Rubisco, the enzyme responsible for assimilating CO_2 into reduced carbon compounds, is an inefficient catalyst under the current atmospheric conditions (Parry et al., 2013; Pottier et al., 2018; Ashida et al., 2019). This inefficiency arises because the carboxylase function of Rubisco can be competitively inhibited by atmospheric oxygen. Thus, many photosynthetic organisms have evolved CO_2 concentrating mechanisms to boost the efficiency of Rubisco by increasing the concentration of CO_2 at the site of carboxylation. Plants using the C_4 photosynthetic pathway accomplish this by dividing the photosynthetic process into two specialized cell types (**Figure 1**). Within mesophyll cells (MC), the initial CO_2 fixation step occurs *via* carboxylation of phosphoenolpyruvate using the enzyme phosphoenolpyruvate carboxylase (PEPC) (Hatch and Slack, 1966; Slack and Hatch, 1967). This is an essential step because PEPC is not inhibited by atmospheric oxygen. The product of this reaction is a four-carbon organic acid that then moves into the bundle sheath (BS) cells, where it is decarboxylated, releasing CO_2 for Rubisco. Because of the low oxygen environment in the BS cells, Rubisco can operate near its maximal efficiency. This pathway has evolved independently in angiosperms at least 66 times, representing three families of monocots and 16 families of dicots (Von Caemmerer and Furbank, 2003; Sage et al., 2012).

While these independent C_4 lineages share many characteristics, there are also significant differences in the C_4-acid decarboxylase enzymes. These differences allow us to classify C_4 plants into three biochemical subgroups. Plants that use NAD malic enzyme (NAD-ME) decarboxylate C_4 acids in the BS mitochondria, while plants using NADP malic enzyme (NADP-ME) decarboxylate C_4 acids in the BS chloroplasts. The third C_4 subtype uses phosphoenolpyruvate carboxykinase (PCK) to decarboxylate C_4 acids primarily in the cytosol of the BS cells. While all of these result in enhanced Rubisco efficiency, these biochemical subtleties reflect differences in the genetic prerequisites for C_4 evolution as well as differences in the selective pressures that favored one subtype over another (Pinto et al., 2016; Sonawane et al., 2017). For example, within C_4 grasses, NADP-ME plants increase in abundance geographically with increasing rainfall, while the number of NAD-ME grasses decreases in these conditions (Taub, 2000; Cabido et al., 2008). Thus, since the discovery of the C_4 photosynthetic pathway, many studies have focused on identifying differences between C_3 and C_4 plants and between different C_4 subtypes to unravel the genetics and evolution of C_4 photosynthesis. Interestingly, nitrogen appears to be an essential component in many of these studies.

Nitrogen

Early research in Poaceae noted that C_4 grasses contained less total nitrogen in their leaves and produced more dry matter per unit of nitrogen fertilizer applied than C_3 grasses. These observations quickly led to the hypothesis that C_4 species utilize nitrogen more efficiently than C_3 species (Brown, 1978). The obvious explanation was the lower investment of nitrogen in Rubisco in C_4 plants (Sage et al., 1987). While this hypothesis is broadly accepted, recent studies suggest minor refinements are justified. For example, Ghannoum et al. (2005) evaluated combinations of various NAD-ME and NADP-ME grass species under high and low nitrogen treatments. They found that while the net CO_2 assimilation rates were similar between these two C_4 subtypes, NAD-ME plants contained more leaf nitrogen than NADP-ME plants with comparable CO_2 assimilation rates. By measuring the total nitrogen in the MC and BS cells, Ghannoum et al. (2005) also showed that in the NAD-ME species, BS cells contained approximately 60% of the nitrogen and chlorophyll. In comparison, only 35% of the total nitrogen and chlorophyll were found in the BS of NADP-ME plants. Analysis of N partitioning suggested that NAD-ME plants invest more nitrogen into the production of Rubisco and other soluble proteins than NADP-ME plants. This seemed to be compensated by significantly greater k_{cat} values of Rubisco in NADP-ME than in NAD-ME species (Ghannoum et al., 2005). Furthermore, a systematic evaluation of several lineages of C_4 grasses encompassing all three biochemical C_4 subtypes found

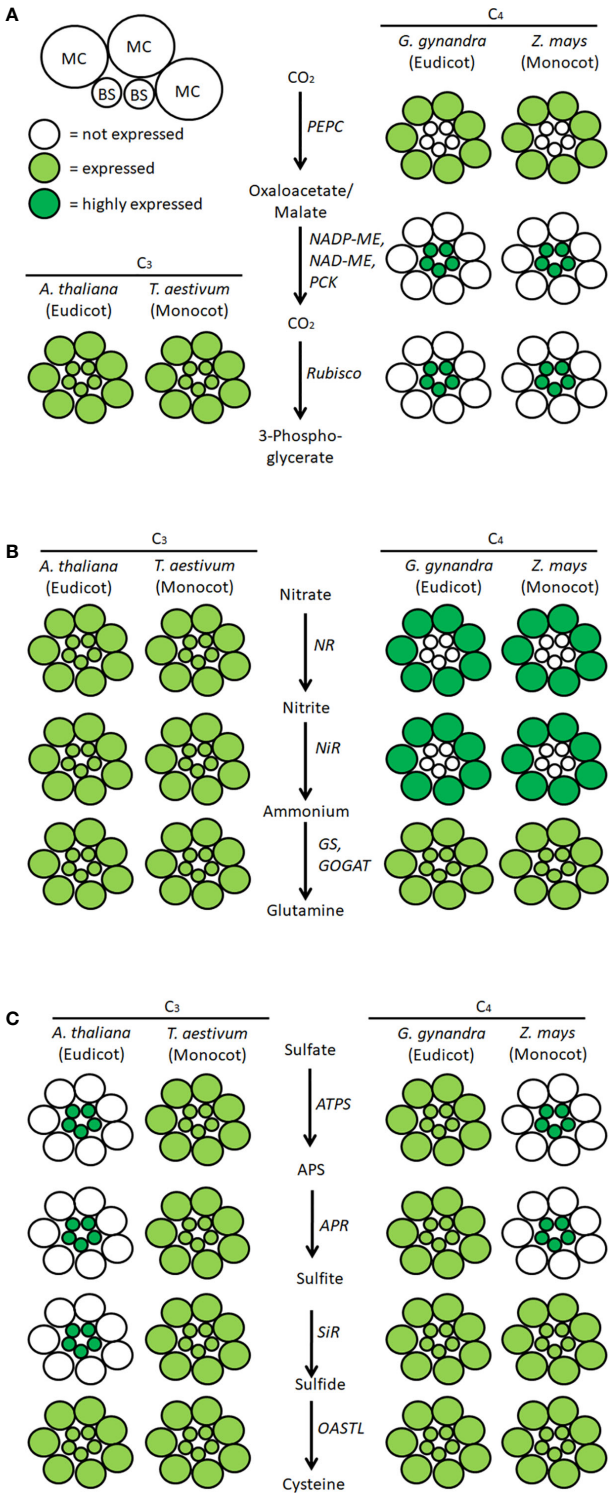

FIGURE 1 | Scheme of localization of the pathways of carbon, nitrate, and sulfate assimilation in exemplary C_3 and C_4 plants. Schematic localization of key enzymes of **(A)** CO_2, **(B)** nitrate, and **(C)** sulfate assimilation in mesophyll cells (MC) and bundle sheath cells (BS) of representative species for C_3 and C_4 monocots and dicots was compiled from the literature described in the manuscript. MC, mesophyll cells; BS, bundle sheath cells; PEPC, phosphoenolpyruvate carboxylase; ME, malic enzyme; PCK, phosphoenolpyruvate carboxykinase; NR, nitrate reductase; NiR, nitrite reductase, GS, glutamine synthetase; GOGAT, glutamate synthase; ATPS, ATP sulfurylase; APS, adenosine 5′-phosphosulfate; APR, APS reductase; SiR, sulfite reductase; OASTL, O-acetylserine (thiol)lyase.

that the nitrogen use efficiency of C_4 grasses is highly correlated with the biochemical subtype with NADP-ME and PCK grasses having higher nitrogen use efficiency than NAD-ME counterparts (Pinto et al., 2016). Thus, while C_4 plants have a higher photosynthetic nitrogen use efficiency (PNUE) than C_3 plants, different biochemical C_4 subtypes vary in their PNUE.

Nitrogen assimilation in C_4 plants differs from C_3 plants not only in PNUE but also in the intercellular compartmentalization of nitrogen assimilation enzymes (**Figure 1**; Kopriva, 2011; Jobe et al., 2019). Already at the onset of C_4 photosynthesis research, it was shown that the activity of nitrate reductase, the key enzyme of nitrate assimilation, is localized mainly in the MC of maize, *Sorghum sudanense*, and *Gomphrena globosa* (Mellor and Tregunna, 1971). Further studies including all three C_4 metabolic subtypes revealed that nitrate reductase was coordinately localized with nitrite reductase, glutamine synthetase, and glutamate synthase, in MC of maize, *Sorghum bicolor*, *Digitaria sanguinalis*, and *Panicum miliaceum*, while in *Panicum maximum*, nitrite reductase was present both in MC and BS (Rathnam and Edwards, 1976). Other studies confirmed the predominant localization of nitrate reductase and nitrite reductase activities in MC, but glutamine synthetase and glutamate synthase were mostly found in both MC and BS (Harel et al., 1977; Moore and Black, 1979). Immunogold labeling confirmed the exclusive localization of maize nitrate reductase in the cytosol of MC (Vaughn and Campbell, 1988) and glutamine synthetase and glutamate synthase in both cell types (Becker et al., 2000). However, whether the spatial distribution of nitrate assimilation in C_4 plants contributes to their PNUE is unknown.

These observations prompted researchers to evaluate the potential role of nitrogen use efficiency and nitrate assimilation as drivers for the evolution of C_4 photosynthesis. Classical schematic models of C_4 evolution suggest that ancestral C_3 plants progressed through a series of discrete stages on the path to C_4 photosynthesis (Edwards et al., 2001; Heckmann et al., 2013; Schluter and Weber, 2016). The first stage is an increase in the BS : MC ratio driven by CO_2 limitation or other environmental factors and a reallocation of glycine decarboxylase (GDC) expression from the MC to the BS. Next is the establishment of C_2 photosynthesis (Mallmann et al., 2014). Next, there is an upregulation of the photorespiratory genes in both the BS and MC, a decrease in Rubisco expression in the MC, and an upregulation of PEPC in the MC (Schluter and Weber, 2016). In the final evolutionary stages, the expression of Rubisco and the photorespiratory genes become confined to the BS. Recent advances in constraint-based modeling have enabled researchers to examine the selective pressures that lead to C_4 photosynthesis *in silico* (Blatke and Brautigam, 2019). These analyses suggested that while light and light distribution were the main drivers governing choice of decarboxylation enzymes, they also predicted that nitrogen limitation might have contributed to C_4 evolution under high levels of photorespiration (Blatke and Brautigam, 2019).

What advantages do these evolutionary adaptations give to C_4 plants over C_3 plants as atmospheric CO_2 increases? Nitrate assimilation was shown to be inhibited by elevated CO_2 in a number of C_3 species but not C_4 plants (Hocking and Meyer, 1991; Bloom et al., 2012). Elevated CO_2 increased PNUE of wheat but not maize, particularly at lower nitrate input, due to enhancing growth, however, at the expense of N accumulation in leaves (Hocking and Meyer, 1991). Nitrate reductase activity was inhibited by the elevated CO_2 in wheat and not in maize (Hocking and Meyer, 1991). Nitrate assimilation can be quantified *in vivo* by an assimilatory quotient, the ratio of net CO_2 consumed over net O_2 evolved (Bloom et al., 2012). Plants assimilating nitrate increase net O_2 evolution while CO_2 consumption is constant, therefore, the assimilatory quotient is low in plants reducing nitrate (Bloom et al., 1989). The quotient is usually determined in comparison with ammonium nutrition after the addition of nitrate as a ΔAQ. In a number of C_3 plants ΔAQ was high at low CO_2 concentrations, but rapidly diminished with increasing CO_2 in accordance with inhibition of nitrate reductase by elevated CO_2 (Bloom et al., 2012). In contrast, in three C_4 species analyzed, the ΔAQ was lower at low CO_2 levels but remained constant with increasing CO_2. Interestingly, in C_3–C_4 intermediate plants the response of ΔAQ to CO_2 was intermediate between C_3 and C_4. Accordingly, FACE experiments have consistently shown that increasing CO_2 negatively impacts nitrogen levels in C_3 plants. This is true for leaves, where often, but not always, Rubisco content diminishes (Bowes, 1991) and for seeds and grains. A recent meta-analysis showed that the average differential effect of increased CO_2 on C_3 plants is—4% (Ebi and Loladze, 2019). In a comparison between several C_3 crops, Myers et al. (2014) found no significant changes in nitrogen content in maize grown under elevated CO_2. While it remains unclear if nitrogen limitation contributed to C_4 evolution, the rising CO_2 levels do not pose a threat for a reduction in nitrogen in C_4 plants as they are already saturated at current CO_2 levels (Von Caemmerer and Furbank, 2003). Although the lower abundance of Rubisco and the identity of the decarboxylation enzyme were shown to impact nitrogen use efficiency, less is known regarding the significance of confining nitrate reduction to the MC. However, it highlights the extensive metabolic rewiring that accompanies C_4 evolution and suggests that multiple mechanisms contribute to enhanced nitrogen use efficiency in C_4 plants. Taken together, both recent and historical studies show that C_4 plants require less total nitrogen, have higher nitrogen use efficiency, and maintain nitrogen levels under elevated CO_2 conditions.

Sulfur

Sulfur is an essential macronutrient for all living organisms, with organic S-compounds representing an important class of metabolites in plant physiology. Sulfate assimilation by plants and microorganisms constitute the entry point of this element into organic molecules in the global sulfur cycle and also in human nutrition. Sulfate is the primary source of S available in nature, and specific H+/sulfate co-transporters from the SULTR family mediate sulfate uptake and mobilization within the plant (reviewed in Takahashi et al., 2011a; Gigolashvili and Kopriva, 2014). Once inside the plant cell, sulfate is initially activated by ATP sulfurylase (ATPS), producing adenosine 5′-phosphosulfate

(APS). In primary S-metabolism, APS undergoes two subsequent reduction reactions catalyzed by APS reductase (APR) to generate sulfite and sulfite reductase (SiR) to produce sulfide. Finally, in a two-step process, serine acetyltransferase catalyzes the transfer of an acetyl moiety from acetyl Coenzyme A to serine resulting in *O*-acetyl-L-serine (OAS). OAS is then used as a substrate for *O*-acetylserine(thiol)lyase (OASTL), which replaces the acetyl group of OAS with sulfide to produce cysteine, the first organic form of sulfur (reviewed in Takahashi et al., 2011b). Cys is the source of reduced S for other metabolites, such as methionine or the tripeptide glutathione (GSH), an essential part of plant redox homeostasis and stress defense (Noctor et al., 2012).

Like nitrate assimilation, sulfate assimilation is differentially localized in MC and BS of C_4 plants. In a number of C_4 species spanning all three C_4 subtypes, most of the total leaf ATPS activity is confined to BS chloroplasts (Gerwick et al., 1980; Passera and Ghisi, 1982). Similar to nitrate assimilation, not all enzymes of the pathway are coordinately expressed. While APR was also found almost exclusively in BS of maize (Schmutz and Brunold, 1984; Burgener et al., 1998), the activities of SiR and OASTL were detected at comparable levels in MC and BS (Passera and Ghisi, 1982; Schmutz and Brunold, 1985). Reduced sulfur needed in MC is transported from maize BS in the form of cysteine (Burgener et al., 1998). Interestingly, GSH synthesis and homeostasis are also differently organized in MC and BS. In maize, GSH synthetase activity is higher in MC than in BS, in line with the export of Cys from BS (Burgener et al., 1998). This results in a higher accumulation of GSH in MC, possibly connected to higher H_2O_2 levels in MC than in BS (Doulis et al., 1997). Given the importance of GSH for maintaining cellular redox potential, it is surprising that glutathione reductase, the key element of the glutathione redox cycle, was also found exclusively in MC of maize (Doulis et al., 1997; Pastori et al., 2000). However, not all C_4 plants follow the same pattern. In the C_4 species of the dicot genus *Flaveria*, APR and ATPS are expressed in both MC and BS (Koprivova et al., 2001). Since the C_4 species analyzed previously were all monocots, BS-exclusive localization of sulfate assimilation could be a trait of C_4 monocots but not C_4 eudicots (**Figure 1**; Koprivova et al., 2001; Kopriva and Koprivova, 2005). Indeed, numerous RNA-seq analyses of MC and BS transcripts showed BS localization of transcripts for ATPS and APR in different C_4 monocots (maize, sorghum, *Setaria viridis*) but a similar transcript abundance in MC and BS of the eudicot C_4 species *Gynandropsis gynandra* (Aubry et al., 2014a; John et al., 2014; Doring et al., 2016; Denton et al., 2017). Thus, the localization of sulfate assimilation in BS cannot be a general C_4 trait. This conclusion was unexpectedly confirmed by experiments with the C_3 model plant, *Arabidopsis thaliana*. In an analysis aimed at discerning the function of the BS cell layer in C_3 plants using a translatome approach, Aubry et al. (2014b) found an enrichment of transcripts for sulfate assimilation genes in the BS. Transcripts of ATPS, APR, SiR, as well as other components of Cys synthesis, in addition to sulfate transporters and genes for synthesizing the sulfur-rich secondary compounds glucosinolates were all overrepresented in RNA from BS compared to the whole leaf (Aubry et al., 2014b). Three obvious questions arise from this study. First, what is the ancestral localization of the sulfate assimilation pathway? Secondly, what is the metabolic significance of the various relocations? Finally, in the C_4 lineages with relocated sulfate assimilation enzymes, was the relocation of sulfate assimilation a prerequisite for C_4 evolution or a consequence of C_4 evolution? These remain key open questions in plant sulfur research.

An analysis of sulfate assimilation in the eudicot genus *Flaveria* revealed another intriguing result. A gradient in the accumulation of leaf Cys and GSH was observed with higher concentrations in the leaves of C_4 species than in C_3 and C_3-C_4 intermediate species (Koprivova et al., 2001; Gerlich et al., 2018). This gradient is sustained through a similar gradient in sulfate uptake, reduction rate, transcript levels, and activity of APR (Koprivova et al., 2001; Weckopp and Kopriva, 2014; Gerlich et al., 2018). Interestingly, expression analyses suggested that sulfate reduction and GSH synthesis are preferentially localized in the roots of C_4 *Flaveria* species. Interspecies grafts of C_3 *F. robusta* and C_4 *F. bidentis* were created to test this hypothesis. The results of this experiment showed that the high GSH accumulation in C_4 leaves is indeed controlled by the roots (Gerlich et al., 2018). While it is plausible that the importance of roots for Cys and GSH synthesis in C_4 *Flaveria* is connected to serine synthesis, which is preferentially synthesized in the roots of C_4 plants through the phosphorylated pathway (Gerlich et al., 2018), this hypothesis should be tested in more C_4 species.

Sulfur is much less abundant in the plant body than nitrogen making it unlikely to be the driving force behind the metabolic adaptations leading to the evolution of C_4 photosynthesis. However, it is possible that the gradient of higher sulfate assimilation flux with increasing C_4 photosynthesis in *Flaveria* is a result of the adaptation to dry and warm habitats typical for C_4 plants. Thus, the higher GSH contents in C_4 *Flaveria* might be a mechanism to cope with increased oxidative stress caused by such environmental conditions. This is consistent with the critical role of GSH in chilling tolerance in maize (Kocsy et al., 2001). However, the importance of the BS-localization of sulfate assimilation in C_4 monocots and possibly in the roots of C_4 dicots is still elusive.

Phosphorus

In addition to carbon, nitrogen, and sulfur, phosphorus is a macronutrient crucial for plant growth and development. As an essential player in cellular energy conversion, an enzymatic substrate, as well as a regulatory factor of enzyme activity, phosphate plays many crucial roles in cellular biochemistry. Moreover, phosphate is responsible for the acidic nature of nucleic acids and is a vital constituent of phospholipid membranes. Plants employ several morphological and physiological adaptations to mitigate phosphorus deficiency, including interconnections with the rhizosphere and soil microbes and diverse molecular mechanisms (Lopez-Arredondo et al., 2014). Phosphate is taken up by various phosphate transporters as an inorganic anion. However, unlike nitrate and sulfate, phosphate is

not reduced and remains in its oxidized state as either a free anion or is incorporated into organic compounds *via* phosphate esters. Disruptions in phosphate homeostasis have intensive footprints on plants. Thus, shoot phosphate concentrations are tightly regulated by systemic control of phosphate uptake and allocation (Bari et al., 2006; Ham et al., 2018; Kopriva and Chu, 2018). Control of phosphate homeostasis is coordinated with the regulation of other nutrients, particularly nitrate and sulfate (Rouached et al., 2011; Hu et al., 2019; Medici et al., 2019).

Phosphate has a vital role in photosynthesis. The metabolic energy of the cell and the energy generated during the light reactions of photosynthesis are stored in phosphate esters and energy-rich pyrophosphate bonds. Inorganic phosphate in the chloroplast regulates the partitioning of photosynthates between starch synthesis and export to the cytosol (Heldt et al., 1977). Moreover, phosphate is indispensable for the function of the triose-phosphate/phosphate translocator (TPT), an antiporter in the inner membrane of the chloroplast (Lee et al., 2017). The TPT exchanges phosphate from the cytosol with triose-phosphates synthesized in the Calvin cycle (Fliege et al., 1978). In C_4 plants, the TPT is even more highly abundant in envelopes of MC chloroplasts as the flux through this transporter is higher in C_4 plants than in C_3 plants (Brautigam et al., 2008). In addition, C_4 plants possess another abundant phosphate driven transporter, the phosphoenolpyruvate phosphate translocator (PPT), which is essential for the transport of PEP from the chloroplast in MC (Brautigam et al., 2008; Majeran et al., 2008). Also, the activities of the critical enzymes involved in C_4 carbon assimilation, such as PEPC, PCK, and pyruvate phosphate dikinase, are modulated by reversible phosphorylation (Ashton and Hatch, 1983; Jiao and Chollet, 1991; Chao et al., 2014).

Although phosphate demand to facilitate transport processes in C_4 plants is high, C_4 specific features of phosphate homeostasis or possible differences in (photosynthetic) phosphate use efficiency (PUE) have not been described. Phosphate deficiency was shown to decrease Rubisco activity in sunflower, but Rubisco activity was not affected by phosphate deficiency in maize (Jacob and Lawlor, 1992). Similarly, C_4 grasses produced higher forage yields on phosphate-limited soil than C_3 grasses (Morris et al., 1982). Accordingly, in a comparative survey of photosynthetic and growth responses to phosphate deficiency in 12 species with diverse photosynthetic characteristics, C_3 species showed more substantial growth retardation in comparison to C_4 species (Halsted and Lynch, 1996). However, no photosynthesis type-dependent changes in photosynthetic PUE could be determined. Although the CO_2 exchange rate was decreased less by phosphate deficiency in C_4 plants than in C_3 ones, due to higher foliar phosphate concentration, the photosynthetic PUE remained unchanged (Halsted and Lynch, 1996). Interestingly, the monocot species were less sensitive to low phosphate stress than dicots irrespective of photosynthesis type, due to a lower phosphate content in the leaf and better maintenance of growth (Halsted and Lynch, 1996). In an independent study focusing on monocots, the response of CO_2 assimilation rates to leaf phosphate concentration was saturated in C_4 species but not in their C_3 relatives (Ghannoum et al., 2008). It seems, therefore, that although C_4 plants require higher amounts of phosphate than C_3 plants, their CO_2 assimilation is less sensitive to phosphate limitation.

HOW DOES ELEVATED CO_2 AFFECT MICRONUTRIENTS IN C_3 AND C_4 PLANTS?

The World Health Organization (WHO) defines malnutrition as deficiencies, excesses, or imbalances in a person´s energy intake and/or nutrient intake (https://www.who.int/news-room/fact-sheets/detail/malnutrition) and recognizes three broad groups of malnutrition conditions - undernutrition, micronutrient-related malnutrition, and overnutrition and noncommunicable diseases. Over the past 60–70 years, plant biologists and plant breeders have focused their attention on alleviating undernutrition by dramatically increasing crop yields by improving plant genetics and intensifying agricultural production systems. However, by focusing on yield, changes in the nutritional value of our food have been largely neglected, especially regarding micronutrient content. Thus, micronutrient levels in plants have decreased for two main reasons. First, intensive agricultural practices have depleted micronutrients from the soil, and second, rising atmospheric carbon dioxide negatively affects the nutrient profiles of C_3 crop plants (Loladze, 2014).

Micronutrient-related malnutrition, sometimes called hidden hunger, is caused by poorly diversified diets that meet the caloric but not the nutritional needs of an individual and is primarily associated with micronutrient deficiency (Myers et al., 2017). In addition to the well documented adverse effects of increasing atmospheric CO_2 on macronutrients in C_3 crops (see above), there is evidence that the effects are equally adverse, or in some cases, much worse for micronutrient levels. For example, a study on the impact of elevated CO_2 on nine diverse rice cultivars showed that growth at elevated CO_2 decreased the manganese (Mn) content in the body of rice plants by 53% (Ujiie et al., 2019). In this same study, the Mn content in the brown rice decreased by 7%, while the polished rice showed a 20.5% decrease in Mn when grown under elevated CO_2 (Ujiie et al., 2019). The vast differences observed in Mn content in different tissues is a significant finding as rice is becoming an important forage crop in some regions of the world (Cheng et al., 2018). While Mn deficiency in forage animals is considered rare, such a significant decrease in micronutrients in the body of the plant suggests that forage animal nutrition will also suffer as a result of rising CO_2. Thus, to accurately assess all the potential impacts of CO_2-induced nutrient depletion on human health, it is crucial to measure nutrients in multiple plant tissues.

Interestingly, a more extensive meta-analysis of 130 plant species/cultivars was unable to detect a significant decrease in Mn content among C_3 crops (Loladze, 2014). However, this study did identify significant decreases in many other micronutrients, namely iron (Fe) and zinc (Zn). Iron is of

particular interest as at least 2 billion people currently suffer from Fe deficiency, making anemia a leading cause of maternal mortality (Micronutrient_Initiative, 2009). Zinc deficiency is also widespread, with approximately 30% of the world population at risk. Zinc deficiency can cause compromised immune responses, stunting during childhood, and increased risk of child mortality (Micronutrient_Initiative, 2009; Livingstone, 2015). While crosstalk between Fe, Zn, P, and S signaling in plants is recognized, not much is known in C_3 or C_4 plants regarding the mechanistic integration of these signaling networks (Mendoza-Cozatl et al., 2019; Xie et al., 2019). However, it was recently proposed that Fe, Zn, P, and S signaling are integrated in a PHR1 dependent manner in the C_3 plant Arabidopsis (Briat et al., 2015). Interestingly, the rice homolog of PHR1, OsPHR2, was also shown to play a role in the integration of P and N signaling networks in rice (Hu et al., 2019). Thus, in C_3 plants, it seems PHR proteins may be essential network hubs integrating signaling from multiple nutrients. When viewed from this perspective, the changes in micronutrient levels observed in C_3 plants under elevated CO_2 could be pleiotropic effects caused by disruption of N and/or P signaling. It remains unknown if these signaling networks are conserved between C_3 and C_4 plants.

Additionally, the genetic diversity in the C_3 crops has a large impact on the effects of elevated CO_2. The variation within species may even exceed the variation between species. For example, a study with 17 rice cultivars grown under controlled conditions in normal or 664 ppm CO_2 showed 10–265% increase in total biomass and even greater—10–350% variation in response of grain yield (Ziska et al., 1996). This is true also for qualitative traits; protein content in grains of 18 field-grown rice cultivars cultivated at ca., 585 ppm CO_2 decreased by 5–20%, whereas grain Zn and Fe concentrations decreased on average, but actually increased in four and two genotypes, respectively, and were not affected in another variety, Nipponbare (Zhu et al., 2018). Similar variation was observed in other species and, interestingly, modern varieties of oat, wheat, or soybean seem to be less responsive to elevated CO_2 than varieties from the 1920s (Ziska and Blumenthal, 2007). There might, therefore, be a potential for the selection of new crop varieties for response to elevated CO_2 levels (Shimono et al., 2018).

The question thus arises, can C_4 crops help to alleviate "hidden hunger"? There are currently only five economically important C_4 food crops—maize, sorghum, sugar cane, onion, and pearl millet. While the list of C_4 crops is small, they account for a large proportion of global crop production. For example, the average annual production of maize from 2008–2010 was 750 million metric tons representing 27% of cereal area, 34% of cereal production and 8% of the value of all primary crop production (Shiferaw et al., 2011). The nutritional quality of these C_4 crops is at best average, e.g., due to low lysine content in maize proteins or poor digestibility of sorghum and millet proteins (Millward, 1999; Galili and Amir, 2013). However, there are also several regionally important C_4 crops, often called orphan crops, that have more desirable nutritional traits for combating hidden hunger. Notable orphan crops include grain amaranth, teff (*Eragrostis tef*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*), and proso millet (*Panicum miliaceum*).

These data suggest that while C_4 crops do not show a CO_2-induced nutritional penalty, the current staple C_4 crops may not be best suited to address dietary deficits and hidden hunger. However, significant genetic advances have been made to improve the nutritional quality of maize and sorghum. For example, a recent genome-wide association study on 923 maize lines identified 46 QTLs significantly associated with seed Zn and Fe concentrations (Hindu et al., 2018). Introgressing favorable alleles of these QTLs into commercial varieties could improve both Zn and Fe levels in maize kernels. Additionally, researchers have developed quality protein maize (QPM), having almost twice the amount of lysine and tryptophan as traditional varieties, and maize lines with enhanced levels of provitamin-A or methionine (Wurtzel et al., 2012; Galili and Amir, 2013; Planta et al., 2017). Thus, biofortification is a viable approach to enhance the nutritional value of C_4 crops and address hidden hunger.

In addition to food crops, there are eight C_4 crops grown for turf, forage, or bioenergy. These include *Miscanthus* x *giganteus*, *Panicum virgatum* (switchgrass), *Chloris gayana* (Rhodes grass), *Cynodon dactylon* (Bermuda grass), *Melinis minutifolia* (molasses grass), *Panicum maximum*, *Cenchrus purpureus* (Napier grass), and *Zoysia japonica*. Collectively, these crops are all known for their high productivity and demonstrate the potential of C_4 plants. Similar to food crops, a nutritional comparison of C_3 and C_4 forage grasses grown under high and low CO_2 levels found that the C_3 grasses had higher levels of protein, nonstructural carbohydrates, and water, but lower levels of fiber when grown under elevated CO_2 compared to the C_4 species (Barbehenn et al., 2004).

Under current environmental conditions, the staple C_4 crops show superior productivity compared to C_3 crops, and some of the C_4 orphan crops seem to have the same or even better nutritional quality (**Table 1**). While the productivity gap can be expected to narrow down, due to elevated atmospheric CO_2 that fertilizes C_3 crops but not C_4 crops, the relative nutritional value of the current C_4 crops may improve because of the lack of the carbon nutrient penalty. Also the rise in temperatures may favor C_4 crops in the future, or at least extend their cultivation areas. However, hidden hunger cannot be combatted without investment into further crop improvement, specifically targeting nutritional quality of staple C_4 crops and improving the productivity of selected local crops with high nutritional value, such as pearl millet.

FUTURE DIRECTIONS

Open Questions on C_4 Mineral Nutrition

To improve the nutritional value of C_4 crops for human food, it is necessary to understand more about the control of their nutrient homeostasis. While some progress has been made, e.g., in the biofortification of maize (Wurtzel et al., 2012; Galili and Amir,

TABLE 1 | Comparison of nutritional composition of grains of several cereal and orphan crops.

per 100 g DW	Crop	Energy (kcal)	Carbohydrate (g)	Protein (g)	Fat (g)	Ash (g)	Fiber (g)	Ca (mg)	Fe (mg)	Thiamin (mg)	Riboflavin (mg)	Niacin (mg)
C_3	Rice (brown)	362	76	7.9	2.7	1.3	1	33	1.8	0.41	0.04	4.3
	Wheat	348	71	11.6	2	1.6	2	30	3.5	0.41	0.1	5.1
C_4	Maize	358	73	9.2	4.6	1.2	2.8	26	2.7	0.38	0.2	3.6
	Sorghum	329	70.7	10.4	3.1	1.6	2	25	5.4	0.38	0.15	4.3
	Pearl millet	363	67	11.8	4.8	2.2	2.3	42	11	0.38	0.21	2.8
	Finger millet	336	72.6	7.7	1.5	2.6	3.6	350	3.9	0.42	0.19	1.1
	Foxtail millet	351	63.2	11.2	4	3.3	6.7	31	2.8	0.59	0.11	3.2
	Common millet	364	63.8	12.5	3.5	3.1	5.2	8	2.9	0.41	0.28	4.5
	Little millet	329	60.9	9.7	5.2	5.4	7.6	17	9.3	0.3	0.09	3.2
	Barnyard millet	300	55	11	3.9	4.5	13.6	22	18.6	0.33	0.1	4.2
	Kodo millet	353	66.6	9.8	3.6	3.3	5.2	35	1.7	0.15	0.09	2
	Teff	357	73	8–11	2.5	2.8	3	17 – 178	9.5 – 37.7	0.19	0.17	1.5
	Quinoa	399	67.6	12.9	5.8	2.2	13.6	148.7	13.2	0.13	0.02	0.6
	Grain Amaranth	371	65.3	13.6	7	2.9	6.7	159	7.6	0.116	0.2	0.92

Data are shown per 100 g dry weight and are taken from Caselato-Sousa and Amaya-Farfán (2012); Saleh et al. (2013), and Niro et al. (2019).

2013; Planta et al., 2017), many questions on mineral nutrition of C_4 plants are still open (**Figure 2**). Probably the biggest set of fundamental questions concerns the drivers and the consequences of the spatial separation of nitrate and sulfate assimilation in C_4 monocots. Does the MC localization of nitrate reductase contribute to the improved nitrogen use efficiency of C_4 plants? Have C_3–C_4 intermediate plants improved nitrogen use efficiency compared to C_3 plants? Are there any differences in sulfur use efficiency between C_3 and C_4 plants? Is the gradient in the accumulation of sulfur compounds found in *Flaveria* conserved in other genera with C_3 and C_4 photosynthesis? Are the pathways of nitrate and sulfate assimilation differently regulated in C_3 and C_4 plants? Why is sulfate assimilation differently localized in C_4 monocots and C_4 dicots?

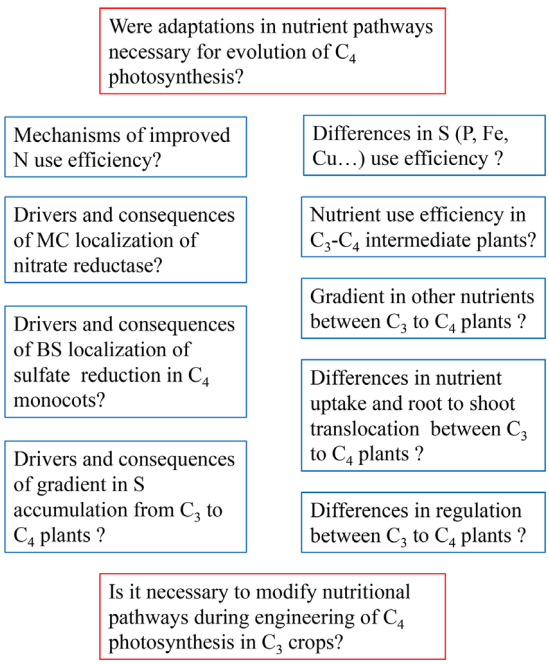

FIGURE 2 | Summary of major research questions in mineral nutrition of C_4 plants.

The other set of questions concerns other nutrients. Is there a gradient similar to that of sulfur compounds in *Flaveria* in accumulation of other nutrients between closely related C_3 and C_4 plants? Does the high flux through TPT and PPT in C_4 plants affect their phosphate needs and homeostasis? Is phosphate homeostasis affected by elevated CO_2? Is there a different need for Fe or Cu in C_4 and C_3 plants given the different arrangements of photosynthetic apparatus?

All these fundamental unknowns lead to one overarching question: Were adaptations in nutrient pathways necessary for the evolution of C_4 photosynthesis? This question has major practical implications for the efforts to improve C_4 crops, but particularly for engineering C_4 photosynthesis to C_3 crops.

Improvement of C_4 Crops by Traditional Breeding

Breeding material with high nutritional value is available for maize (Newell et al., 2014; Wang et al., 2019) and thus breeding for improved nutritional value is feasible. A few approaches for maintaining the nutritional levels of crops under elevated CO_2 have been proposed. For example, the negative effect of elevated CO_2 on nitrate assimilation and nitrogen content might be attenuated by increasing the proportion of ammonium as the nitrogen source (Bloom et al., 2010). However, crop species differ in their tolerance to ammonium, therefore, as discussed above, the most straightforward approach is to incorporate FACE studies into modern breeding programs. This approach would be useful for both C_3 and C_4 crops and would allow us to accomplish two goals. First, we could screen specifically for traits that improve the nutritional levels of crops under elevated CO_2 and select for these traits in future cultivars. Secondly, we could

ensure that traits selected to meet other breeding goals (i.e., pathogen resistance traits or drought resistance traits) are not negatively affected by elevated CO_2 levels and do not further decrease the nutritional standards of our crops. While this approach would be technically challenging for breeding programs due to the expense and large space requirements associated with field-scale FACE studies, it has a high likelihood of success in the short term. As noted by Ujiie et al. (2019), carbohydrates, nitrogen, and sulfur resources are all transported through the phloem during nutrient reallocation and grain filling. Thus, improving nutrient translocation or the strength of the sink organ could counteract the nutritional decrease in crops grown under elevated CO_2. These goals are well within the scope of modern breeding programs.

C_4 Engineering

The conversion of C_3 crops to full C_4 photosynthesis is a long-standing goal of plant biologists, and significant advances have been made with the help of both systems biology and synthetic biology (Schuler et al., 2016; Ermakova et al., 2020). To achieve this, at least five major milestones have been identified that are necessary to convert C_3 crops to C_4 photosynthesis: 1) induction of higher-order veins, 2) increase BS:M ratio, 3) adaptation of BS morphology, 4) engineering of dimorphic chloroplasts in BS and M cells, and 5) compartmentalization of the photosynthetic enzymes between BS and M cells (reviewed in Schuler et al., 2016). However, significant hurdles remain, especially in identifying a suitable C_3 chassis for engineering, establishing Kranz anatomy, and the establishment of a carbon concentrating mechanism (Hennacy and Jonikas, 2020). Despite these challenges, consortiums like the C_4 Rice Project, a global collaboration between leading researchers in photosynthesis, aim to engineer C_4 photosynthesis into rice. Increasing rice yield and decreasing water and nitrogen fertilization requirements would significantly increase the sustainability of rice, a staple crop for 50% of the world population (see c4rice.com). Furthermore, additional C_3 and C_4 plant species are being developed for comparative studies to better understand the evolution of C_4 traits. Potential model species of interest include the C_3 panicoid grass *Dichanthelium oligosanthes*, which diverged from the C_4 species *Setaria viridis* approximately 15 million years ago, representing a more recent divergence than most other C_3 and C_4 panicoid grasses (Studer et al., 2016).

An alternative to engineering C_4 photosynthesis into C_3 plants is using synthetic biology for improving photosynthesis (Kubis and Bar-Even, 2019). Possible mechanisms include engineering carbon concentrating mechanisms (Long et al., 2018), exploiting CAM mechanisms (DePaoli et al., 2014), or manipulating photorespiration (Maurino, 2019). Another possibility is to increase the performance of C_4 crops directly. Indeed, it was possible to increase CO_2 assimilation in maize by overexpressing Rubisco together with a chaperon, RUBISCO ASSEMBLY FACTOR 1 (RAF1), which resulted in fresh weight gain of the transgenic plants (Salesse-Smith et al., 2018). Alternatively, CO_2 assimilation was increased by overexpression of Rieske FeS protein of the Cytochrome b6f complex in *Setaria viridis* (Ermakova et al., 2019). These efforts, however, concentrate fully on carbon fixation and do not consider the nutritional aspects, neither with respect to the crop nutritional value nor the mineral nutrient homeostasis and use efficiency of the new crops. Nevertheless, while engineering C_4 crops is a very active area of research, it is unlikely to contribute significantly to food security or improved crop nutrition in the short term.

C_2 Engineering

Recently, Lundgren (2020) presented a compelling case for engineering C_2 photosynthesis into C_3 crop plants to improve photosynthetic performance in the face of climate change. The main argument made in favor of this approach is that C_2 photosynthesis is a stable intermediate physiological state between C_3 and C_4 metabolism that increases net carbon assimilation under high temperatures (Monson, 1989; Bellasio and Farquhar, 2019). But, importantly, C_2 photosynthesis does not require the complex anatomical changes associated with C_4 photosynthesis (Lundgren, 2020). This strategy could be useful in improving crop yields (or in mitigating yield declines) in the medium term. However, it is unclear how C_2 engineering will impact the nutritional status of crops, particularly under elevated CO_2, and the photosynthetic nutrient use efficiency. To the best of our knowledge, there are no FACE experiments evaluating the effects of elevated CO_2 on the yield or nutritional status of C_2 plants. Despite these limitations, this approach seems feasible for two reasons. First, C_2 engineering appears to be a necessary step toward C_4 engineering, suggesting that these efforts will not be wasted in the long term. Secondly, even if initial C_2 engineering has a negative impact on plant nutrition, when combined with traditional breeding approaches and additional engineering efforts, there is a high likelihood that these can be reverted. Thus, C_2 engineering of C_3 crops is likely to increase yield while maintaining or improving nutritional quality.

De Novo Domestication

Of the approximately 150 commonly cultivated crops worldwide, humans obtain almost 50% of their calories from just three crops - rice, wheat, and maize (Ross-Ibarra et al., 2007). This is in stark contrast to preagricultural humans who had significantly more diverse diets and achieved some level of domestication in approximately 2,500 plant species (Khoury et al., 2014; Smykal et al., 2017). Recent advances in genome editing technology have made *de novo* domestication of wild plants a viable option to design ideal crops for the future (Fernie and Yan, 2019). For example, a recent study targeting a small number of critical genes in the orphan Solanaceae crop "groundcherry" (*Physalis pruinosa*) was able to rapidly improve plant architecture and productivity (Lemmon et al., 2018). Because groundcherry is a semi-domesticated orphan crop in the same family as tomato, researchers quickly identified homologues of two domestication genes—*SELF PRUNING 5* and *CLAVATA1*. Using genome editing techniques to induce mutations in these genes resulted in an increased fruit size of over 20% and improved plant architecture (more compact growth), making groundcherry easier to grow and harvest. Furthermore, advances in multiplexing platforms that allow simultaneous genome editing

of six or more genes in a single transformation open the door for similar improvements to be made quickly in wild species (Zhang et al., 2016). Considering the small number of C_4 plant species that have been domesticated and the growing list of known domestication genes to target, there is good reason to believe the weeds of today could be the nutritious and sustainable foods of tomorrow.

CONCLUSIONS

C_4 crops play an essential role in human nutrition, and this role will probably be even stronger in the future. They are characterized by high productivity and adaptability to warm and dry climates and by their better water and nitrogen use efficiency than C_3 crops. While their yields will not directly benefit from elevated CO_2, their nutritional value is not predicted to be negatively affected. However, to unlock the full potential of C_4 crops for the future, more fundamental knowledge on the connection between mineral nutrition and C_4 photosynthesis needs to be generated. As outlined above, in particular nitrogen metabolism underwent significant alterations in the course of evolution of C_4 photosynthesis and might have been one of the evolutionary drivers. The increasing number and availability of new genomic and genetic resources and tools will enable us to extend the investigations of plant nutrition to a wider variety of C_4 and C_3–C_4 intermediate species, and at the same time, to include investigations of nutrient homeostasis in the general framework of C_4 photosynthesis research.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version.

FUNDING

Research in SK's lab is funded by the Deutsche Forschungsgemeinschaft (DFG) under Germany´s Excellence Strategy – EXC 2048/1 – project 390686111. IZ is supported by a DAAD fellowship.

REFERENCES

Amoroso, L. (2016). The Second International Conference on Nutrition: Implications for Hidden Hunger. *World Rev. Nutr. Diet* 115, 142–152. doi: 10.1159/000442100

Ashida, H., Mizohata, E., and Yokota, A. (2019). Learning RuBisCO's birth and subsequent environmental adaptation. *Biochem. Soc. Trans.* 47, 179–185. doi: 10.1042/BST20180449

Ashton, A. R., and Hatch, M. D. (1983). Regulation of C4 photosynthesis: regulation of pyruvate, Pi dikinase by ADP-dependent phosphorylation and dephosphorylation. *Biochem. Biophys. Res. Commun.* 115, 53–60. doi: 10.1016/0006-291x(83)90967-1

Aubry, S., Kelly, S., Kumpers, B. M., Smith-Unna, R. D., and Hibberd, J. M. (2014a). Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of C4 photosynthesis. *PLoS Genet.* 10, e1004365. doi: 10.1371/journal.pgen.1004365

Aubry, S., Smith-Unna, R. D., Boursnell, C. M., Kopriva, S., and Hibberd, J. M. (2014b). Transcript residency on ribosomes reveals a key role for the *Arabidopsis thaliana* bundle sheath in sulfur and glucosinolate metabolism. *Plant J.* 78, 659–673. doi: 10.1111/tpj.12502

Barbehenn, R. V., Chen, Z., Karowe, D. N., and Spickard, A. (2004). C3 grasses have higher nutritional quality than C4 grasses under ambient and elevated atmospheric CO2. *Global Change Biol.* 10, 1565–1575. doi: 10.1111/j.1365-2486.2004.00833.x

Bari, R., Datt Pant, B., Stitt, M., and Scheible, W. R. (2006). PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. *Plant Physiol.* 141, 988–999. doi: 10.1104/pp.106.079707

Becker, T. W., Carrayol, E., and Hirel, B. (2000). Glutamine synthetase and glutamate dehydrogenase isoforms in maize leaves: localization, relative proportion and their role in ammonium assimilation or nitrogen transport. *Planta* 211, 800–806. doi: 10.1007/s004250000355

Bellasio, C., and Farquhar, G. D. (2019). A leaf-level biochemical model simulating the introduction of C2 and C4 photosynthesis in C3 rice: gains, losses and metabolite fluxes. *New Phytol.* 223, 150–166. doi: 10.1111/nph.15787

Blatke, M. A., and Brautigam, A. (2019). Evolution of C4 photosynthesis predicted by constraint-based modelling. *Elife* 8, e49305. doi: 10.7554/eLife.49305

Bloom, A. J., Caldwell, R. M., Finazzo, J., Warner, R. L., and Weissbart, J. (1989). Oxygen and carbon dioxide fluxes from barley shoots depend on nitrate assimilation. *Plant Physiol.* 91, 352–356. doi: 10.1104/pp.91.1.352

Bloom, A. J., Burger, M., Rubio Asensio, J. S., and Cousins, A. B. (2010). Carbon dioxide enrichment inhibits nitrate assimilation in wheat and Arabidopsis. *Science* 328, 899–903. doi: 10.1126/science.1186440

Bloom, A. J., Asensio, J. S., Randall, L., Rachmilevitch, S., Cousins, A. B., and Carlisle, E. A. (2012). CO2 enrichment inhibits shoot nitrate assimilation in C3 but not C4 plants and slows growth under nitrate in C3 plants. *Ecology* 93, 355–367. doi: 10.1890/11-0485.1

Bowes, G. (1991). Growth at elevated CO2: photosynthetic responses mediated through Rubisco. *Plant Cell Environ.* 14, 795–806. doi: 10.1111/j.1365-3040.1991.tb01443.x

Brautigam, A., Hoffmann-Benning, S., and Weber, A. P. (2008). Comparative proteomics of chloroplast envelopes from C3 and C4 plants reveals specific adaptations of the plastid envelope to C4 photosynthesis and candidate proteins required for maintaining C4 metabolite fluxes. *Plant Physiol.* 148, 568–579. doi: 10.1104/pp.108.121012

Briat, J. F., Rouached, H., Tissot, N., Gaymard, F., and Dubos, C. (2015). Integration of P, S, Fe, and Zn nutrition signals in *Arabidopsis thaliana*: potential involvement of PHOSPHATE STARVATION RESPONSE 1 (PHR1). *Front. Plant Sci.* 6:290. doi: 10.3389/fpls.2015.00290

Brown, R. H. (1978). A difference in the nitrogen use efficiency of C3 and C4 plants and its implications in adaptation and evolution. *Crop Sci.* 18, 93–98. doi: 10.2135/cropsci1978.0011183X001800010025x

Burgener, M., Suter, M., Jones, S., and Brunold, C. (1998). Cyst(e)ine is the transport metabolite of assimilated sulfur from bundle-sheath to mesophyll cells in maize leaves. *Plant Physiol.* 116, 1315–1322. doi: 10.1104/pp.116.4.1315

Cabido, M., Pons, E., Cantero, J. J., Lewis, J. P., and Anton, A. (2008). Photosynthetic pathway variation among C4 grasses along a precipitation gradient in Argentina. *JBiogeog* 35, 131–140. doi: 10.1111/j.1365-2699.2007.01760.x

Caselato-Sousa, V. M., and Amaya-Farfán, J. (2012). State of knowledge on amaranth grain: a comprehensive review. *J. Food Sci.* 77, R93–R104. doi: 10.1111/j.1750-3841.2012.02645.x

Chao, Q., Liu, X. Y., Mei, Y. C., Gao, Z. F., Chen, Y. B., Qian, C. R., et al. (2014). Light-regulated phosphorylation of maize phosphoenolpyruvate carboxykinase

plays a vital role in its activity. *Plant Mol. Biol.* 85, 95–105. doi: 10.1007/s11103-014-0171-3

Cheng, W., Kimani, S. M., Kanno, T., Tang, S., Oo, A. Z., Tawaraya, K., et al. (2018). Forage rice varieties Fukuhibiki and Tachisuzuka emit larger CH4 than edible rice Haenuki. *Soil Sci. Plant Nutr.* 64, 77–83. doi: 10.1080/00380768.2017.1378569

Denton, A. K., Mass, J., Kulahoglu, C., Lercher, M. J., Brautigam, A., and Weber, A. P. (2017). Freeze-quenched maize mesophyll and bundle sheath separation uncovers bias in previous tissue-specific RNA-Seq data. *J. Exp. Bot.* 68, 147–160. doi: 10.1093/jxb/erw463

DePaoli, H. C., Borland, A. M., Tuskan, G. A., Cushman, J. C., and Yang, X. (2014). Synthetic biology as it relates to CAM photosynthesis: challenges and opportunities. *J. Exp. Bot.* 65, 3381–3393. doi: 10.1093/jxb/eru038

Doring, F., Streubel, M., Brautigam, A., and Gowik, U. (2016). Most photorespiratory genes are preferentially expressed in the bundle sheath cells of the C4 grass *Sorghum bicolor*. *J. Exp. Bot.* 67, 3053–3064. doi: 10.1093/jxb/erw041

Doulis, A. G., Debian, N., Kingston-Smith, A. H., and Foyer, C. H. (1997). Differential Localization of Antioxidants in Maize Leaves. *Plant Physiol.* 114, 1031–1037. doi: 10.1104/pp.114.3.1031

Ebi, K. L., and Loladze, I. (2019). Elevated atmospheric CO2 concentrations and climate change will affect our food's quality and quantity. *Lancet Planet Health* 3, e283–e284. doi: 10.1016/S2542-5196(19)30108-1

Edwards, G. E., Furbank, R. T., Hatch, M. D., and Osmond, C. B. (2001). What does it take to be C4? Lessons from the evolution of C4 photosynthesis. *Plant Physiol.* 125, 46–49. doi: 10.1104/pp.125.1.46

Ermakova, M., Lopez-Calcagno, P. E., Raines, C. A., Furbank, R. T., and Von Caemmerer, S. (2019). Overexpression of the Rieske FeS protein of the Cytochrome b 6 f complex increases C4 photosynthesis in *Setaria viridis*. *Commun. Biol.* 2, 314. doi: 10.1038/s42003-019-0561-9

Ermakova, M., Danila, F. R., Furbank, R. T., and Von Caemmerer, S. (2020). On the road to C4 rice: advances and perspectives. *Plant J.* 101, 940–950. doi: 10.1111/tpj.14562

Fernando, N., Panozzo, J., Tausz, M., Norton, R., Fitzgerald, G., and Seneweera, S. (2012). Rising atmospheric CO2 concentration affects mineral nutrient and protein concentration of wheat grain. *Food Chem.* 133, 1307–1311. doi: 10.1016/j.foodchem.2012.01.105

Fernie, A. R., and Yan, J. (2019). De Novo Domestication: An Alternative Route toward New Crops for the Future. *Mol. Plant* 12, 615–631. doi: 10.1016/j.molp.2019.03.016

Fliege, R., Flugge, U. I., Werdan, K., and Heldt, H. W. (1978). Specific transport of inorganic phosphate, 3-phosphoglycerate and triosephosphates across the inner membrane of the envelope in spinach chloroplasts. *Biochim. Biophys. Acta* 502, 232–247. doi: 10.1016/0005-2728(78)90045-2

Galili, G., and Amir, R. (2013). Fortifying plants with the essential amino acids lysine and methionine to improve nutritional quality. *Plant Biotechnol. J.* 11, 211–222. doi: 10.1111/pbi.12025

Gerlich, S. C., Walker, B. J., Krueger, S., and Kopriva, S. (2018). Sulfate Metabolism in C4 *Flaveria* Species Is Controlled by the Root and Connected to Serine Biosynthesis. *Plant Physiol.* 178, 565–582. doi: 10.1104/pp.18.00520

Gerwick, B. C., Ku, S. B., and Black, C. C. (1980). Initiation of sulfate activation: a variation in c4 photosynthesis plants. *Science* 209, 513–515. doi: 10.1126/science.209.4455.513

Ghannoum, O., Evans, J. R., Chow, W. S., Andrews, T. J., Conroy, J. P., and Von Caemmerer, S. (2005). Faster Rubisco is the key to superior nitrogen-use efficiency in NADP-malic enzyme relative to NAD-malic enzyme C4 grasses. *Plant Physiol.* 137, 638–650. doi: 10.1104/pp.104.054759

Ghannoum, O., Paul, M. J., Ward, J. L., Beale, M. H., Corol, D. I., and Conroy, J. P. (2008). The sensitivity of photosynthesis to phosphorus deficiency differs between C3 and C4 tropical grasses. *Func. Plant Biol.* 35, 213–221. doi: 10.1071/FP07256

Gigolashvili, T., and Kopriva, S. (2014). Transporters in plant sulfur metabolism. *Front. Plant Sci.* 5:442. doi: 10.3389/fpls.2014.00442

Halsted, M., and Lynch, J. (1996). Phosphorus responses of C3 and C4 species. *J. Exp. Bot.* 47, 497–505. doi: 10.1093/jxb/47.4.497

Ham, B. K., Chen, J., Yan, Y., and Lucas, W. J. (2018). Insights into plant phosphate sensing and signaling. *Curr. Opin. Biotechnol.* 49, 1–9. doi: 10.1016/j.copbio.2017.07.005

Harel, E., Lea, P. J., and Miflin, B. J. (1977). The localisation of enzymes of nitrogen assimilation in maize leaves and their activities during greening. *Planta* 134, 195–200. doi: 10.1007/BF00384971

Hatch, M. D., and Slack, C. R. (1966). Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. *Biochem. J.* 101, 103–111. doi: 10.1042/bj1010103

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P., et al. (2013). Predicting C4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell* 153, 1579–1588. doi: 10.1016/j.cell.2013.04.058

Heldt, H. W., Chon, C. J., Maronde, D., Herold, A., Stankovic, Z. S., Walker, D. A., et al. (1977). Role of orthophosphate and other factors in the regulation of starch formation in leaves and isolated chloroplasts. *Plant Physiol.* 59, 1146–1155. doi: 10.1104/pp.59.6.1146

Hennacy, J. H., and Jonikas, M. C. (2020). Prospects for Engineering Biophysical CO2 Concentrating Mechanisms into Land Plants to Enhance Yields. *Annu. Rev. Plant Biol.* 71, 461–485. doi: 10.1146/annurev-arplant-081519-040100

Hindu, V., Palacios-Rojas, N., Babu, R., Suwarno, W. B., Rashid, Z., Usha, R., et al. (2018). Identification and validation of genomic regions influencing kernel zinc and iron in maize. *Theor. Appl. Genet.* 131, 1443–1457. doi: 10.1007/s00122-018-3089-3

Hocking, P. J., and Meyer, C. P. (1991). Enrichment and Nitrogen Stress on Growth, and Partitioning of Dry Matter and Nitrogen in Wheat and Maize. *Austr. J. Plant Physiol.* 18, 339–356. doi: 10.1071/PP9910339

Hogy, P., Wieser, H., Kohler, P., Schwadorf, K., Breuer, J., Franzaring, J., et al. (2009). Effects of elevated CO2 on grain yield and quality of wheat: results from a 3-year free-air CO2 enrichment experiment. *Plant Biol. (Stuttg)* 11 (Suppl 1), 60–69. doi: 10.1111/j.1438-8677.2009.00230.x

Hu, B., Jiang, Z., Wang, W., Qiu, Y., Zhang, Z., Liu, Y., et al. (2019). Nitrate-NRT1.1B-SPX4 cascade integrates nitrogen and phosphorus signalling networks in plants. *Nat. Plants* 5, 401–413. doi: 10.1038/s41477-019-0384-1

Jacob, J., and Lawlor, D. W. (1992). Dependence of photosynthesis of sunflower and maize leaves on phosphate supply, ribulose-1,5-bisphosphate carboxylase/oxygenase activity, and ribulose-1,5-bisphosphate pool size. *Plant Physiol.* 98, 801–807. doi: 10.1104/pp.98.3.801

Jiao, J. A., and Chollet, R. (1991). Posttranslational regulation of phosphoenolpyruvate carboxylase in c(4) and crassulacean Acid metabolism plants. *Plant Physiol.* 95, 981–985. doi: 10.1104/pp.95.4.981

Jobe, T. O., Zenzen, I., Rahimzadeh Karvansara, P., and Kopriva, S. (2019). Integration of sulfate assimilation with carbon and nitrogen metabolism in transition from C3 to C4 photosynthesis. *J. Exp. Bot.* 70, 4211–4221. doi: 10.1093/jxb/erz250

John, C. R., Smith-Unna, R. D., Woodfield, H., Covshoff, S., and Hibberd, J. M. (2014). Evolutionary convergence of cell-specific gene expression in independent lineages of C4 grasses. *Plant Physiol.* 165, 62–75. doi: 10.1104/pp.114.238667

Khoury, C. K., Bjorkman, A. D., Dempewolf, H., Ramirez-Villegas, J., Guarino, L., Jarvis, A., et al. (2014). Increasing homogeneity in global food supplies and the implications for food security. *Proc. Natl. Acad. Sci. U. S. A.* 111, 4001–4006. doi: 10.1073/pnas.1313490111

Kocsy, G., Galiba, G., and Brunold, C. (2001). Role of glutathione in adaptation and signalling during chilling and cold acclimation in plants. *Physiol. Plant* 113, 158–164. doi: 10.1034/j.1399-3054.2001.1130202.x

Kopriva, S., and Chu, C. (2018). Are we ready to improve phosphorus homeostasis in rice? *J. Exp. Bot.* 69, 3515–3522. doi: 10.1093/jxb/ery163

Kopriva, S., and Koprivova, A. (2005). Sulfate assimilation and glutathione synthesis in C-4 plants. *Photosynt. Res.* 86, 363–372. doi: 10.1007/s11120-005-3482-z

Kopriva, S. (2011). "Nitrogen and Sulfur Metabolism in C-4 Plants," In: *C4 Photosynthesis and Related CO2 Concentrating Mechanisms. Advances in Photosynthesis and Respiration*. Raghavendra A., Sage R. (eds), (Dordrecht: Springer), vol 32, pp. 109–128. doi: 10.1007/978-90-481-9407-0_7

Koprivova, A., Melzer, M., Von Ballmoos, P., Mandel, T., Brunold, C., and Kopriva, S. (2001). Assimilatory sulfate reduction in C-3, C-3-C-4, and C-4 species of *Flaveria*. *Plant Physiol.* 127, 543–550. doi: 10.1104/pp.010144

Kubis, A., and Bar-Even, A. (2019). Synthetic biology approaches for improving photosynthesis. *J. Exp. Bot.* 70, 1425–1433. doi: 10.1093/jxb/erz029

Lee, Y., Nishizawa, T., Takemoto, M., Kumazaki, K., Yamashita, K., Hirata, K., et al. (2017). Structure of the triose-phosphate/phosphate translocator reveals the basis of substrate specificity. *Nat. Plants* 3, 825–832. doi: 10.1038/s41477-017-0022-8

Lemmon, Z. H., Reem, N. T., Dalrymple, J., Soyk, S., Swartwood, K. E., Rodriguez-Leal, D., et al. (2018). Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat. Plants* 4, 766–770. doi: 10.1038/s41477-018-0259-x

Livingstone, C. (2015). Zinc: physiology, deficiency, and parenteral nutrition. *Nutr. Clin. Pract.* 30, 371–382. doi: 10.1177/0884533615570376

Loladze, I. (2014). Hidden shift of the ionome of plants exposed to elevated CO(2) depletes minerals at the base of human nutrition. *Elife* 3, e02245. doi: 10.7554/eLife.02245

Long, B. M., Hee, W. Y., Sharwood, R. E., Rae, B. D., Kaines, S., Lim, Y. L., et al. (2018). Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in tobacco chloroplasts. *Nat. Commun.* 9, 3570. doi: 10.1038/s41467-018-06044-0

Lopez-Arredondo, D. L., Leyva-Gonzalez, M. A., Gonzalez-Morales, S. I., Lopez-Bucio, J., and Herrera-Estrella, L. (2014). Phosphate nutrition: improving low-phosphate tolerance in crops. *Annu. Rev. Plant Biol.* 65, 95–123. doi: 10.1146/annurev-arplant-050213-035949

Lundgren, M. R. (2020). C2 photosynthesis: a promising route towards crop improvement? *New Phytol.* in press. doi: 10.1111/nph.16494

Luo, Y., Su, B., Currie, W. S., Dukes, J. S., Finzi, A., Hartwig, U., et al. (2004). Progressive Nitrogen Limitation of Ecosystem Responses to Rising Atmospheric Carbon Dioxide. *BioScience* 54, 731–739. doi: 10.1641/0006-3568(2004)054[0731:PNLOER]2.0.CO;2

Majeran, W., Zybailov, B., Ytterberg, A. J., Dunsmore, J., Sun, Q., and Van Wijk, K. J. (2008). Consequences of C4 differentiation for chloroplast membrane proteomes in maize mesophyll and bundle sheath cells. *Mol. Cell Proteomics* 7, 1609–1638. doi: 10.1074/mcp.M800016-MCP200

Mallmann, J., Heckmann, D., Brautigam, A., Lercher, M. J., Weber, A. P., Westhoff, P., et al. (2014). The role of photorespiration during the evolution of C4 photosynthesis in the genus *Flaveria*. *Elife* 3, e02478. doi: 10.7554/eLife.02478

Maurino, V. G. (2019). Using energy-efficient synthetic biochemical pathways to bypass photorespiration. *Biochem. Soc. Trans.* 47, 1805–1813. doi: 10.1042/BST20190322

Medici, A., Szponarski, W., Dangeville, P., Safi, A., Dissanayake, I. M., Saenchai, C., et al. (2019). Identification of Molecular Integrators Shows that Nitrogen Actively Controls the Phosphate Starvation Response in Plants. *Plant Cell* 31, 1171–1184. doi: 10.1105/tpc.18.00656

Mellor, G. E., and Tregunna, E. B. (1971). The localization of nitrate-assimilating enzymes in leaves of plants with the C4-pathway of photosynthesis. *Can. J. Bot.* 49, 137–142. doi: 10.1139/b71-024

Mendoza-Cozatl, D. G., Gokul, A., Carelse, M. F., Jobe, T. O., Long, T. A., and Keyster, M. (2019). Keep talking: crosstalk between iron and sulfur networks fine-tunes growth and development to promote survival under iron limitation. *J. Exp. Bot.* 70, 4197–4210. doi: 10.1093/jxb/erz290

Micronutrient_Initiative (2009). *Investing in the future: A united call to action on vitamin and mineral deficiencies* (Ottawa: Micronutrient Initiative).

Millward, D. J. (1999). The nutritional value of plant-based diets in relation to human amino acid and protein requirements. *Proc. Nutr. Soc.* 58, 249–260. doi: 10.1017/s0029665199000348

Monson, R. K. (1989). The relative contributions of reduced photorespiration, and improved water-and nitrogen-use efficiencies, to the advantages of C3-C4 intermediate photosynthesis in *Flaveria*. *Oecologia* 80, 215–221. doi: 10.1007/BF00380154

Moore, R., and Black, C. C. (1979). Nitrogen Assimilation Pathways in Leaf Mesophyll and Bundle Sheath Cells of C(4) Photosynthesis Plants Formulated from Comparative Studies with Digitaria sanguinalis (L.) Scop. *Plant Physiol.* 64, 309–313. doi: 10.1104/pp.64.2.309

Morris, R. J., Fox, R. H., and Jung, G. A. (1982). Growth, P uptake and quality of warm and cool season grasses on a low available P soil. *Agron. J.* 74, 125–129. doi: 10.2134/agronj1982.00021962007400010032x

Myers, S. S., Zanobetti, A., Kloog, I., Huybers, P., Leakey, A. D., Bloom, A. J., et al. (2014). Increasing CO2 threatens human nutrition. *Nature* 510, 139–142. doi: 10.1038/nature13179

Myers, S. S., Smith, M. R., Guth, S., Golden, C. D., Vaitla, B., Mueller, N. D., et al. (2017). Climate Change and Global Food Systems: Potential Impacts on Food Security and Undernutrition. *Annu. Rev. Public Health* 38, 259–277. doi: 10.1146/annurev-publhealth-031816-044356

Newell, M. A., Vogel, K. E., Adams, M., Aydin, N., Bodnar, A. L., Ali, M., et al. (2014). Genetic and biochemical differences in populations bred for extremes in maize grain methionine concentration. *BMC Plant Biol.* 14:49. doi: 10.1186/1471-2229-14-49

Niro, S., D'Agostino, A., Fratianni, A., Cinquanta, L., and Panfili, G. (2019). Gluten-Free Alternative Grains: Nutritional Evaluation and Bioactive Compounds. *Foods* 8, E208. doi: 10.3390/foods8060208

Noctor, G., Mhamdi, A., Chaouch, S., Han, Y., Neukermans, J., Marquez-Garcia, B., et al. (2012). Glutathione in plants: an integrated overview. *Plant Cell Environ.* 35, 454–484. doi: 10.1111/j.1365-3040.2011.02400.x

Parcell, S. (2002). Sulfur in human nutrition and applications in medicine. *Altern. Med. Rev.* 7, 22–44.

Parry, M. A., Andralojc, P. J., Scales, J. C., Salvucci, M. E., Carmo-Silva, A. E., Alonso, H., et al. (2013). Rubisco activity and regulation as targets for crop improvement. *J. Exp. Bot.* 64, 717–730. doi: 10.1093/jxb/ers336

Passera, C., and Ghisi, R. (1982). ATP sulphurylase and O-acetylserine sulphydrylase in isolated mesophyll protoplasts and bundle sheath strands of S-deprived maize leaves. *J. Exp. Bot.* 33, 432–438. doi: 10.1093/jxb/33.3.432

Pastori, G. M., Mullineaux, P. M., and Foyer, C. H. (2000). Post-transcriptional regulation prevents accumulation of glutathione reductase protein and activity in the bundle sheath cells of maize. *Plant Physiol.* 122, 667–675. doi: 10.1104/pp.122.3.667

Pinto, H., Powell, J. R., Sharwood, R. E., Tissue, D. T., and Ghannoum, O. (2016). Variations in nitrogen use efficiency reflect the biochemical subtype while variations in water use efficiency reflect the evolutionary lineage of C4 grasses at inter-glacial CO2. *Plant Cell Environ.* 39, 514–526. doi: 10.1111/pce.12636

Planta, J., Xiang, X., Leustek, T., and Messing, J. (2017). Engineering sulfur storage in maize seed proteins without apparent yield loss. *Proc. Natl. Acad. Sci. U. S. A.* 114, 11386–11391. doi: 10.1073/pnas.1714805114

Pottier, M., Gilis, D., and Boutry, M. (2018). The Hidden Face of Rubisco. *Trends Plant Sci.* 23, 382–392. doi: 10.1016/j.tplants.2018.02.006

Rathnam, C. K., and Edwards, G. E. (1976). Distribution of Nitrate-assimilating Enzymes between Mesophyll Protoplasts and Bundle Sheath Cells in Leaves of Three Groups of C(4) Plants. *Plant Physiol.* 57, 881–885. doi: 10.1104/pp.57.6.881

Ross-Ibarra, J., Morrell, P. L., and Gaut, B. S. (2007). Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. Natl. Acad. Sci. U. S. A.* 104 (Suppl 1), 8641–8648. doi: 10.1073/pnas.0700643104

Rouached, H., Secco, D., Arpat, B., and Poirier, Y. (2011). The transcription factor PHR1 plays a key role in the regulation of sulfate shoot-to-root flux upon phosphate starvation in Arabidopsis. *BMC Plant Biol.* 11:19. doi: 10.1186/1471-2229-11-19

Sage, R. F., Pearcy, R. W., and Seemann, J. R. (1987). The Nitrogen Use Efficiency of C(3) and C(4) Plants : III. Leaf Nitrogen Effects on the Activity of Carboxylating Enzymes in *Chenopodium album* (L.) and *Amaranthus retroflexus* (L.). *Plant Physiol.* 85, 355–359. doi: 10.1104/pp.85.2.355

Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and the evolution of C4 photosynthesis. *Annu. Rev. Plant Biol.* 63, 19–47. doi: 10.1146/annurev-arplant-042811-105511

Saleh, A. S., Zhang, Q., Chen, J., and Shen, Q. (2013). Millet grains: nutritional quality, processing, and potential health benefits. *Compr. Rev. Food Sci. Food Saf.* 12, 281–295. doi: 10.1111/1541-4337.12012

Salesse-Smith, C. E., Sharwood, R. E., Busch, F. A., Kromdijk, J., Bardal, V., and Stern, D. B. (2018). Overexpression of Rubisco subunits with RAF1 increases Rubisco content in maize. *Nat. Plants* 4, 802–810. doi: 10.1038/s41477-018-0252-4

Schluter, U., and Weber, A. P. (2016). The Road to C4 Photosynthesis: Evolution of a Complex Trait via Intermediary States. *Plant Cell Physiol.* 57, 881–889. doi: 10.1093/pcp/pcw009

Schmutz, D., and Brunold, C. (1984). Intercellular Localization of Assimilatory Sulfate Reduction in Leaves of *Zea mays* and *Triticum aestivum*. *Plant Physiol.* 74, 866–870. doi: 10.1104/pp.74.4.866

Schmutz, D., and Brunold, C. (1985). Localization of nitrite and sulfite reductase in bundle sheath and mesophyll cells of maize leaves. *Physiol. Plant* 64, 523–528. doi: 10.1111/j.1399-3054.1985.tb08533.x

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. doi: 10.1111/tpj.13155

Shiferaw, B., Prasanna, B. M., Hellin, J., and Bänziger, M. (2011). Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Secur.* 3, 307–327. doi: 10.1007/s12571-011-0140-5

Shimono, H., Farquhar, G., Brookhouse, M., Busch, F. A., Grady, A. O., Tausz, M., et al. (2018). Prescreening in large populations as a tool for identifying elevated CO2-responsive genotypes in plants. *Funct. Plant Biol.* 46, 1–14. doi: 10.1071/FP18087

Slack, C. R., and Hatch, M. D. (1967). Comparative studies on the activity of carboxylases and other enzymes in relation to the new pathway of photosynthetic carbon dioxide fixation in tropical grasses. *Biochem. J.* 103, 660–665. doi: 10.1042/bj1030660

Smykal, P., Hradilova, I., Trneny, O., Brus, J., Rathore, A., Bariotakis, M., et al. (2017). Genomic diversity and macroecology of the crop wild relatives of domesticated pea. *Sci. Rep.* 7, 17384. doi: 10.1038/s41598-017-17623-4

Sonawane, B. V., Sharwood, R. E., Von Caemmerer, S., Whitney, S. M., and Ghannoum, O. (2017). Short-term thermal photosynthetic responses of C4 grasses are independent of the biochemical subtype. *J. Exp. Bot.* 68, 5583–5597. doi: 10.1093/jxb/erx350

Studer, A. J., Schnable, J. C., Weissmann, S., Kolbe, A. R., Mckain, M. R., Shao, Y., et al. (2016). The draft genome of the C3 panicoid grass species *Dichanthelium oligosanthes*. *Genome Biol.* 17, 223. doi: 10.1186/s13059-016-1080-3

Takahashi, H., Buchner, P., Yoshimoto, N., Hawkesford, M. J., and Shiu, S. H. (2011a). Evolutionary relationships and functional diversity of plant sulfate transporters. *Front. Plant Sci.* 2:119. doi: 10.3389/fpls.2011.00119

Takahashi, H., Kopriva, S., Giordano, M., Saito, K., and Hell, R. (2011b). Sulfur Assimilation in Photosynthetic Organisms: Molecular Functions and Regulations of Transporters and Assimilatory Enzymes. *Annu. Rev. Plant Biol.* 62, 157–184. doi: 10.1146/annurev-arplant-042110-103921

Taub, D. R. (2000). Climate and the U.S. distribution of C4 grass subfamilies and decarboxylation variants of C4 photosynthesis. *Am. J. Bot.* 87, 1211–1215. doi: 10.2307/2656659

Ujiie, K., Ishimaru, K., Hirotsu, N., Nagasaka, S., Miyakoshi, Y., Ota, M., et al. (2019). How elevated CO2 affects our nutrition in rice, and how we can deal with it. *PLoS One* 14, e0212840. doi: 10.1371/journal.pone.0212840

Vaughn, K. C., and Campbell, W. H. (1988). Immunogold localization of nitrate reductase in maize leaves. *Plant Physiol.* 88, 1354–1357. doi: 10.1104/pp.88.4.1354

Von Caemmerer, S., and Furbank, R. T. (2003). The C(4) pathway: an efficient CO (2) pump. *Photosynth. Res.* 77, 191–207. doi: 10.1023/A:1025830019591

Wang, W., Niu, S., Dai, Y., Zhai, X., Wang, M., Ding, Y., et al. (2019). Molecular Mechanisms Underlying Increase in Lysine Content of Waxy Maize through the Introgression of the opaque2 Allele. *Int. J. Mol. Sci.* 20, E684. doi: 10.3390/ijms20030684

Weckopp, S. C., and Kopriva, S. (2014). Are changes in sulfate assimilation pathway needed for evolution of C4 photosynthesis? *Front. Plant Sci.* 5:773. doi: 10.3389/fpls.2014.00773

Wurtzel, E. T., Cuttriss, A., and Vallabhaneni, R. (2012). Maize provitamin a carotenoids, current resources, and future metabolic engineering challenges. *Front. Plant Sci.* 3:29. doi: 10.3389/fpls.2012.00029

Xie, X., Hu, W., Fan, X., Chen, H., and Tang, M. (2019). Interactions Between Phosphorus, Zinc, and Iron Homeostasis in Nonmycorrhizal and Mycorrhizal Plants. *Front. Plant Sci.* 10:1172. doi: 10.3389/fpls.2019.01172

Zhang, G., Sakai, H., Tokida, T., Usui, Y., Zhu, C., Nakamura, H., et al. (2013). The effects of free-air CO(2) enrichment (FACE) on carbon and nitrogen accumulation in grains of rice (*Oryza sativa* L.). *J. Exp. Bot.* 64, 3179–3188. doi: 10.1093/jxb/ert154

Zhang, Z., Mao, Y., Ha, S., Liu, W., Botella, J. R., and Zhu, J. K. (2016). A multiplex CRISPR/Cas9 platform for fast and efficient editing of multiple genes in Arabidopsis. *Plant Cell Rep.* 35, 1519–1533. doi: 10.1007/s00299-015-1900-z

Zhu, C., Kobayashi, K., Loladze, I., Zhu, J., Jiang, Q., Xu, X., et al. (2018). Carbon dioxide (CO2) levels this century will alter the protein, micronutrients, and vitamin content of rice grains with potential health consequences for the poorest rice-dependent countries. *Sci. Adv.* 4, eaaq1012. doi: 10.1126/sciadv.aaq1012

Ziska, L. H., and Blumenthal, D. M. (2007). Empirical selection of cultivated oat in response to rising atmospheric carbon dioxide. *Crop Sci.* 47, 1547–1552. doi: 10.2135/cropsci2006.09.0616

Ziska, L. H., Manalo, P. A., and Ordonex, R. A. (1996). Intraspecific variation in the response of rice (Oryza sativa L.) to increased CO2 and temperature: growth and yield response of 17 cultivars. *J. Exp. Bot.* 47, 1353–1359. doi: 10.1093/jxb/47.9.1353

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 06 October 2020
doi: 10.3389/fpls.2020.572080

Independent Recruitment of Duplicated β-Subunit-Coding NAD-ME Genes Aided the Evolution of C4 Photosynthesis in Cleomaceae

Marcos A. Tronconi[1†], Meike Hüdig[2†], M. Eric Schranz[3] and Veronica G. Maurino[2]*

[1] Centro de Estudios Fotosintéticos y Bioquímicos (CEFOBI-CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Argentina, [2] Abteilung Molekulare Pflanzenphysiologie, Institut für Molekulare Physiologie und Biotechnologie der Pflanzen, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany, [3] Biosystematics Group, Wageningen University, Wageningen, Netherlands

OPEN ACCESS

Edited by:
Martha Ludwig, University of Western Australia, Australia

Reviewed by:
Gerald E. Edwards, Washington State University, United States
Sylvain Aubry, University of Zurich, Switzerland

***Correspondence:**
Veronica G. Maurino
vero.maurino@uni-bonn.de

[†] *These authors have contributed equally to this work*

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *15 June 2020*
Accepted: *14 September 2020*
Published: *06 October 2020*

Citation:
Tronconi MA, Hüdig M, Schranz ME and Maurino VG (2020) Independent Recruitment of Duplicated β-Subunit-Coding NAD-ME Genes Aided the Evolution of C4 Photosynthesis in Cleomaceae. Front. Plant Sci. 11:572080. doi: 10.3389/fpls.2020.572080

In different lineages of C_4 plants, the release of CO_2 by decarboxylation of a C_4 acid near rubisco is catalyzed by NADP-malic enzyme (ME) or NAD-ME, and the facultative use of phosphoenolpyruvate carboxykinase. The co-option of gene lineages during the evolution of C_4-NADP-ME has been thoroughly investigated, whereas that of C_4-NAD-ME has received less attention. In this work, we aimed at elucidating the mechanism of recruitment of *NAD-ME* for its function in the C_4 pathway by focusing on the eudicot family Cleomaceae. We identified a duplication of *NAD-ME* in vascular plants that generated the two paralogs lineages: α- and β-*NAD-ME*. Both gene lineages were retained across seed plants, and their fixation was likely driven by a degenerative process of sub-functionalization, which resulted in a NAD-ME operating primarily as a heteromer of α- and β-subunits. We found most angiosperm genomes maintain a 1:1 β-*NAD-ME*/α-*NAD-ME* (β/α) relative gene dosage, but with some notable exceptions mainly due to additional duplications of β-NAD-ME subunits. For example, a significantly high proportion of species with C_4-NAD-ME-type photosynthesis have a non-1:1 ratio of β/α. In the Brassicales, we found C_4 species with a 2:1 ratio due to a β-*NAD-ME* duplication (β*1* and β*2*); this was also observed in the C_3 *Tarenaya hassleriana* and Brassica crops. In the independently evolved C_4 species, *Gynandropsis gynandra* and *Cleome angustifolia*, all three genes were affected by C_4 evolution with α- and β1-*NAD-ME* driven by adaptive selection. In particular, the β1-NAD-MEs possess many differentially substituted amino acids compared with other species and the β2-NAD-MEs of the same species. Five of these amino acids are identically substituted in β1-NAD-ME of *G. gynandra* and *C. angustifolia*, two of them were identified as positively selected. Using synteny analysis, we established that β-*NAD-ME* duplications were derived from ancient polyploidy events and that α-*NAD-ME* is in a unique syntenic context in both Cleomaceae and Brassicaceae. We discuss our hypotheses for the evolution of NAD-ME and its recruitment for C_4 photosynthesis. We propose that gene duplications

provided the basis for the recruitment of NAD-ME in C_4 Cleomaceae and that all members of the *NAD-ME* gene family have been adapted to fit the C_4-biochemistry. Also, one of the β-*NAD-ME* gene copies was independently co-opted for its function in the C_4 pathway.

Keywords: C_4-photosynthesis, C_4-evolution, Cleomaceae, gene duplication, NAD-malic enzyme, subfunctionalization, neofunctionalization

INTRODUCTION

C_3 photosynthesis (Bassham et al., 1954) relies exclusively on ribulose-1,5-bisphosphate carboxylase oxygenase (rubisco) for carboxylase activity and evolved early in the history of life (Hayes, 1994). Rubisco is a bifunctional enzyme that catalyzes both the carboxylation and the oxygenation of its substrate ribulose-1,5-bisphosphate. One of the products of the oxygenase activity, 2-phosphoglycolate, is a toxic metabolite (Anderson, 1971; Kelly and Latzko, 1976; González-Moro et al., 1997). As a selective response to rubisco's promiscuity, plants evolved the energetically costly photorespiratory pathway (Maurino and Peterhansel, 2010). Until ∼400 million years ago (Mya), the rubisco oxygenase reaction was negligible due to elevated CO_2 and low O_2 levels in the atmosphere (Sage and Monson, 1999; Leakey and Lau, 2012). After this time, the onset of oxygenic photosynthesis introduced changes in atmospheric conditions such as high levels of O_2 which led to significant levels of costly photorespiration. Some land plants evolved a carbon concentrating mechanism, known as the C4 photosynthetic pathway, which resulted in reduction of the high-levels of photorespiration (Hatch, 1971; Heckmann et al., 2013). Nearly all the independent C_3 to C_4 transitions are dated to the mid-Oligocene (25–30 Mya), a time that was proceeded by a massive depletion of atmospheric CO_2 (Christin et al., 2008; Vicentini et al., 2008; Edwards et al., 2010). Under a wide range of environmental conditions, such as high temperatures, dryness, and high light intensities, plants possessing the C_4 biochemical pump are more efficient in terms of water and nitrogen use (Furbank and Hatch, 1987).

The initial step of the C_4 photosynthetic pathway is the fixation of inorganic carbon onto phosphoenolpyruvate (PEP) by PEP-carboxylase (PEPCase) to produce a four−carbon (C_4) acid (Hatch, 1987; Kanai and Edwards, 1999). The C_4 acid moves to the site of rubisco where specific decarboxylases release CO_2 (Drincovich et al., 2011). The decarboxylation reaction also produces a three−carbon acid, which diffuses back to the site of PEPCase where it is recycled to PEP. The C_4 cycle effectively acts as a CO_2 pump, increasing CO_2 levels around rubisco such that it nearly saturates the active site and thus reduces photorespiration to a minimal level (von Caemmerer, 2000).

Two major C_4 photosynthetic metabolic routes, known as the NAD-malic enzyme (ME) and NADP-ME subtypes, are distinguished by the primary ME decarboxylase used (Maier et al., 2011). In grasses, either of these subtypes can also make use of the facultative activity of PEPCase (Wang et al., 2014). C_4 photosynthesis is a complex convergent trait that arose independently in at least 66 plant linages (Sage et al., 2011, 2012), the majority of which use the NADP-ME subtype (Sage et al., 2011). The NAD-ME subtype is mostly found in eudicot species, where it is found in approximately 20 C_4-linages (Sage et al., 2011).

In plants using the NADP-ME subtype of C_4 metabolism, C_4-NADP-ME is present as a unique plastidial isoform with tissue specific expression and special regulatory properties, such as having pH dependent changes in oligomerization and substrate inhibition (Detarsio et al., 2007; Alvarez et al., 2019). The underlying molecular determinants for the distinction of C_4-NADP-ME from the non-photosynthetic isoform (nonC_4-NADP-ME) were recently characterized (Alvarez et al., 2019). The nonC_4-NADP-ME is present as multiple isoforms in higher plants, which are located to plastids and cytosol and show tissue specific expression (Saigo et al., 2004; Gerrard Wheeler et al., 2005; Detarsio et al., 2008; Gerrard Wheeler et al., 2008; Maurino et al., 2009).

In contrast to the well-described C_4-NADP-ME, C_4-NAD-ME has yet to be characterized at the molecular level. Early studies of *Amaranthus hypochondriacus* and various monocot species showed contradictory data for subunit composition and oligomeric states (Murata et al., 1989; Long et al., 1994). In plants, NAD-ME is exclusively present in mitochondria, where its core function is in L-malate respiration, as an associated enzyme of the tricarboxylic acid cycle (Grover et al., 1981; Artus and Edwards, 1985; Tronconi et al., 2008; Fuchs et al., 2020). In *Arabidopsis thaliana*, NAD-ME functions as a homo- and/or heterodimer of two distinct, homologs proteins (∼65% sequence identity): known as the α-subunits (to which AtNAD-ME1 belongs) and β-subunits (to which AtNAD-ME2 belong) and with molecular masses in the range of 58 and 63 kDa (Tronconi et al., 2008, 2010a). The α- and β-NAD-ME share only 40% of identity with the NADP-ME isoforms, owing to the fact that the *NAD-ME* and *NADP-ME* genes were acquired in independent evolutionary processes in plants (Tronconi et al., 2018).

Modifications in function and expression are key components of the evolutionary transition of several enzymes involved in the C_4 carbon concentrating mechanism. Gene duplication was proposed as a precondition for the evolution of C_4 activities, as it facilitates C_4-specific adaptive changes in *cis*-regulatory control regions as well as in coding regions (Lynch and Conery, 2000; Moore and Purugganan, 2005). The evolution of C_4-NADP-ME likely followed this process of gene duplication and neofunctionalization, starting from a plastidial non-C_4 isoform-coding gene and including the acquisition of a bundle sheath cell-specific expression pattern (Maurino et al., 2001; Tausta et al., 2002; Saigo et al., 2004; Christin et al., 2009). The C_4 specific isoforms of NADP-ME evolved at least five times independently in this way in grasses (Christin et al., 2009). In contrast, the

evolutionary history of C_4-NAD-ME remains elusive, as to date a gene coding for a C_4-specific isoform has not been identified (Murata et al., 1989; Long et al., 1994).

Here, we aimed at elucidating the mechanism of recruitment of NAD-ME for its function in the C_4 pathway by focusing on the eudicot family Cleomaceae. This family contains species spanning a developmental progression from C_3 to C_4 photosynthesis and at least three separate origins of C_4 lineages (Feodorova et al., 2010; Koteyeva et al., 2011). Cleomaceae belongs to the order Brassicales and is a sister group of the Brassicaceae, which contains one of the best studied plant species, *Arabidopsis thaliana*, for which vast amounts of -omics data are available for comparative analyses. Cleomaceae and Brassicaceae share the At-beta whole-genome duplication (WGD) event, which was estimated to have occurred 75–100 Mya (Edger et al., 2015). The Cleomaceae and Brassicaceae lineages diverged 41 Mya and more recently underwent the independent At-alpha (23–34 Mya) and Cs-alpha (14–20 Mya) WGDs (Schranz and Mitchell-Olds, 2006; Barker et al., 2009). Despite different patterns of gene loss and retention and chromosomal rearrangements after polyploidy, the genomes of Cleomaceae and Brassicaceae species show detectable synteny (Schranz and Mitchell-Olds, 2006). Moreover, there exist no significant differences in gene copy numbers between C_3 and C_4 *Cleome* species (van den Bergh et al., 2014).

Our comprehensive analyses indicate that a duplication of the *NAD-ME* gene during the evolution of vascular plants resulted in two paralogs lineages, α- and β-*NAD-ME*, which were retained during seed plant evolution and diversification. We propose that the heteromeric assembly of NAD-ME was established by sub-functionalization of the duplicated NAD-ME genes. Later, neo-functionalization optimized the α- and β-NAD-ME functions and changes in the subunit-specific duplications provided the basis for the recruitment of NAD-ME in C_4 biochemistry. We found that in Cleomaceae all *NAD-ME* genes were affected by C_4 evolution, where one of the β-*NAD-ME* gene copies was co-opted for its function in the C_4 pathway.

MATERIALS AND METHODS

Sequence Retrieval

For species with entire genome information, NAD-ME coding sequences were extracted from primary gene models www.phytozome.net. Sequences from Cleomaceae species were acquired from transcriptome data (Kulahoglu et al., 2014; Mabry et al., 2019); in case of ***Gynandropsis gynandra*** (C_4), ***Cleome angustifolia*** (C_4), and ***Tarenaya hassleriana*** (C_3) the sequences were verified and correctly assembled using cDNA-based sequencing as the transcriptomes showed misassembled transcripts for several NAD-ME genes. Sequences for Chara braunii, Azolla filiculoides, Salvinia cucullata, and Panicum miliaceum were identified from their respective genome publications (Li et al., 2018; Nishiyama et al., 2018; Zou et al., 2019). Additional fern sequences were extracted from available large-scale transcriptomic data (Shen et al., 2018). Ginkgo biloba, Amborella trichopoda, Taxus baccata, Pinus pinaster, Pinus sylvestris, and Pseudotsuga menziesii sequences were collected via PLAZA 3.0 (Proost et al., 2015). Accession numbers of all NAD-ME coding sequences used in this work are listed in **Supplementary Table 1**. BLASTP with the BLOSUM62 as default scoring matrix and a minimal e-value of 0.0001 was implemented to obtain homologs using AtNAD-ME1 (AT2G13560) and AtNAD-ME2 (AT4G00570) as query. All sequences were manually checked for correct translation start sites and the presence of conserved amino acid regions found in all NAD(P)-ME (Tronconi et al., 2018). Mitochondrial localization was verified using the program Target P (Emanuelsson et al., 2000).

Multiple Sequence Alignments

A data set of the coding sequences was assembled using MEGA X (v.10.0.5) (Kumar et al., 2018). The sequences were then translated into amino acids and aligned using Muscle (Edgar, 2004) with the gap opening penalty value of −2.9 and without penalizing its extension. Once retranslated into nucleotides, the alignment was manually edited to select the most-reliable positions in the alignment, assisted by Gblocks[1] and TrimAl[2] programs. Since the different phylogenetic methods consider columns with gaps in different ways, we applied a stringent criterion by eliminating codons with coverage less than 95%. The final multiple sequence alignment (MSA) consisted of 240 coding sequences from 118 species with 1,731 nucleotides positions corresponding to 577 codons.

Phylogenetic Analyses

Bayesian inference (BI) was performed using MrBayes 3.1.2 software (Ronquist and Huelsenbeck, 2003). Two parallel runs, each including four Metropolis-coupled Markov chain Monte Carlo (MC3) analyses, were run for 5,000,000 generations and sampled every 1,000 generations. This generated an output of 5,000 trees per run. A site-specific rate model (partition scheme) was used. The characters in the MSA were divided into three sets corresponding to the codon positions. Each position has its own rate labeled m1 in case of the first codons site, m2 in case of the second codons site and m3 in case of the third codons site. For the tree inferred from the third positions we defined a one-partition scheme by excluding the characters represented by the first and second sites in the MSA. For an efficient Metropolis coupling, an incremental heating scheme of three heated chains and one cold chain in each run was used with a temperature parameter setting of 0.1. The final average standard deviation of split frequencies was used as the convergence index (values <0.01 indicated good convergence). The convergence of clade posterior probabilities within and between runs was checked using the potential scale reduction factor. The initial 25% of the sampled trees for each MC3 run were discarded as "burn-in" and the post-burn-in trees from the two runs were integrated to generate a 50% majority-rule consensus tree. The percentage of samples recovering any particular clade in a BI analysis represents the posterior probability (BPP) of a

[1]http://molevol.cmima.csic.es/castresana/Gblocks.html
[2]http://trimal.cgenomics.org/

clade. In all cases, a GTR (General Time Reversible) model with base frequencies gamma shape parameter (G) and proportion of invariants sites (I) was set. All the active parameters in the GTR + G + I model were optimized separately for each position of the codons. For analyses using third codon positions, the number of chains in each run of was increased from four to five due to convergence conflicts. Nodes with BPP values >90% were considered highly supported.

Maximum likelihood (ML) and neighbor joining (NJ) analyses on the whole data set were conducted using MEGA X (v.10.0.5). The goodness of fit of each model to the data was measured by the Bayesian information criterion (BIC) and the model with the lowest BIC score was considered the best description for a specific substitution pattern. The initial tree for the ML search was generated automatically by applying the NJ and BIONJ algorithms, and its branch lengths were adjusted to maximize the likelihood of the data set for that tree topology under the selected model of evolution. Heuristic searches were conducted with the initial tree based on the nearest neighbor interchange (NNI) search where the alternative trees differ in one branching pattern. Reliability of interior branches was assessed with 2,000 bootstrap (B) re-samplings. Nodes with MLB or NJB values 50–69% were regarded as weakly supported, 70–84% as moderately supported, and 85–100% as strongly supported (Hillis and Bull, 1993). The tree files were saved in Newick format (.nwk) containing all the relevant clade support values and branch length information. The trees were displayed using the FigTree v1.4.4 software and edited by rotating nodes and compressing lineages that were designated by their subdivision, class, order, or family names.

Differential Substitution Analysis

A strictly differentially substituted position is one for which the NAD-ME sequences of the C_4 species, *G. gynandra* and *Cleome angustifolia* (recently reclassified as *Coalisina angustifolia*), contained an identical amino acid, while a second, different amino acid was shared in all other NAD-ME sequences. For the differential substitution analysis, we used the whole data set of NAD-MEs of the Brassicales obtained in this work. MSAs of α- and β-NAD-MEs were computed using the MAFFT algorithm (v7.427) with the iterative refinement method L-INS-i (Katoh et al., 2005; Kuraku et al., 2013). The sequences were aligned using the online tool integrated MKT[3]. The best amino acid substitution model based on each MSA was estimated using MEGA X (v.10.0.5) (Kumar et al., 2018). We used the MSAs to identify amino acid positions that are strictly differentially conserved in the α-NAD-ME and β-NAD-ME sequences of the C_4 species as previously performed by Alvarez et al. (2019).

Synteny Analysis

To further investigate the evolutionary relationships and duplication of Brassicaceae and Cleomaceae β-*NAD-ME* and α-*NAD-ME* genes, syntenic analysis was performed with the SynFind tool using the default parameters (last algorithm, window size set to 40 genes and with a minimum number of collinear anchors of 4) (Tang et al., 2015). For this analysis, a limited set of representative taxa was used based on phylogenetic breadth (see **Figure 3**), presence of ancient polyploidy (i.e., *Brassica* and *Gynandropsis*) and the availability of high-quality annotated draft genomes. For Cleomaceae, the genomes of *T. hassleriana*, *G. gynandra*, and *Cleome violacea* were included. For Brassicaceae we included *Arabidopsis thaliana*, *Arabis alpina*, *Brassica rapa*, and *Eutrema salsugineum*. As an outgroup representative, we included *Citrus clementina*. For our SynFind analyses, we used either *G. gynandra*, *A. thaliana* or *C. violacea* orthologs of β-NAD-ME and α-NAD-ME genes. The results of our synteny analyses were compared and visualized using GEvo (Tang et al., 2015).

Positive Selection Tests and Statistical Analysis

For testing sites that underwent positive selection during the evolution of C_4 species in the α-NAD-ME and β-NAD-ME coding sequences, different site-class specific models were employed using the software codeml, implemented in the PAML package (Yang, 2007). Because no gene lineages leading to C_4-specific NAD-MEs were identified, we conducted a site-class-specific approach. The models assume that the ω = dN/dS (non-synonymous to synonymous substitution rates) take a value of 1 under neutral evolution. Positive and purifying (negative) selection are indicated when $\omega > 1$ and $\omega < 1$, respectively. The codon substitution models were: M0 (one-ratio) M1a (nearly neutral), M2a (positive selection), M3 (discrete), M7 (beta), M8 (beta and $\omega > 1$), and M8a (beta and $\omega = 1$) (Yang et al., 2000). The fit of these models to the sequence data was compared using likelihood-ratio test (LRT). When an LRT test yielded a significant result for any of the pairwise comparisons, the Bayes empirical Bayes (BEB) method was used to identify amino acids residues that have evolved under selection. Posterior probability of 0.90 was selected as the standard threshold for identifying residues under selection (Scheffler and Seoighe, 2005).

Significance of deviations from 1:1 β-NAD-ME/α-NAD-ME was determined according to the Fisher's exact test using the SigmaStat software (Systat Software, Inc.).

RESULTS

Identification of Two Major Clades of NAD-ME Genes (α- and β) in all Seed Plants

To cover the early and late evolution of NAD-ME proteins of land plants (Embryophyta), we analyzed a dataset of 253 coding sequences (**Supplementary Table 1**) from genomes and assembled transcripts from liverworts (Hepaticophyta), mosses (Bryophyta), early branching vascular plants (Lycophyta), ferns (Filicopsida), gymnosperms, angiosperms, and streptophyte algae (Charophytes), with chlorophyte algae (Chlorophyta) as the out-group. The coding sequences were aligned and used for BI and ML analyses to infer the phylogenetic gene and protein trees.

[3]https://imkt.uab.cat

We found all phylogenetic trees to be globally congruent regardless of the phylogenetic approach used (coding sequences-based BI tree in **Figure 1** and **Supplementary Data 1**, coding sequences-based ML tree in **Supplementary Data 2**, and protein-based ML tree in **Supplementary Data 3**). The α- and β-*NAD-ME* genes formed well supported orthologs clades within the seed plants (spermatophytes). Both paralogs genes were retained across all gymnosperms and angiosperms analyzed and orthologs of α- and β-*NAD-ME* were not found in all other non-seed plants. In all the trees we found fern homologs confidently placed as a sister group to the α-*NAD-ME* of seed plants with Bayesian posterior probability (BPP) = 100% (**Figure 1** and **Supplementary Data 1**) and MLB = 92 and 86% (**Supplementary Data 2** and **Supplementary Data 3**, respectively). Similarly, we found no evidence for α- and β-*NAD-ME* duplicates in streptophyte algae, liverwort, mosses, nor the early branching tracheophyte *S. moellendorffii*. Given these data, the duplication of the *NAD-ME* gene and the fixation of both paralogs occurred at the foundation of the seed plants.

Deviations From 1:1 β-*NAD-ME*/α-*NAD-ME* (β/α) Among NAD-ME of C_3 and C_4-Species

All 92 angiosperm species analyzed have retained orthologs of both NAD-ME α- and β coding genes and with most having a 1:1 β-*NAD-ME*/α-*NAD-ME* (β/α) relative gene dosage. However, changes in the relative genetic dosage could be identified (**Figure 2**). Twenty species have at least two β copies for each α gene, and only one species, *Amaranthus hypochondriacus* (Carophylles), has two α genes for one β (**Figure 2**). The most pronounced change was found in *Glycine max* with a β/α gene ratio of 4:1. In eudicots, the NFC (Nitrogen Fixing Clade including: Rosales, Fabales, Cucurbitales, and Fagales) and COM (Celastrales, Oxalideles, and Malpighiales) clades contain 42% of the species with a non-1:1 ratio. Species having C_4-NAD-ME photosynthesis, with the exception of *Panicum halli* and *Panicum miliaceum* (Poaceae), have a β/α relation deviating from 1:1. This proportion (71.5%) is significantly higher than that observed for

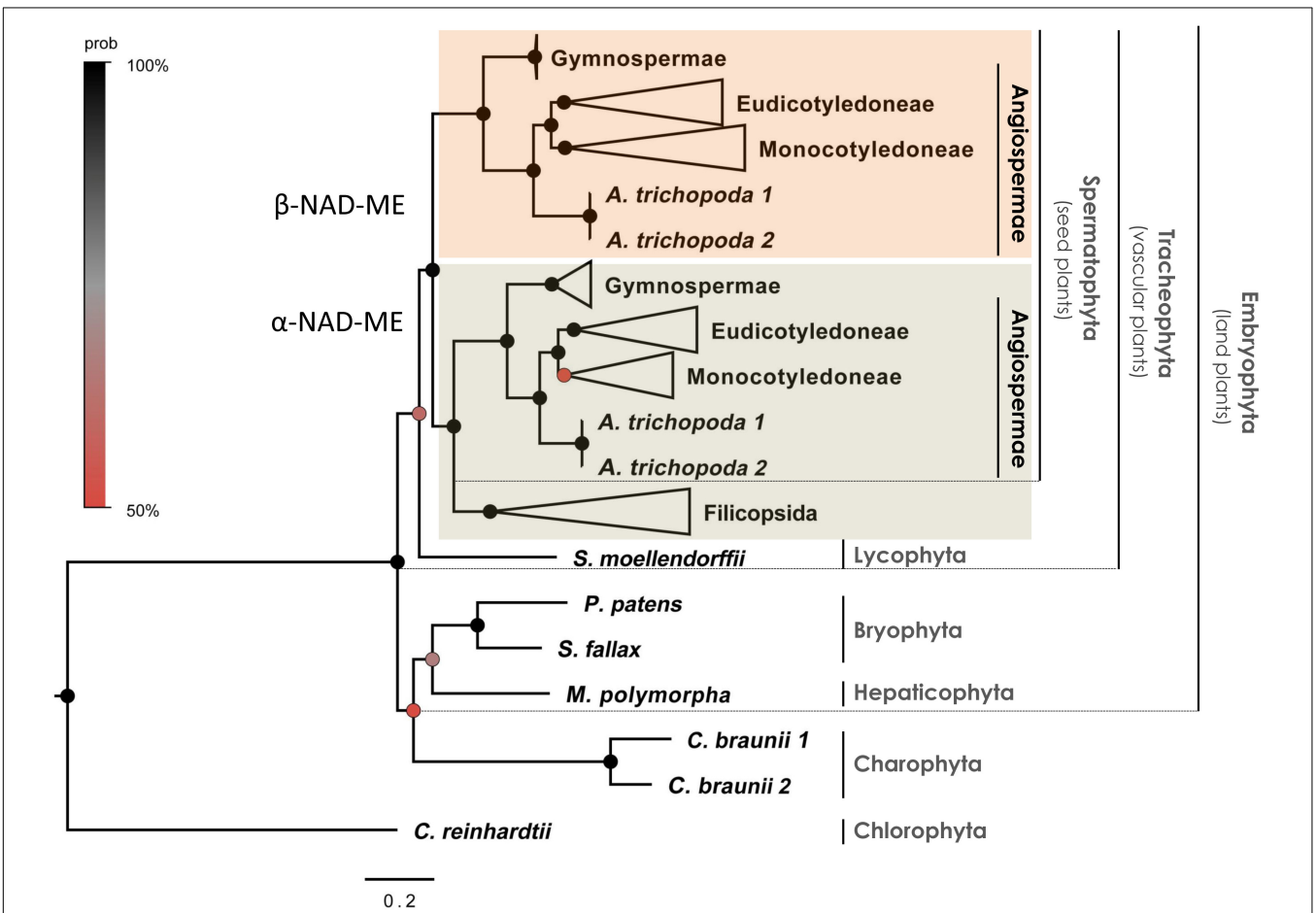

FIGURE 1 | Phylogenetic gene tree of α- and β-*NAD-ME* in the green lineage. The phylogenetic relationships were inferred from coding sequences using BI. The main clades are compressed and designated by their subdivision, class, order or family names. Black circles highlight nodes with Bayesian Posterior Probability (BPP) values higher than 90%, gray circles indicate BPP values higher than 70% and red circles indicate BPP values higher than 50%. The three positions of codons were optimized separately (partitioned tree). Best-fit substitution model was a GRT + G submodel (BPP = 0.90 ± 0.01) with individual rates values: m1 = 0.42, m2 = 0.33, and m3 = 2.74 for the 1st, 2nd, and 3rd position of codons, respectively. The full tree is available in **Supplementary Data 1**.

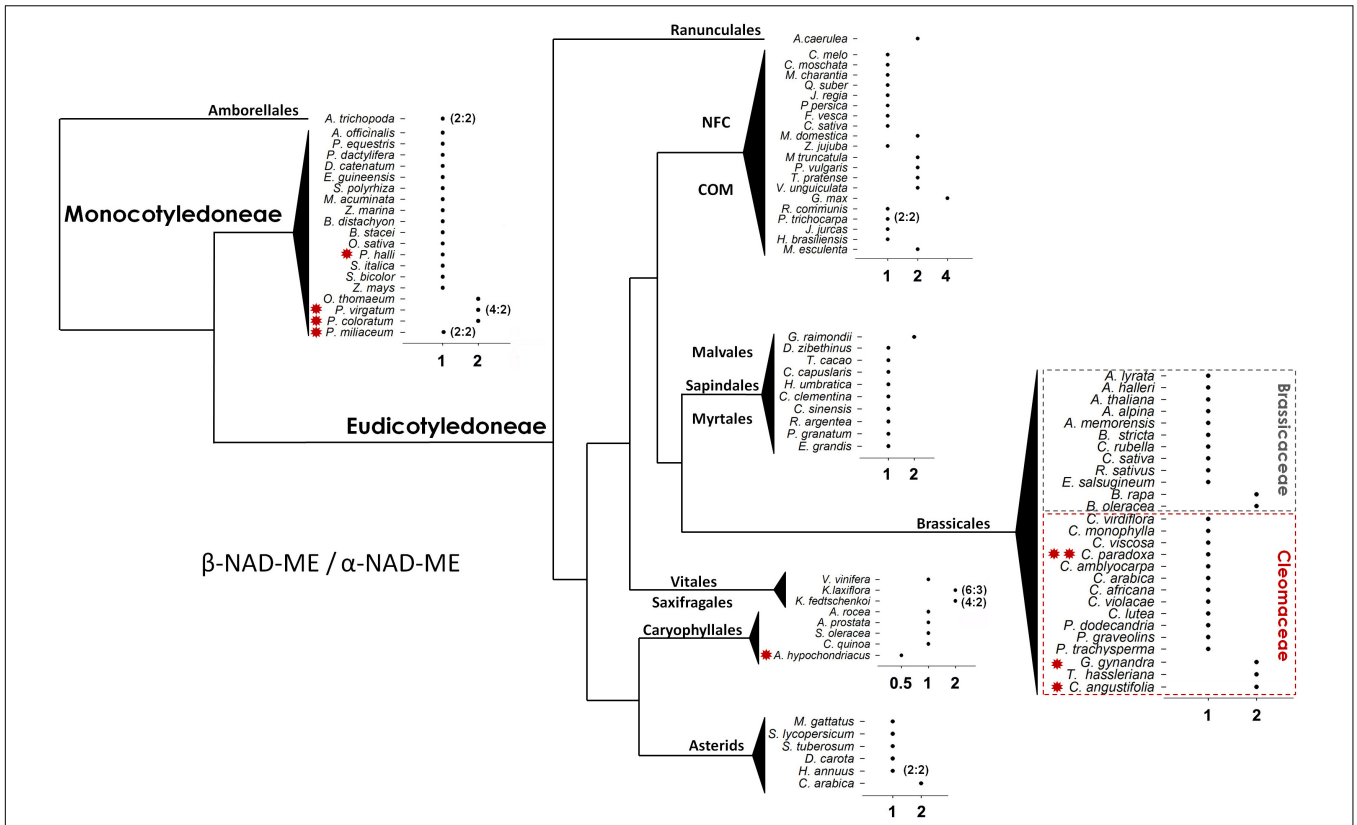

FIGURE 2 | Scheme showing the ratio of β-*NAD-ME* to α-*NAD-ME* gene composition in Angiosperms. Most species have the same copy number (1 to 1). In the cases where the relative proportion and absolute composition differ, the absolute β to α gene dosage is indicated in parentheses. Schematic representation of the values with the species grouped according to their order and family interrelationships (APG IV, 2016). COM: Celastrales, Malpighiales, and Oxalideles. NFC (Nitrogen Fixing Clades): Rosales, Fabales, Cucurbitales, and Fagales. Red * and ** indicate C_4-NAD-ME and C_3–C_4 photosynthetic metabolism, respectively. The scheme was created as a free art illustration where the lines do not represent branch lengths.

non-C_4 NAD-ME species (18.6%) (Fisher's exact test, $P = 0.006$). The C_3–C_4 intermediate *Cleome paradoxa* does maintain the 1:1 β/α ratio.

Evolution of α- and β-NAD-ME Proteins in Cleomaceae and Brassicaceae

Gene duplication plays a critical role in generating the diversity needed for the evolution of protein through neo-functionalization. We wanted to investigate the role of gene duplication to the evolution of NAD-ME type C_4 photosynthetic metabolism in Eudicot species. NAD-ME has been independently co-opted for C_4 photosynthesis in Caryophyllales and Brassicales species. Thus, we further investigated the pattern of the NAD-ME protein evolution in the Cleomaceae (having both C_3 and C_4 species) and in its sister-family the Brassicaceae (including Arabidopsis and Brassica crops). Most species in both families have a 1:1 ratio of β/α relative gene dosage (**Figure 2**). However, both families possess species carrying an additional copy of the β-*NAD-ME* gene (β*1* and β*2*) (**Figure 2**) that correlate with mesopolyploidy events. In Cleomaceae, the β-*NAD-ME* gene duplication is found in the C_3 species *T. hassleriana* and the C_4 species *C. angustifolia* and *G. gynandra* (**Figure 2**). *T. hassleriana* and *G. gynandra* share the Th-alpha ancient polyploidy event, but *C. angustifolia* underwent an independent polyploidy event (Mabry et al., 2019). In Brassicaceae, both *B. rapa* and *B. oleracea* have duplicated β-*NAD-ME* gene associated with the Br-α genome triplication.

The NJ (**Figure 3** and **Supplementary Data 4**) and ML (**Supplementary Data 5**) phylogenetic protein tree topologies were very consistent despite the low support values of some branches (MLB and NJB <50%). For closely related species, the protein-based tree has fewer informative sites than the CDS-based tree (**Figure 1**). However, we were interested in evaluating unusual branch positions and lengths that would suggest evolutionary rate shifts and altered amino acidic sequences. In the Brassicaceae, the relationships of the orthologs in the α-NAD-ME and β-NAD-ME clades agree with recent assessments of the species phylogenetic relationships (Nikolov and Tsiantis, 2017; Bayat et al., 2018). Conversely, α- and β-NAD-ME protein trees of Cleomaceae are incongruent with current species trees (Feodorova et al., 2010; Patchell et al., 2014).

Regarding the α-NAD-ME orthologs, we recovered the core *Cleome sensu stricto* (*s. str.*) (*C. violacea*, *C. arabica*, *C. africana*, and *C. amblyocarpa*) and the clade containing *Polanisia* species

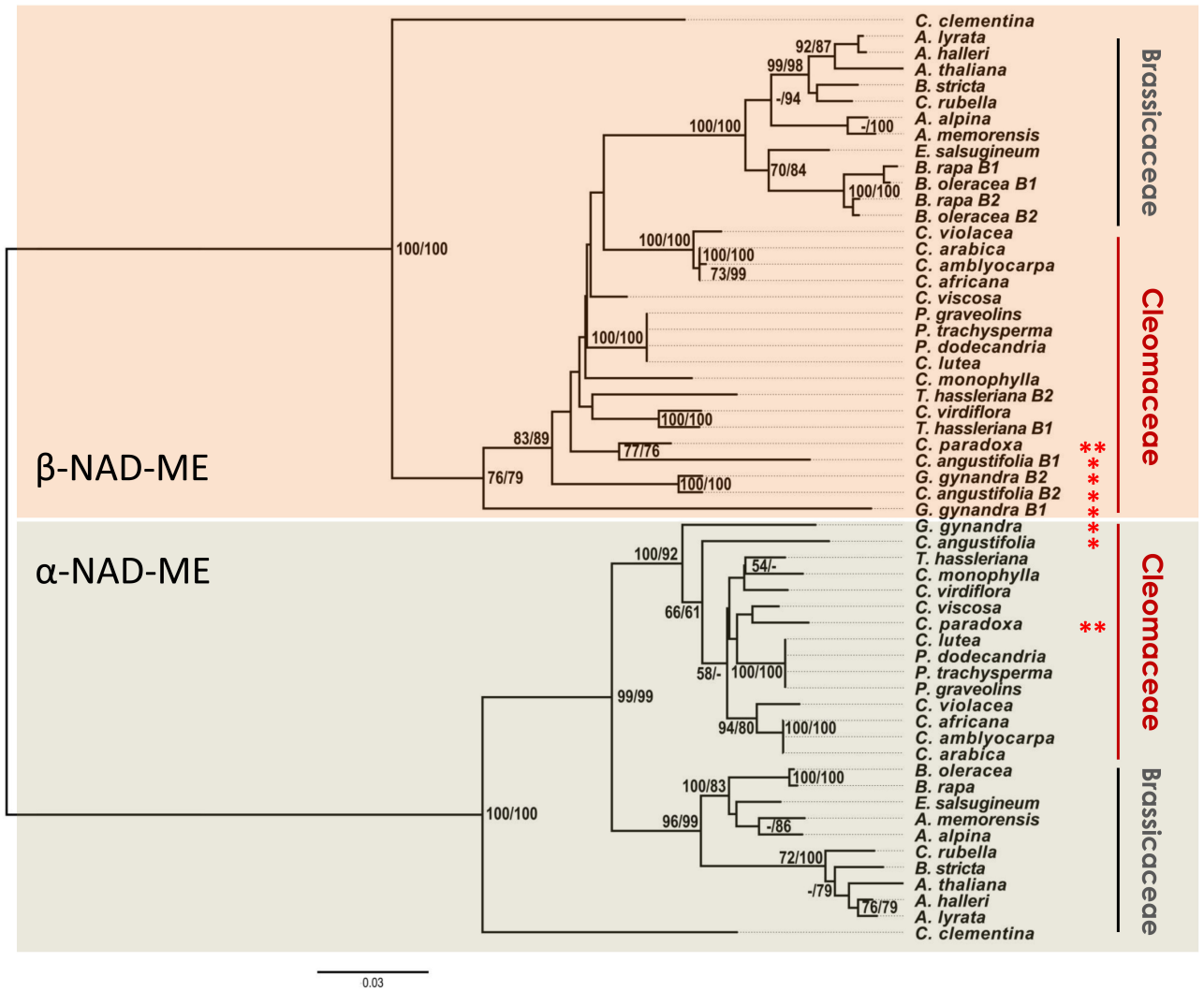

FIGURE 3 | Phylogenetic gene tree of NAD-ME proteins of Cleomaceae and Brassicaceae. The evolutionary history was inferred using the ML and NJ methods. The NJ optimal tree with the sum of branch length = 1.58 is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The evolutionary distances were computed using the JTT matrix-based method with a discrete Gamma distribution to model evolutionary rate differences among sites (G, parameter = 0.52). Support values of 2,000 bootstrap replicates are given as MLB/NJB next to the branches when greater than 50%. The analysis involved 57 amino acid sequences and there were 227 parsimony-informative sites over a total of 597 positions (less than 40%) in the final dataset. Red ∗ and ∗∗ indicate C_4 and C_3–C_4 photosynthetic metabolism, respectively. *Citrus clementina* NAD-ME isoforms were used as out groups. The full trees are available in **Supplementary Data 4** and **Supplementary Data 5**.

with well-supported MLB and NJB values (Patchell et al., 2014). However, the α-NAD-ME of the C_4 species *G. gynandra* and *C. angustifolia* were placed as the base of the clade within the Cleomaceae (**Figure 3**), which disagrees with known species relationships.

In the β-NAD-ME clade, the conflicting pattern of protein relationships from known species relationships is even more pronounced. Here, the β-NAD-ME of Cleomaceae was recovered as a non-monophyletic group with Brassicaceae as a sister group of the Cleome *s. str.* (**Figure 3**). Within the Brassicaceae we recovered the same groups as described in the α-NAD-ME clade. Even more so, the β-NAD-ME of the C_4 species *G. gynandra* and *C. angustifolia* appear as first-branching lineages that precede the formation of Brassicaceae. This whole pattern could be explained by an unrealistic evolutionary scenario in which the β-NAD-ME of the C_4 species *G. gynandra* and *C. angustifolia* are encoded by ancient genes retained only in the genome of these species. Instead, the differences in rates of molecular evolution within the Cleomaceae are most probably due to selective pressures on the NAD-ME genes of the C_4 species. In addition to the incongruent positioning in the protein-based trees, the α- and β-NAD-ME of the C_4 species show branch lengths that do not correlate with those of the C_3 orthologs in Cleomaceae (**Supplementary Data 4**). At least in *G. gynandra*, such a high number of non-synonymous substitutions cannot be

a consequence of pseudogenization, as α- and β*2-NAD-ME* transcripts are highly abundant in leaves (Brown et al., 2011; Kulahoglu et al., 2014).

Congruency Between NAD-ME Gene-Based Trees and Cleomaceae Species Trees

We hypothesize that evolutionary forces driving the evolution of C_4 in the Cleomaceae could be responsible for the incongruent NAD-ME protein-based phylogenetic trees. To address this hypothesis, we inferred phylogenetic trees from the third position of the codons, which is considered as a nearly neutral marker (Christin et al., 2007; Hilu et al., 2014). We conducted this analysis focused on the NAD-ME coding sequences of the Rosids, a well-defined monophyletic group capable of reconstructing true phylogenetic relationships given the high phylogenetic signals and low homoplasy (Hilu et al., 2014).

The BI (**Figure 4**, right and **Supplementary Data 6**) and ML (**Figure 4**, left and **Supplementary Data 7**) phylogenetic trees deduced from unconstrained sites were widely congruent with well-supported nodes. In the α-*NAD-ME* clade, we recovered the expected topology based on published species phylogenies inferred from plastid, mitochondrial and nuclear markers (Feodorova et al., 2010; Patchell et al., 2014). We found that *Polanisia* and the *Cleome s. str.* species cluster as basal lineages of C_3 species. The C_4 species *G. gynandra* and *C. angustifolia* form paraphyletic groups, with the C_3-C_4 intermediate *C. paradoxa* confidently placed in the Angustifolia clade and *G. gynandra* nested close to the African species *C. monophylla* (**Figure 4**). Finally, the C_3 species *T. hassleriana* and *C. virdiflora* confidently branch together as members of the *Tarenaya* clade. For the β-*NAD-ME* gene tree, the topology of the *Polanisia* and *Cleome s. str.* clades is consistent with species relationships. Again, *Polanisia* and the *Cleome s. str.* cluster together as first-branching C_3 species (**Figure 4**). However, β-*NAD-ME* duplicated genes of *T. hassleriana*, *G. gynandra*, and *C. angustifolia* do not group as paralogs gene pairs or a cluster of orthologs. Instead a complex species-dependent arrangement of β*1* and β*2* genes is observed. In the C_3 *T. hassleriana*, the β*1* copy nests with its *C. virdiflora* ortholog, in line with the α-*NAD-ME* grouping, and *Th*β*2-NAD-ME* is found in a separate clade with *C. monophylla*. A plausible explanation is that the duplication of β-*NAD-ME* occurred in a common ancestor to these three C_3 related species and, after speciation, *C. virdiflora* lost the β*1* gene and *C. monophylla* lost the β*2* gene. In the C_4 species, the β*1* gene of *C. angustifolia* groups with the *C. paraxoda* ortholog, in concordance with the α-*NAD-ME* lineage, but the β*2* copy groups with β*2* of *G. gynandra*. Finally, β*1* of *G. gynandra* is placed basal to the *Gg*β*2-NAD-ME–Ca*β*2-NAD-ME* cluster (**Figure 4**). Hence, the β-*NAD-ME* duplication and subsequent loss of one copy seems to be a possible common phenomenon in Cleomaceae, which overshadows the true gene relationship. However, because of the essential role of NAD-ME in C4 photosynthesis, we are confident to propose that strong selective pressures accelerated the mutational rate of the α- and β-*NAD-ME* genes, as the species tree was globally recovered when nearly neutral markers are used.

Synteny Analysis of β-NAD-ME and α-NAD-ME Genes

The analysis of syntenic relationships of β-*NAD-ME* and α-*NAD-ME* genes in the genomes of Cleomaceae and Brassicaceae allowed us to clarify the evolutionary dynamics and duplication histories in these two families. We found that all retained copies of β-*NAD-ME* genes are syntenic across both families and with outgroup species (**Figure 5A**). Furthermore, it was evident that duplicate copies of β-*NAD-ME* in *Tarenaya, Gynandropsis*, and *Brassica* can be attributed to the known ancient polyploid histories of these species as duplicate copies of the genes were found in larger intra-genomic blocks. We also detected syntenic regions in all species derived from ancient polyploidy events, such as the At-alpha event at the origin the Brassicaceae, where a duplicated copy of β-*NAD-ME* was lost due to genome fractionation/gene loss processes that are known to have occurred.

We found a very different evolutionary pattern when examining the syntenic relationships for α-*NAD-ME* homologs. We found that α-*NAD-ME* genes are in a different genomic context in both families compared to all other angiosperms and in a specific and conserved context in Brassicaceae (**Figure 5B**) and in Cleomaceae (**Figure 5C**). The α-*NAD-ME* homologs are all syntenic with one another within the Brassicaceae, but not with Cleomaceae nor out-group species. Thus, we conclude that the ancestral α-*NAD-ME* gene transposed to a new genomic context in Brassicaceae after its divergence from the ancestor of Cleomaceae, potentially associated with the dramatic genome repatterning that would have occurred after the At-alpha WGD event. Whereas all three examined Cleomaceae species (*C. violacea*, *G. gynandra*, and *T. hassleriana*) α-*NAD-ME* orthologs are syntenic. *C. violacea* lacks evidence of the ancient polyploidy event shared between *G. gynandra* and *T. hassleriana*. Thus, the ancestor of all Cleomaceae α-*NAD-ME* must have transposed before the splitting of these two lineages (perhaps even earlier during the evolution of the Brassicales). The unique transpositions of α-*NAD-ME* to independent new genomic contexts may have had an impact of the expression and function of the gene in both Brassicaceae and Cleomaceae species compared to other eudicot species. As mentioned above, all Brassicaceae and Cleomaceae species retain only a single copy of α-*NAD-ME* and consistently we do not detect any retained and syntenic gene duplicates due to ancient genome duplications. Like with β-*NAD-ME*, we detected syntenic regions derived from ancient polyploidy events where a duplicated copy of α-*NAD-ME* was lost due to genome fractionation.

Amino Acid Substitutions in NAD-ME During the Evolution of C_4 Photosynthesis in Cleomaceae

We found a high number of non-synonymous substitutions in the β-NAD-ME sequences of the C_4 species (**Figure 3**). We further analyzed the patterns of amino acid positions in the β-NAD-ME sequences that accumulated changes during C_4 evolution in Cleomaceae compared to all Brassicales β-NAD-MEs in a MSA (**Supplementary Figure 1**). We found that

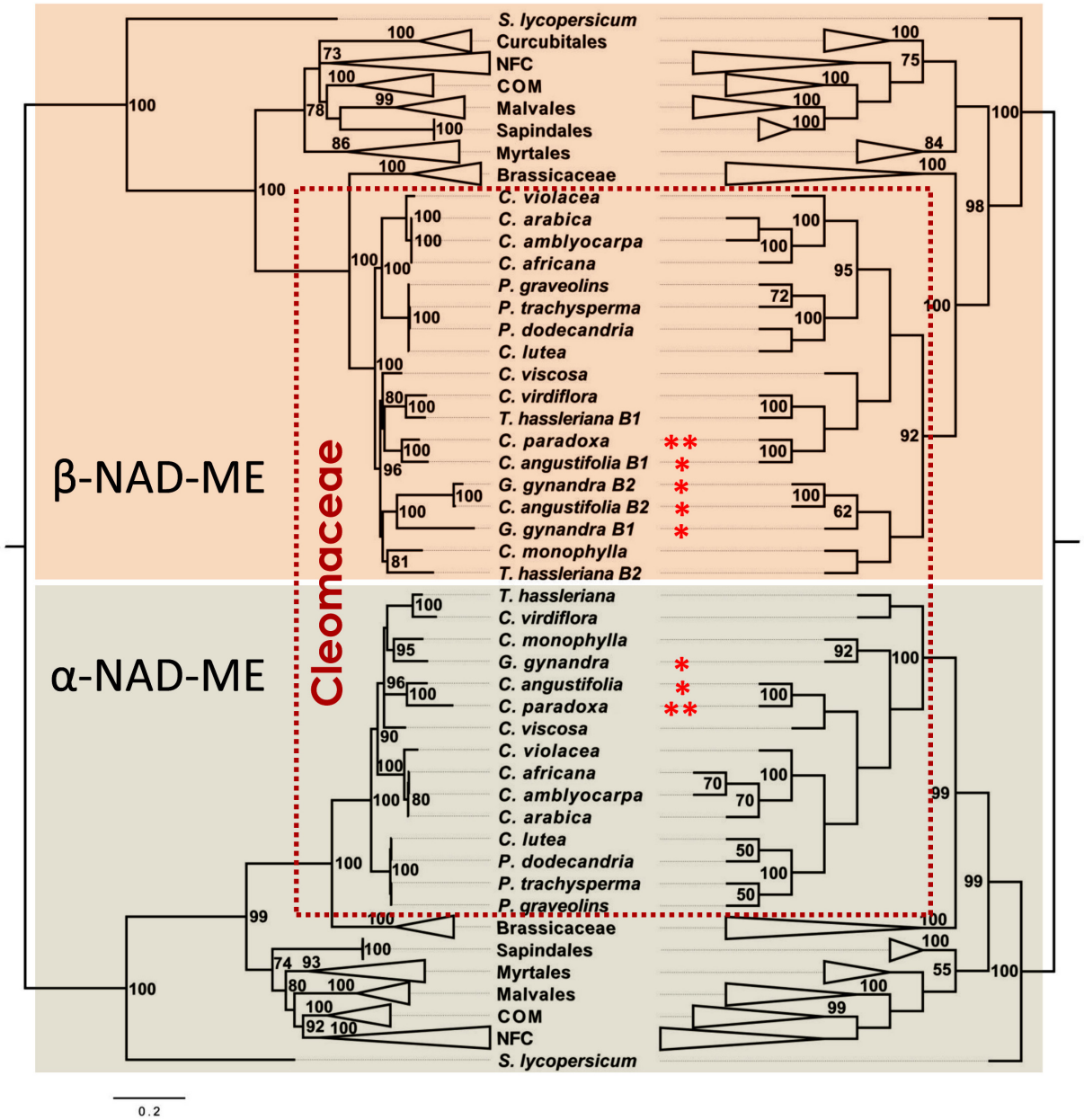

FIGURE 4 | Evolutionary history of *NAD-ME* genes in Cleomaceae deduced from third position of the codons. Bayesian **(left)** and Maximum Likelihood **(right)** phylogenetic tree of NAD-ME of the Rosid lineage. Most of the clades are compressed and designated by their Order or Family names. BPP and MLB values higher than 70 and 50%, respectively, are given next to the branches. In the BI analysis, the tree is drawn to scale with branch lengths measured in the number of substitutions per site. In the ML analysis, the bootstrap consensus tree (dendogram) inferred from 2,000 replicates is taken to represent the evolutionary history of the taxa, in which the partitions reproduced in less than 50% of the bootstrap replicates are collapsed. In both analyses, the best-fit substitution model was a GRT + G (3.70) model involving 124 nucleotide sequences and a total of 497 positions in the final dataset. ∗ and ∗∗ indicate C_4 and C_3–C_4 photosynthetic metabolism, respectively. *Solanum lycopersicum* NAD-MEs coding sequences were used as out groups. The full trees are available in **Supplementary Data 6** and **Supplementary Data 7**.

393 of 579 amino acids (∼68%) are identical in all mature β-NAD-ME protein sequences (**Supplementary Figure 1**). The most diverged protein sequences are the sequences of C_4 species *G. gynandra* β1-NAD-ME (186 amino acids, ∼32%) and *C. angustifolia* β1-NAD-ME (158 amino acids, ∼27%). Interestingly, β1-NAD-MEs of *G. gynandra* and *C. angustifolia* share many changed amino acids and both sequences differ only in 83 positions.

To identify amino acid residues potentially involved in C_4-optimization, we compared the NAD-MEs of the Brassicales in the MSA searching for differentially substituted amino acids in the NAD-ME protein sequences of the C_4 species. A differentially

FIGURE 5 | Syntenic relationships of β-*NAD-ME* and α-*NAD-ME* genes across Cleomaceae, Brassicaceae and outgroup genomes using SynFind and GEvo analysis. **(A)** A search for syntenic regions based on the β-*NAD-ME1* gene in *G. gynandra* (as reference region) identified syntenic homologs in all species (shown in red box), including the outgroup species *C. clementina*. Also, all duplicated copies found by phylogenetic analysis of β-*NAD-ME*, for example in *B. rapa* and *T. hassleriana*, were also syntenic which supports that they are derived from ancient polyploidy events. **(B)** A search for syntenic regions based on the *A. thaliana* (Brassicaceae) copy of α-NAD-ME identified only single copy syntenic homologs (shown within red box) in other Brassicaceae species (*A. alpina*, *E. salsugineum*, and *B. rapa*) but not with Cleomaceae species (*C. violacea*, *G. gynandra*, and *T. hassleriana*) nor the outgroup species *C. clementenia*. Similarly, **(C)** A search for syntenic homologs of the *C. violacea* α-*NAD-ME* gene (shown within red box) found them only in the other Cleomaceae species (*G. gynandra* and *T. hassleriana*) but not with Brassicaceae or the outgroup species. These combined results **(B, C)** suggest independent translocations of α-*NAD-ME* genes in the Brassicaceae and Cleomaceae to new genomic contexts.

substituted amino acid is a position in the MSA at which a NAD-ME protein of a C_4 species has an amino acid that differs from all other NAD-ME proteins, which all share the same amino acid (Alvarez et al., 2019).

We found three amino acid positions, F127, Q205, and N466 (numbered according to the full-length sequence of *G. gynandra*; **Supplementary Figure 2**), strictly differentially substituted in both *G. gynandra* and *C. angustifolia* α-NAD-ME proteins (**Supplementary Figure 2**, orange amino acids). Interestingly, we found that 36 amino acids are differentially substituted in the β1-NAD-ME from *G. gynandra* (**Supplementary Figure 1**, light blue + orange amino acids). Similarly, we identified 13 amino acids changes in the β1-NAD-ME of *C. angustifolia* (**Supplementary Figure 1**, dark blue + orange amino acids). Interestingly, 5 amino acid positions, V131, S132, H195, V297, and K605 (numbered according to the full-length sequence of *G. gynandra*) are identically substituted in both β1-NAD-ME of the C_4 species *G. gynandra* and *C. angustifolia* (**Supplementary Figure 1**, orange amino acids). This kind of amino acid replacement was not observed in the β2-NAD-ME isoforms of the C_4 species.

The strictly differentially substituted amino acids in the α-NAD-ME and β1-NAD-ME of *G. gynandra* and *C. angustifolia* suggest that these amino acids probably evolved under positive selection. To address this, we carried out model tests to identified residues that underwent adaptive changes in the α-NAD-ME and β1-NAD-MEs of the C_4 species. The models were optimized using the tree topology inferred from third positions of codons (**Figure 4**, right) and the mature α-NAD-ME and β-NAD-ME coding sequences of Cleomaceae in the MSA. For both paralogs, one model (M8), allowing a proportion of codons evolving under positive selection, provided better fit than the null model (M7) (**Supplementary Tables 2,3**). We found five amino acid positions having a posterior probability greater than 0.9 in the α-NAD-MEs: 80, 205, 423, 500, and 587 (numbered according to the full-length sequence of *G. gynandra*; **Supplementary Figure 2**). The position 205 corresponds to the differentially substituted Q205 in the α-NAD-ME of *G. gynandra* and *C. angustifolia* (**Supplementary Figure 2**). For the β1-NAD-MEs, three amino acid positions had a posterior probability greater than 0.9: 132, 297, and 395 (numbered according to the full-length sequence of *G. gynandra*, **Supplementary Figure 1**). The positively selected positions 132 and 297 correspond to the differentially substituted S132 and V297 residues in β1-NAD-ME of *G. gynandra* and *C. angustifolia*.

DISCUSSION

Sub-Functionalization Set Up the Heteromeric Assembly of NAD-ME and Neo-Functionalization Optimized the α- and β-Subunit Functions

We identified a duplication of *NAD-ME* in vascular plants that generated the two paralogs lineages: α- and β-*NAD-ME*. All seed plants examined maintained at least one α-*NAD-ME* and one β-*NADME* –homolog (**Figure 1** and **Supplementary Data 1**), with most species maintaining a 1:1 α/β relative gene

dosage (**Figure 2**). This is a strong indication that the α- and β-NAD-ME homologs diversified and evolved non-redundant, alternative and critical (housekeeping) functions. In higher plants the heteromeric assembly of NAD-ME is most probably a requirement to fully fit into the central carbon metabolism, and specific duplication of the NAD-ME subunit-coding genes are evolutionarily allowed.

The α-*NAD-ME* and β-*NAD-ME* gene lineages evolved by duplication of an ancestral *NAD-ME* gene that occurred late during the evolution of vascular plants, as the paralogs were not found in the early branching tracheophyte *S. moellendorffii* (**Figure 1**). The well supported positioning of the single *NAD-ME* genes of ferns as a sister group of spermatophytes α-NAD-ME (**Figure 1**, **Supplementary Data 2** and **Supplementary Data 3**) suggests two alternative scenarios for the duplication of *NAD-ME* (**Figure 6**): (A) an ancestral α-*NAD-ME-like* gene likely diverged in Tracheophytes after the Lycophytes separation. This ancestral gene duplicated at the origin of the seed plants giving rise to *pre*-α-*NAD-ME* and *pre*-β-*NAD-ME* paralogs sequences, which afterward diverged from the ancestral α-*NAD-ME-like* gene. Finally, the paralogs were fixed (possibly by genetic drifts) and adaptive selection preserved them in seed plants. In this scenario, a primitive α-*NAD-ME-like* gene was retained in lycophytes; alternatively, (B) a *NAD-ME* duplication giving rise to *pre*-α-*NAD-ME* and *pre*-β-*NAD-ME* paralogs took place in Tracheophytes directly after the Lycophytes separation. The pre-β-*NAD-ME* was then lost in the ferns before adaptive selection could preserve both genes (**Figure 6**). Instead, *pre*-α-*NAD-ME*

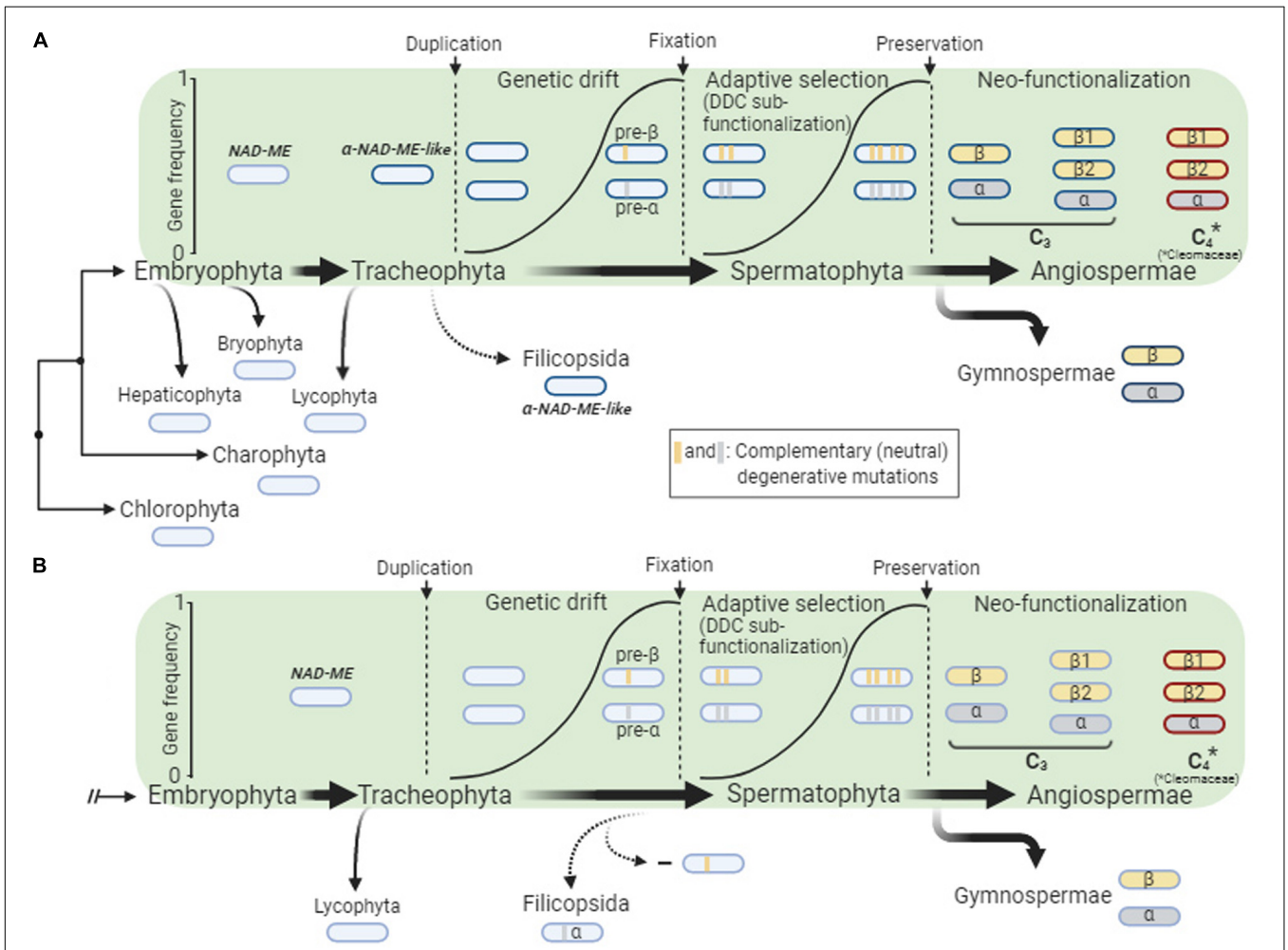

FIGURE 6 | Scheme showing our hypothesis for the evolution of the α-*NAD-ME* and β-*NAD-ME* genes in seed plants. The green panel shows the general course of the gene duplication event and the frequency of the both genes in the population until their fixation (genetic drift) and preservation (adaptive selection). The *NAD-ME* duplication took place late during the evolution of vascular plants, as the α- and β-*NAD-ME* paralogs were not found in Lycophytha, an early branching division of Tracheophyta. After that, one possible scenario is that one α-*NAD-ME-like* gene lineage was duplicated and fixed in seed plants but not in ferns, which kept a single copy **(A)**. Alternatively, the NAD-ME duplication could have occurred before Filicopsida and Spermatophyta split and the pre-β-*NAD-ME* copy was lost in ferns, as no selective pressure was exerted at that point **(B)**. In seed plants, adaptive selection through a duplication-degeneration-complementation (DDC) sub-functionalization process gave rise to the α- and β-*NAD-ME* gene lineages. Once preserved, both paralogs were double down and split, providing raw material for functional innovations.

and *pre*-β-*NAD-ME* diverged from each other and were preserved as the α- and β-*NAD-ME* paralogs in seed plants.

Four major models have been proposed for preservation of duplicated genes: neo-functionalization, specialization, dosage selection and duplication-degeneration-complementation (DDC) sub-functionalization (Conant and Wolfe, 2008; Innan and Kondrashov, 2010). Neo-functionalization and specialization provide an explanation for the cases where after fixation by genetic drift, one or both duplicate genes diverge from the ancestral function (e.g., in expression patterns, substrate specificities, etc.). The dosage selection model supposes that the new duplicated gene evolves by positive selection (e.g., the increase in the amount of protein is beneficial; the new copy emerges with a novel function or by shielding against deleterious mutations). The DDC sub-functionalization is a conservative evolutionary model that assumes that the functions of an ancestral gene have been neutrally divided among the daughter copies due to complementary degenerative mutations: neither copy is able to fulfill the original functions on their own.

Previous findings support the DDC sub-functionalization process by which the duplicated α- and β-*NAD-ME* copies could be preserved: (i) To fulfill the housekeeping function in L-malate respiration in seed plants, the NAD-ME is predominantly assembled as an heteromer of α- and β-subunits, with both catalyzing the same reaction (Tronconi et al., 2008, 2010b); (ii) both α-*NAD-ME* and β-*NAD-ME* genes are constitutively expressed and coordinately regulated and the encoded proteins accumulate at similar levels in the most plant tissues (Tronconi et al., 2008, 2010b; Fuchs et al., 2020); (iii) knockout lines of Arabidopsis lacking either α-NAD-ME or β-NAD-ME show residual NAD-ME activities, which in sum do not reach the activity measured in wildtype. Moreover, the mutants plants lacking β-NAD-ME retain less than 10% of the total NAD-ME activity (Tronconi et al., 2008; Brown et al., 2011); (iv) The identification of coevolved connected amino acidic residues belonging to the α-NAD-ME and β-NAD-ME subunits (Tronconi et al., 2018) indicates that compensatory neutral mutations marked the evolution toward a functional heteromeric NAD-ME in higher plants.

Following the preservation phase, neo-functionalization can occur in one or both duplicated genes (Conant and Wolfe, 2008). In this regard, the comprehensive characterization of *A. thaliana* NAD-ME indicated that the α-NAD-ME and β-NAD-ME homodimers and the α/β-NAD-ME heterodimer behave differently in terms of catalytic mechanism, interaction with the substrates and allosteric regulation (Tronconi et al., 2008, 2010a,b, 2015), pointing out an adaptive advantage in terms of metabolic flexibility. This metabolic flexibility might have played a role in the adaption of a C_4-specific version of the NAD-ME. Moreover, α- and β-NAD-ME differentially accumulate in the separate components of the floral organ (Tronconi et al., 2010b). In sepals, the α-NAD-ME is present at a slightly higher proportion than β-NAD-ME. On the other hand, β-NAD-ME is the only protein present in anthers. All these observations suggest that NAD-ME activity may be regulated by variations of the native association *in vivo*, rendering enzymatic entities with distinct allosteric regulation to fulfill additional metabolic roles (Maurino et al., 2009; Tronconi et al., 2010b; Maier et al., 2011).

Changes in the Subunit-Specific Duplications Provide the Basis for the Recruitment of NAD-ME in C_4 Biochemistry

In multiple families of the angiosperms we found a higher number of *NAD-ME* genes coding for β-NAD-ME (20 species) than for α-NAD-ME (one species) (**Figure 2**). This indicates that if small-scale gene duplications or WGDs occurred, the gene coding for a β-NAD-ME is tolerant of independent copy number variation (possibly fixed by genetic drift) while the α-*NAD-ME* gene is under evolutionary pressure to return to a 1:1 β/α relative status. It seems that the α-*NAD-ME* duplication imparts a detrimental effect or does not increase plant fitness and thus, the duplicated gene is not fixed in the population. A new copy can escape the negative effect of the dosage if during the fixation phase (**Figure 6**) a mutation conferring an adaptive advantage arises (Conant and Wolfe, 2008). This can explain the observation that the C_4-NAD-ME species *A. hypochondriacus* has retained two α-*NAD-ME* genes (**Figure 2**). Probably, shortly after the duplication, a copy quickly diverged from the original function by neo-functionalization to fit to the C_4 biochemistry.

The significantly higher proportion of C_4-NAD-ME species possessing an additional *NAD-ME* gene copy (**Figure 2**) is consistent with the general notion that gene duplication is a precondition for the evolution of the C_4 functions (Lynch and Conery, 2000; Moore and Purugganan, 2005). For other C_4 genes, one copy of a duplicated gene is neo-functionalized without affecting the other members of the genetic family (Christin et al., 2013). However, in the C_4 species of Cleomaceae all three genes were potentially affected by adaptive selection as suggested by the inconsistencies observed in the protein-based tree topology (**Figure 3**). This could be rectified for the α-*NAD-ME* genes, albeit partially for the β-*NAD-ME* genes, when a nearly neutral marker (third codon positions) was used (**Figure 4**). Higher amino acid substitution rates are associated with an accelerated evolution or potential positive selection due to neo-functionalization of the resulting protein sequence. Neo-functionalization from a duplicated gene is a classic driver of protein evolution.

We detected a high proportion of base substitutions in the Cleomaceae C_4 species β-*NAD-ME* genes. The β-*NAD-ME* genes are duplicated in the Cleomaceae C_4 species *C. angustifolia* and *G. gynandra* and the C_3 species *T. hassleriana*, and in the Brassicaceae *B. rapa* and *B. oleracea*. Synteny analysis clearly showed that duplication of β-*NAD-ME* in both families (**Figure 5A**) was due to the ancient polyploidy events, such as the *Brassica* lineage genome triplication and the *Tarenaya* WGD that is shared with *G. gynandra* (Cheng et al., 2014; van den Bergh et al., 2014). Importantly, changes in copy number of genes can maintain dosage-balance relationships if generated by polyploidy events. Nevertheless, only the β1-NAD-ME of both C_4 species has accumulated a high number of amino acid changes. Intriguingly, five positions show the same amino acid changes in both β1-NAD-ME of these C_4 (**Supplementary Figure 1**),

two of which (S132 and V295) evolved under positive selection (**Supplementary Table 3**). We hypothesize that these amino acids play a role in the C_4 function of NAD-ME, as *C. angustifolia* and *G. gynandra* belong to different C_4 lineages and the five amino acids are differentially conserved in the β-NAD-ME of C_3 Cleome species, Brassicales β-NAD-ME or the respective β2-NAD-ME of *C. angustifolia* and *G. gynandra*. Interesting, four of these five amino acid are involved in substrate coordination or are located in the L-malate binding domain (Chang and Tong, 2003). Finally, we found three differentially conserved amino acid substitutions in the α-NAD-ME of the C_4 species (**Supplementary Figure 2**), one of them (Q205) identified as an adaptive change (**Supplementary Table 2**). This adaptive and differentially conserved position in the α-NAD-ME is neither part of the enzymatic active center nor was shown to participate in the enzymatic mechanism of reaction (Chang and Tong, 2003). Because of the heteromeric NAD-ME assembly, this substitution most likely represents a change necessary to compensate for the evolutionary changes in the β1-NAD-ME, or enable another kind of function that is necessary in the adaption of NAD-ME in the C_4 context.

CONCLUSION

It appears that the genes encoding C_4 enzymes evolved by simply duplication of an original metabolic enzyme and further neo-functionalization (Christin et al., 2013; Ludwig, 2016; Alvarez et al., 2019). NAD-ME turns out to be an exceptional case, probably due to its heteromeric structure. NAD-ME followed an intricated molecular mechanism of evolution marked by sub-functionalization and differences in the frequency of α- and β-*NAD-ME* gene duplication (**Figure 6**). Future work should focus on how NAD-ME in C4 plant mitochondria has been adapted to perform both housekeeping and C_4-associated functions.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

VM conceived the project in active discussion with all co-authors. MT and MH performed the phylogenetic analysis. MS performed the syntenic analysis. All authors contributed to data analysis and writing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft under grant MA2379/18-1 to VM and the National Agency for Scientific and Technological Promotion PICT2015-1064 to MT.

ACKNOWLEDGMENTS

We thank J. Chris Pires and Makenzie Mabry for access to Cleomaceae transcriptome data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.572080/full#supplementary-material

REFERENCES

Alvarez, C. E., Bovdilova, A., Hoppner, A., Wolff, C. C., Saigo, M., Trajtenberg, F., et al. (2019). Molecular adaptations of NADP-malic enzyme for its function in C4 photosynthesis in grasses. *Nat. Plants* 5, 755–765. doi: 10.1038/s41477-019-0451-7

Anderson, L. E. (1971). Chloroplast and cytoplasmic enzymes. II. Pea leaf triose phosphate isomerases. *Biochim. Biophys. Acta* 235, 237–244. doi: 10.1016/0005-2744(71)90051-9

Artus, N. N., and Edwards, G. E. (1985). NAD-malic enzyme from plants. *FEBS Lett.* 182, 225–233. doi: 10.1016/0014-5793(85)80305-7

Barker, M. S., Vogel, H., and Schranz, M. E. (2009). Paleopolyploidy in the brassicales: analyses of the cleome transcriptome elucidate the history of genome duplications in *Arabidopsis* and other brassicales. *Genome Biol. Evol.* 1, 391–399. doi: 10.1093/gbe/evp040

Bassham, J. A., Benson, A. A., Kay, L. D., Harris, A. Z., Wilson, A. T., and Calvin, M. (1954). The path of carbon in photosynthesis. XXI. The cyclic regeneration of carbon dioxide acceptor. *J. Am. Chem. Soc.* 76, 1760–1770. doi: 10.1021/ja01636a012

Bayat, S., Schranz, M. E., Roalson, E. H., and Hall, J. C. (2018). Lessons from Cleomaceae, the sister of crucifers. *Trend. Plant Sci.* 23, 808–821. doi: 10.1016/j.tplants.2018.06.010

Brown, N. J., Newell, C. A., Stanley, S., Chen, J. E., Perrin, A. J., Kajala, K., et al. (2011). Independent and parallel recruitment of preexisting mechanisms underlying C-4 photosynthesis. *Science* 331, 1436–1439. doi: 10.1126/Science.1201248

Chang, G. G., and Tong, L. (2003). Structure and function of malic enzymes, a new class of oxidative decarboxylases. *Biochemistry* 42, 12721–12733. doi: 10.1021/bi035251+

Cheng, F., Wu, J., and Wang, X. (2014). Genome triplication drove the diversification of *Brassica* plants. *Hortic. Res.* 1:14024. doi: 10.1038/hortres.2014.24

Christin, P. A., Besnard, G., Samaritani, E., Duvall, M. R., Hodkinson, T. R., Savolainen, V., et al. (2008). Oligocene CO2 decline promoted C4 photosynthesis in grasses. *Curr. Biol.* 18, 37–43. doi: 10.1016/j.cub.2007.11.058

Christin, P. A., Boxall, S. F., Gregory, R., Edwards, E. J., Hartwell, J., and Osborne, C. P. (2013). Parallel recruitment of multiple genes into c4 photosynthesis. *Genome Biol. Evol.* 5, 2174–2187. doi: 10.1093/gbe/evt168

Christin, P. A., Salamin, N., Savolainen, V., Duvall, M. R., and Besnard, G. (2007). C4 Photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr. Biol.* 17, 1241–1247. doi: 10.1016/j.cub.2007.06.036

Christin, P. A., Samaritani, E., Petitpierre, B., Salamin, N., and Besnard, G. (2009). Evolutionary insights on C4 photosynthetic subtypes in grasses from genomics and phylogenetics. *Genome Biol. Evol.* 1, 221–230. doi: 10.1093/gbe/evp020

Conant, G. C., and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950. doi: 10.1038/nrg2482

Detarsio, E., Alvarez, C. E., Saigo, M., Andreo, C. S., and Drincovich, M. F. (2007). Identification of domains involved in tetramerization and malate inhibition of maize C4-NADP-malic enzyme. *J. Biol. Chem.* 282, 6053–6060. doi: 10.1074/jbc.M609436200

Detarsio, E., Maurino, V. G., Alvarez, C. E., Muller, G. L., Andreo, C. S., and Drincovich, M. F. (2008). Maize cytosolic NADP-malic enzyme (ZmCytNADP-ME): a phylogenetically distant isoform specifically expressed in embryo and emerging roots. *Plant Mol. Biol.* 68, 355–367. doi: 10.1007/s11103-008-9375-8

Drincovich, M. F., Lara, M. V., Andreo, C. S., and Maurino, V. G. (2011). "Evolution of C4 decarboxylases: different solutions for the same biochemical problem: provision of CO2 in bundle sheath cells," in *C4 Photosynthesis and Related CO2 Concentration Mechanisms; Advances in Photosynthesis and Respiration*, ed. S. Govindjee (Berlin: Springer), 277–300. doi: 10.1007/978-90-481-9407-0_14

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Glockner, G., Platts, A. E., et al. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8362–8366. doi: 10.1073/pnas.1503926112

Edwards, E. J., Osborne, C. P., Strömberg, C. A. E., and Smith, S. A. (2010). The origins of C4 grasslands: integrating evolutionary and ecosystem science. *Science* 328, 587–591. doi: 10.1126/science.1177216

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016. doi: 10.1006/jmbi.2000.3903

Feodorova, T. A., Voznesenskaya, E. V., Edwards, G. E., and Roalson, E. H. (2010). Biogeographic patterns of diversification and the origins of C-4 in cleome (Cleomaceae). *Syst. Bot.* 35, 811–826. doi: 10.1600/036364410x539880

Fuchs, P., Rugen, N., Carrie, C., Elsasser, M., Finkemeier, I., Giese, J., et al. (2020). Single organelle function and organization as estimated from *Arabidopsis* mitochondrial proteomics. *Plant J.* 101, 420–441. doi: 10.1111/tpj.14534

Furbank, R. T., and Hatch, M. D. (1987). Mechanism of c(4) photosynthesis: the size and composition of the inorganic carbon pool in bundle sheath cells. *Plant Physiol.* 85, 958–964. doi: 10.1104/pp.85.4.958

Gerrard Wheeler, M. C., Arias, C. L., Tronconi, M. A., Maurino, V. G., Andreo, C. S., and Drincovitch, M. F. (2008). *Arabidopsis* thaliana NADP-malic enzyme isoforms: high degree of identity but clearly distinct properties. *Plant Mol. Biol.* 67, 231–242. doi: 10.1007/s11103-008-9313-9

Gerrard Wheeler, M. C., Tronconi, M. A., Drincovich, M. F., Andreo, C. S., Flugge, U. I., and Maurino, V. G. (2005). A comprehensive analysis of the NADP-malic enzyme gene family of *Arabidopsis*. *Plant Physiol.* 139, 39–51. doi: 10.1104/pp.105.065953

González-Moro, B., Lacuesta, M., Becerril, J. M., Gonzálelz-Murua, C., and Muñoz-Rueda, A. (1997). Glycolate accumulation causes a decrease of photosynthesis by inhibiting RUBISCO activity in maize. *J. Plant Physiol.* 150, 388–394. doi: 10.1016/s0176-1617(97)80087-9

Grover, S. D., Canellas, P. F., and Wedding, R. T. (1981). Purification of NAD malic enzyme from potato and investigation of some physical and kinetic properties. *Arch. Biochem. Biophys.* 209, 396–407. doi: 10.1016/0003-9861(81)90297-6

Hatch, M. D. (1971). "Mechanism and function of C4 photosynthesis," in *Photosynthesis and Photorespiration*, eds M. D. Hatch, C. B. Osmond, and R. O. Slatyer (New York: Wiley-Interscience), 139–152.

Hatch, M. D. (1987). C4 photosynthesis: a unique blend of midified biochemistry, anatomy and ultrastructure. *Biochim. Biophys. Acta* 895, 81–106. doi: 10.1016/s0304-4173(87)80009-5

Hayes, J. M. (1994). "Global methanothophy at the Archean-Proterozoic transition," in *Early life on Earth*, ed. S. Bengston (New York, NY: Columbia University Press), 220–236.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P. M., et al. (2013). Predicting C-4 photosynthesis evolution: modular, individually adaptive steps on a mount fuji fitness landscape. *Cell* 153, 1579–1588. doi: 10.1016/J.Cell.2013.04.058

Hillis, D. M., and Bull, J. J. (1993). An empirical-test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42, 182–192. doi: 10.2307/2992540

Hilu, K. W., Black, C. M., and Oza, D. (2014). Impact of gene molecular evolution on phylogenetic reconstruction: a case study in the rosids (Superorder Rosanae, Angiosperms). *PLoS one* 9:e99725. doi: 10.1371/journal.pone.0099725

Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11, 97–108. doi: 10.1038/nrg2689

Kanai, R., and Edwards, E. J. (1999). "The biochemistry of C4 photosynthesis," in *C4 Plant Biology*, eds R. F. Sage and R. K. Monson (San Diego, CA: Academic Press), 49–87. doi: 10.1016/b978-012614440-6/50004-5

Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33, 511–518. doi: 10.1093/nar/gki198

Kelly, G. J., and Latzko, E. (1976). Inhibition of spinach-leaf phosphofructokinase by 2-phosphoglycollate. *FEBS Lett.* 68, 55–58. doi: 10.1016/0014-5793(76)80403-6

Koteyeva, N. K., Voznesenskaya, E. V., Roalson, E. H., and Edwards, G. E. (2011). Diversity in forms of C4 in the genus Cleome (Cleomaceae). *Ann. Bot.* 107, 269–283. doi: 10.1093/aob/mcq239

Kulahoglu, C., Denton, A. K., Sommer, M., Mass, J., Schliesky, S., Wrobel, T. J., et al. (2014). Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C3 and C4 plant species. *Plant Cell* 26, 3243–3260. doi: 10.1105/tpc.114.123752

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Kuraku, S., Zmasek, C. M., Nishimura, O., and Katoh, K. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res.* 41, W22–W28. doi: 10.1093/nar/gkt389

Leakey, A. D., and Lau, J. A. (2012). Evolutionary context for understanding and manipulating plant responses to past, present and future atmospheric [CO2]. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 613–629. doi: 10.1098/rstb.2011.0248

Li, F. W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P. M., et al. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat. Plants* 4, 460–472. doi: 10.1038/s41477-018-0188-8

Long, J. J., Wang, J. L., and Berry, J. O. (1994). Cloning and analysis of the C4 photosynthetic NAD-dependent malic enzyme of amaranth mitochondria. *J. Biol. Chem.* 269, 2827–2833.

Ludwig, M. (2016). Evolution of carbonic anhydrase in C4 plants. *Curr. Opin. Plant Biol.* 31, 16–22. doi: 10.1016/j.pbi.2016.03.003

Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151

Mabry, M. E., Brose, J. M., Blischak, P. D., Sutherland, B., Dismukes, W. T., Bottoms, C. A., et al. (2019). Phylogeny and multiple independent whole-genome duplication events in the brassicales. *bioRxiv* [Preprint]. doi: 10.1101/789040

Maier, A., Zell, M. B., and Maurino, V. G. (2011). Malate decarboxylases: evolution and roles of NAD(P)-ME isoforms in species performing C(4) and C(3) photosynthesis. *J. Exp. Bot.* 62, 3061–3069. doi: 10.1093/jxb/err024

Maurino, V. G., and Peterhansel, C. (2010). Photorespiration: current status and approaches for metabolic engineering. *Curr. Opin. Plant Biol.* 13, 249–256. doi: 10.1016/j.pbi.2010.01.006

Maurino, V. G., Saigo, M., Andreo, C. S., and Drincovich, M. F. (2001). Non-photosynthetic 'malic enzyme' from maize: a constitutively expressed enzyme that responds to plant defence inducers. *Plant Mol. Biol.* 45, 409–420. doi: 10.1023/A:1010665910095

Maurino, V. G., Wheeler, M. C. G., Andreo, C. S., and Drincovich, M. F. (2009). Redundancy is sometimes seen only by the uncritical: does *Arabidopsis* need six malic enzyme isoforms? *Plant Sci.* 176, 715–721. doi: 10.1016/J.Plantsci.2009.02.012

Moore, R. C., and Purugganan, M. D. (2005). The evolutionary dynamics of plant duplicate genes. *Curr. Opin. Plant Biol.* 8, 122–128. doi: 10.1016/j.pbi.2004.12.001

Murata, T., Ohsugi, R., Matsuoka, M., and Nakamoto, H. (1989). Purification and Characterization of NAD Malic Enzyme from Leaves of Eleusine coracana and *Panicum dichotomiflorum*. *Plant Physiol.* 89, 316–324. doi: 10.1104/pp.89.1.316

Nikolov, L. A., and Tsiantis, M. (2017). Using mustard genomes to explore the genetic basis of evolutionary change. *Curr. Opin. Plant Biol.* 36, 119–128. doi: 10.1016/j.pbi.2017.02.005

Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K. K., et al. (2018). The chara genome: secondary complexity and implications for plant terrestrialization. *Cell* 174, 448.e24–464.e24. doi: 10.1016/j.cell.2018.06.033

Patchell, M. J., Roalson, E. H., and Hall, C. H. (2014). Resolved phylogeny of Cleomaceae based on all three genomes. *Taxon* 63, 315–328. doi: 10.12705/632.17

Proost, S., Van Bel, M., Vaneechoutte, D., Van de Peer, Y., Inze, D., Mueller-Roeber, B., et al. (2015). PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Res.* 43, D974–D981. doi: 10.1093/nar/gku986

Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180

Sage, R. F., Christin, P. A., and Edwards, E. J. (2011). The C(4) plant lineages of planet Earth. *J. Exp. Bot.* 62, 3155–3169. doi: 10.1093/jxb/err048

Sage, R. F., and Monson, R. K. (1999). *C4 Plant Biology*. San Diego: Academic.

Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and the evolution of C4 photosynthesis. *Annu. Rev. Plant Biol.* 63, 19–47. doi: 10.1146/annurev-arplant-042811-105511

Saigo, M., Bologna, F. P., Maurino, V. G., Detarsio, E., Andreo, C. S., and Drincovich, M. F. (2004). Maize recombinant non-C4 NADP-malic enzyme: a novel dimeric malic enzyme with high specific activity. *Plant Mol. Biol.* 55, 97–107. doi: 10.1007/s11103-004-0472-z

Scheffler, K., and Seoighe, C. (2005). A Bayesian model comparison approach to inferring positive selection. *Mol. Biol. Evol.* 22, 2531–2540. doi: 10.1093/molbev/msi250

Schranz, M. E., and Mitchell-Olds, T. (2006). Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell* 18, 1152–1165. doi: 10.1105/tpc.106.041111

Shen, H., Jin, D., Shu, J. P., Zhou, X. L., Lei, M., Wei, R., et al. (2018). Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns. *Gigascience* 7, 1–11. doi: 10.1093/gigascience/gix116

Tang, H., Bomhoff, M. D., Briones, E., Zhang, L., Schnable, J. C., and Lyons, E. (2015). SynFind: compiling syntenic regions across any set of genomes on demand. *Genome Biol. Evol.* 7, 3286–3298. doi: 10.1093/gbe/evv219

Tausta, S. L., Coyle, H. M., Rothermel, B., Stiefel, V., and Nelson, T. (2002). Maize C4 and non-C4 NADP-dependent malic enzymes are encoded by distinct genes derived from a plastid-localized ancestor. *Plant Mol. Biol.* 50, 635–652.

Tronconi, M. A., Andreo, C. S., and Drincovich, M. F. (2018). Chimeric structure of plant malic enzyme family: different evolutionary scenarios for NAD- and NADP-dependent isoforms. *Front. Plant Sci.* 9:565. doi: 10.3389/fpls.2018.00565

Tronconi, M. A., Fahnenstich, H., Gerrard Weehler, M. C., Andreo, C. S., Flugge, U. I., Drincovich, M. F., et al. (2008). *Arabidopsis* NAD-malic enzyme functions as a homodimer and heterodimer and has a major impact on nocturnal metabolism. *Plant Physiol.* 146, 1540–1552. doi: 10.1104/pp.107.114975

Tronconi, M. A., Gerrard Wheeler, M. C., Maurino, V. G., Drincovich, M. F., and Andreo, C. S. (2010a). NAD-malic enzymes of *Arabidopsis* thaliana display distinct kinetic mechanisms that support differences in physiological control. *Biochem. J.* 430, 295–303. doi: 10.1042/BJ20100497

Tronconi, M. A., Maurino, V. G., Andreo, C. S., and Drincovich, M. F. (2010b). Three different and tissue-specific NAD-malic enzymes generated by alternative subunit association in *Arabidopsis thaliana*. *J. Biol. Chem.* 285, 11870–11879. doi: 10.1074/jbc.M109.097477

Tronconi, M. A., Wheeler, M. C., Martinatto, A., Zubimendi, J. P., Andreo, C. S., and Drincovich, M. F. (2015). Allosteric substrate inhibition of *Arabidopsis* NAD-dependent malic enzyme 1 is released by fumarate. *Phytochemistry* 111, 37–47. doi: 10.1016/j.phytochem.2014.11.009

van den Bergh, E., Külahoglu, C., Bräutigam, A., Hibberd, J. M., Weber, A. P. M., Zhu, X.-G., et al. (2014). Gene and genome duplications and the origin of C4 photosynthesis: birth of a trait in the Cleomaceae. *Curr. Plant Biol.* 1, 2–9. doi: 10.1016/j.cpb.2014.08.001

Vicentini, F. C., Denes, F. T., Borges, L. L., Silva, F. A., Machado, M. G., and Srougi, M. (2008). Laparoscopic pyeloplasty in children: is the outcome different in children under 2 years of age? *J. Pediatr. Urol.* 4, 348–351. doi: 10.1016/j.jpurol.2008.03.001

von Caemmerer, S. (2000). *Biochemical Models of Leaf Photosynthesis*. Collingwood: Csiro Publishing.

Wang, Y., Brautigam, A., Weber, A. P., and Zhu, X. G. (2014). Three distinct biochemical subtypes of C4 photosynthesis? A modelling analysis. *J. Exp. Bot.* 65, 3567–3578. doi: 10.1093/jxb/eru058

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A. M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449.

Zou, C. S., Li, L. T., Miki, D., Li, D. L., Tang, Q. M., Xiao, L. H., et al. (2019). The genome of broomcorn millet. *Nat. Commun.* 10:436. doi: 10.1038/s41467-019-08409-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 15 October 2020
doi: 10.3389/fpls.2020.564463

A Partial C_4 Photosynthetic Biochemical Pathway in Rice

HsiangChun Lin[1], Stéphanie Arrivault[2†], Robert A. Coe[1], Shanta Karki[3], Sarah Covshoff[4], Efren Bagunu[1], John E. Lunn[2†], Mark Stitt[2†], Robert T. Furbank[5], Julian M. Hibberd[4] and William Paul Quick[1,6]*

[1] C_4 Rice Centre, International Rice Research Institute (IRRI), Los Baños, Philippines, [2] Max Planck Institute of Molecular Plant Physiology (MPI-MP), Potsdam, Germany, [3] National Centre for Fruit Development, Kirtipur, Nepal, [4] Department of Plant Sciences, University of Cambridge, Cambridge, United Kingdom, [5] ARC Centre of Excellence for Translational Photosynthesis, Research School of Biology, The Australian National University, Acton, ACT, Australia, [6] Department of Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom

OPEN ACCESS

Edited by:
Veronica Graciela Maurino,
University of Bonn, Germany

Reviewed by:
Sascha Offermann,
Leibniz University Hannover, Germany
Tsuyoshi Furumoto,
Ryukoku University, Japan

***Correspondence:**
William Paul Quick
w.p.quick@irri.org

†ORCID:
Mark Stitt
orcid.org/0000-0002-4900-1763
Stéphanie Arrivault
orcid.org/0000-0003-0516-6950
John E. Lunn
orcid.org/0000-0001-8533-3004

Specialty section:
This article was submitted to Plant Physiology, a section of the journal Frontiers in Plant Science

Received: 24 June 2020
Accepted: 25 September 2020
Published: 15 October 2020

Citation:
Lin H, Arrivault S, Coe RA, Karki S, Covshoff S, Bagunu E, Lunn JE, Stitt M, Furbank RT, Hibberd JM and Quick WP (2020) A Partial C_4 Photosynthetic Biochemical Pathway in Rice. Front. Plant Sci. 11:564463. doi: 10.3389/fpls.2020.564463

Introduction of a C_4 photosynthetic pathway into C_3 rice (*Oryza sativa*) requires installation of a biochemical pump that concentrates CO_2 at the site of carboxylation in modified bundle sheath cells. To investigate the feasibility of this, we generated a quadruple line that simultaneously accumulates four of the core C_4 photosynthetic enzymes from the NADP-malic enzyme subtype, phospho*enol*pyruvate carboxylase (*Zm*PEPC), NADP-malate dehydrogenase (*Zm*NADP-MDH), NADP-malic enzyme (*Zm*NADP-ME), and pyruvate phosphate dikinase (*Zm*PPDK). This led to enhanced enzyme activity and mild phenotypic perturbations but was largely neutral in its effects on photosynthetic rate. Measurements of the flux of $^{13}CO_2$ through photosynthetic metabolism revealed a significant increase in the incorporation of ^{13}C into malate, consistent with increased fixation of $^{13}CO_2$ via PEP carboxylase in lines expressing the maize PEPC enzyme. However, there was no significant differences in labeling of 3-phosphoglycerate (3PGA) indicating that there was no carbon flux through NADP-ME into the Calvin-Benson cycle. There was also no significant difference in labeling of phospho*enol*pyruvate (PEP) indicating that there was no carbon flux through PPDK. Crossing the quadruple line with a line with reduced glycine decarboxylase H-protein (*Os*GDCH) abundance led to a photosynthetic phenotype characteristic of the reduced *Os*GDCH line and higher labeling of malate, aspartate and citrate than in the quintuple line. There was evidence of ^{13}C labeling of aspartate indicating $^{13}CO_2$ fixation into oxaloacetate by PEPC and conversion to aspartate by the endogenous aspartate aminotransferase activity. While Kranz anatomy or other anatomical modifications have not yet been installed in these plants to enable a fully functional C_4 cycle, these results demonstrate for the first-time a partial flux through the carboxylation phase of NADP-ME C_4 metabolism in transgenic rice containing two of the key metabolic steps in the C_4 pathway.

Keywords: C_4 rice, C_4 photosynthesis, ^{13}C labeling, NADP-malic enzyme, malate, *Oryza sativa* (rice), transgenic rice, metabolic engineering

INTRODUCTION

A major recent research objective has been the engineering of a C_4 photosynthetic pathway into rice [1](Kajala et al., 2011; von Caemmerer et al., 2012; Ermakova et al., 2019), potentially leading to an increase in radiation use efficiency and yield of up to 50% (Hibberd et al., 2008). The C_4 pathway represents a complex combination of both biochemical and anatomical adaptations that suppresses photorespiration by effectively saturating ribulose bisphosphate carboxylase/oxygenase (Rubisco) with CO_2. In the majority of C_4 plants, this is achieved by compartmentalization of photosynthetic reactions between two morphologically distinct cell types: the mesophyll cells (MCs) and the bundle sheath cells (BSCs). Operating across these cells is a biochemical CO_2 pump elevating the CO_2 concentration in the BSCs where Rubisco is located (Hatch, 1987).

There are three primary variants of this pump characterized by the main decarboxylase reaction (Hatch et al., 1975). The NADP-ME subtype was chosen for engineering C_4 photosynthesis into rice as it is well-characterized in the C_4 model crop species maize (*Zea mays*) and potentially requires the fewest biochemical enzymes among all C_4 subtypes (Weber and von Caemmerer, 2010; Kajala et al., 2011; Ermakova et al., 2019). Each molecule of CO_2 entering the cytosol of the MCs is first converted to bicarbonate (HCO_3^-) by the activity of carbonic anhydrase (CA) and then incorporated into phospho*enol*pyruvate (PEP) by PEP carboxylase (PEPC), yielding the C_4 acid oxaloacetate (OAA). OAA is taken up into the chloroplast of the MCs where it is reduced to malate by the NADP-dependent malate dehydrogenase (NADP-MDH). Malate is exported back to the cytosol and then diffuses into BSCs through plasmodesmata along a steep concentration gradient. In the BSCs, malate is transported into the chloroplast by an unknown transporter and oxidatively decarboxylated by NADP-dependent malic enzyme (NADP-ME), yielding CO_2, NADPH, and pyruvate. CO_2 is assimilated by Rubisco, yielding two molecules of 3PGA, about half of which is reduced to triose-phosphate (TPs) using the NADPH provided by NADP-ME in the BSC chloroplast to regenerate RuBP in the Calvin-Benson cycle. The other half of the 3PGA moves to the MCs for reduction to TP in the MC chloroplast and then returns to the BSCs to enter the Calvin-Benson cycle. Pyruvate moves from the BSCs into the chloroplasts of the MCs where it is converted to PEP by pyruvate:phosphate dikinase (PPDK).

Previous attempts to introduce a single-cell C_4 pathway into rice led to increased photoinhibition of photosynthesis, leaf chlorophyll bleaching and serious stunting with no evidence of CO_2 concentration in chloroplasts (Taniguchi et al., 2008; Miyao et al., 2011). This work highlighted the need to achieve the correct activity, regulation, kinetic properties, and location of the enzymes. To address this, in this study, we report on the introduction of part of two-celled biochemical pathway into rice. It has previously been shown that genomic sequences encoding C_4 proteins give stronger expression in rice than cDNAs (Matsuoka et al., 1994). Therefore, we decided to express individual full-length genes of *Zm*PEPC, *Zm*PPDK, *Zm*NADP-MDH, and *Zm*NADP-ME (including promoters, untranslated regions, exons, and introns) from maize (Kajala et al., 2011; Karki et al., 2020) in a bid to achieve a C_4-like pattern of C_4 gene expression, enzyme localization, enzyme activity, and enzyme kinetic properties (Matsuoka et al., 1994; Miyao, 2003; Hibberd and Covshoff, 2010; Miyao et al., 2011). Individual lines were then crossed to generate a plant overexpressing all four of these core C_4 cycle enzymes to investigate the feasibility of installing a functional C_4 biochemical pathway into rice. This quadruple transgenic line was also crossed with a line with decreased *Os*GDCH protein (Lin et al., 2016). We investigated the effect on plant growth and photosynthesis.

To evaluate photosynthetic functionality, we used $^{13}CO_2$ labeling experiments (Arrivault et al., 2017), similar in concept to the radiolabeling experiments originally performed to characterize flux in C_4 photosynthesis (Hatch et al., 1967; Hatch, 1971). Flux of $^{13}CO_2$ through photosynthetic metabolism, in particular into C_4 acids, was determined for the quadruple and quintuple lines, compared to untransformed controls. We show that there was increased labeling of C_4 acids in both sets of plants compared to wild type, consistent with partial low-level function of a portion of the C_4 pathway.

MATERIALS AND METHODS

Plant Materials

Individual transgenic lines were generated overexpressing four of the core C_4 cycle enzymes required for a functional NADP-ME C_4 cycle (Kajala et al., 2011; **Supplementary Figure 1**), *Zm*PEPC (GRMZM2G083841), *Zm*PPDK (GRMZM2G306345), *Zm*NADP-MDH (GRMZM2G129513), and *Zm*NADP-ME (GRMZM2G085019. Generations of pSC0/*Zm*PEPC, pSC0/*Zm*PPDK, pSC0/ZmNADP-MDH, and pSC0/NADP-ME vectors were previously described (Giuliani et al., 2019b; Karki et al., 2020). In almost all cases, three independent single insertion homozygous transgenic lines with high transgene expression were selected for molecular and biochemical evaluation. However, for *Zm*NADP-ME, protein expression was only detected in a single transgenic line containing >6 copies of the overexpression construct and so this was the only line that could be taken forward (Karki et al., 2020). The overexpression constructs were stacked into single lines through conventional crossing to create two triple cross line (PEPC-28/PPDK-11/MDH-40 and PEPC-62/PPDK-2/MDH-22) each with an independent transgenic event for each gene. The crossing strategy was presented in **Supplementary Figure 2**. The *Zm*PEPC (PEPC-28 and PEPC-62) and *Zm*NADP-MDH (MDH-40 and MDH-22) single transgenic lines were initially crossed to create two double transgene lines (PEPC-28/MDH-40 and PEPC-62/MDH-22) that were then crossed with the *Zm*PPDK (PPDK-11 and PPDK-2) single transgenic lines. These two triple lines (PEPC-28/PPDK-11/MDH-40 and PEPC-62/PPDK-2/MDH-22) were then crossed with the single *Zm*NADP-ME (ME-116) transgenic line to produce two quadruple lines (PEPC-28/PPDK-11/MDH-40/ME-116 and

[1]https://C4rice.com

PEPC-62/PPDK-2/MDH-22/ME-116). Progeny of the PEPC-28/PPDK-11/MDH-40 showed that detectable *Zm*PEPC localizes to MCs, but progeny of the PEPC-62/PPDK-2/MDH-22 showed that *Zm*PEPC localizes to both MCs and BSCs (**Supplementary Figure 3**). The PEPC-28/PPDK-11/MDH-40/ME-116 line was selected for analysis in the present study since its parent PEPC-28/PPDK-11/MDH-40 showed that correct *Zm*PEPC MCs expression and its progenitor PEPC-28 had been detail characterized in Giuliani et al. (2019b). Two quintuple crosses (PEPC-28/PPDK-11/MDH-40/ME-116/*gdch*-31) and (PEPC-28/PPDK-11/MDH-40/ME-116/*gdch*-38) were then generated by crossing the quadruple F_2 line (PEPC-28/PPDK-11/MDH-40/ME-116) with single *Osgdch* knockdown lines (*gdch*-31 and *gdch*-38) described by Lin et al. (2016). The PEPC-28/PPDK-11/MDH-40/ME-116/*gdch*-38 line was chosen for analysis in present study since its progenitor *gdch*-38 had shown a more consistent photorespiratory-deficient phenotype under different O_2:CO_2 growing and measuring conditions compared with *gdch*-31 line, and had been detail characterized in Giuliani et al. (2019a). The presence of transgenes was determined by genomic PCR and protein accumulation by immunoblotting in each crossed line (**Supplementary Figures 4–6**).

Plant Growth

Plants were grown under natural light conditions in a screenhouse with a day/night temperature of 35/28 ± 3°C at the International Rice Research Institute (Los Baños, Philippines: 14° 10019.900N, 121° 15022.300E). Maximum irradiance was 2000 μmol photons m^{-2} s^{-1} on a sunny day. Plants were grown in 7-liter pots filled with soil from the IRRI upland farm.

Immunoblotting

Leaf samples for soluble protein extraction were harvested from the youngest fully expanded leaf at the mid-tillering stage between 09:00 h and 11:00 h, and stored on ice immediately. Leaves were homogenized to a fine powder using a nitrogen-cooled mortar and pestle. Proteins were extracted and fractionated by SDS-PAGE as described previously (Lin et al., 2016). Samples were loaded based on equal leaf area (0.2364 mm^2 for *Zm*PEPC and *Zm*PPDK, and 2.364 mm^2 for *Zm*NADP-MDH, *Zm*NADP-ME and *Os*GDCH). After electrophoresis, proteins were electroblotted onto a polyvinylidene difluoride membrane and probed with rabbit antisera against *Zm*PEPC, *Zm*NADP-MDH, *Zm*NADP-ME (all provided by Richard Leegood, University of Sheffield, United Kingdom), *Zm*PPDK (provided by Chris Chastain, Minnesota State University, United States), and *Os*GDCH protein (provided by Asaph Cousins, Washington State University, United States). The dilutions of *Zm*PEPC, *Zm*PPDK, *Zm*NADP-MDH, *Zm*NADP-ME, and *Os*GDCH antisera were 1:20,000, 1:20,000, 1:5,000, 1:2,000, and 1:100, respectively. A peroxidase-conjugated goat anti-rabbit IgG secondary antibody (Sigma-Aldrich, United States)[2] was used at a dilution of 1:5,000 and immunoreactive bands were visualized with ECL Western Blotting Detection Reagents (GE Healthcare, United Kingdom)[3].

Immunolocalization

The middle portion of the youngest fully expanded leaf at the mid-tillering stage was sampled between 09:00 h and 11:00 h and processed as described previously by Lin et al. (2016). After fixation and cutting, the thin leaf sections were probed with the antisera against *Zm*NADP-MDH, *Zm*NADP-ME, *Zm*PPDK, and *Zm*PEPC at dilutions of 1:500, 1:25, 1:10, and 1:200, respectively. The secondary Alexa Fluor 488 goat anti-rabbit IgG (Invitrogen, United States)[4] antibody was used at a dilution of 1:200. The sections were visualized on a BX61 microscope fitted with a Disk Scanning Unit attachment microscope (Olympus, United States)[5] with fluorescence function under DAPI, RFP, and GFP filters.

Enzyme Activity Measurement

Leaf samples were harvested between 09:00 h and 11:00 h from the youngest fully expanded leaf of plants at the mid-tillering stage, and frozen immediately. Leaves were homogenized to a fine powder using a nitrogen-cooled mortar and pestle and extracted in 250 μL of buffer containing: 50 mM HEPES-KOH, pH 7.4, 5 mM $MgCl_2$, 1 mM EDTA, 1 mM Dithiothreitol, 1% (v/v) glycerol. After centrifugation at 10,000 × *g* for 2 min at 4°C, the supernatant was collected for enzyme activity measurements. PEPC enzyme activity was assayed using a method modified from Meyer et al. (1988) and Ueno et al. (1997). The PEPC reaction mixture contained: 100 mM HEPES-NaOH, pH 7.5, 10 mM $MgCl_2$, 1 mM $NaHCO_3$, 5 mM G6P, 0.2 mM NADH, 12 unit/mL MDH (from pig heart; Roche Diagnostics, Basel, Switzerland)[6], and the reaction was started by adding PEP to a final concentration of 4 mM. PPDK enzyme activity was assayed as described by Fukayama et al. (2001). NADP-MDH activity was determined by a method modified from Tsuchida et al. (2001). NADP-MDH reaction mixture contained: 50 mM HEPES-KOH, pH 8, 70 mM KCl, 1 mM EDTA, 1 mM DTT, and 0.2 mM NADPH, and the reaction was started by adding OAA to a final concentration of 1 mM. NADP-ME activity was measured by a method modified from Tsuchida et al. (2001; protocol 1). The activities of PEPC, PPDK, NADP-MDH, and NADP-ME were measured spectrophotometrically at 340 nm at 25°C, 30°C, 25°C, and 25°C, respectively.

Gas Exchange Measurements

Leaf gas-exchange measurements were made using a Li-6400XT infrared gas analyzer (LI-COR Biosciences, United States)[7] fitted with a standard 2 × 3 cm leaf chamber and a 6400-02B light source. Measurements were made at a constant airflow rate of 400 μmol s^{-1}, leaf temperature of 25°C, leaf-to-air vapor pressure deficit between 1.0 and 1.5 kPa and relative humidity of 60–65%. Data were acquired between 08:00 h and 13:00 h.

[2]https://www.sigmaaldrich.com/

[3]https://www.gelifesciences.com

[4]https://www.thermofisher.com/ph/en/home/brands/invitrogen.html

[5]https://www.olympus-global.com/

[6]https://www.roche.com/

[7]https://www.licor.com/

Measurements were made from the two youngest fully expanded leaves for each plant during the tillering stage. The mid-portions of leaves were acclimated in the cuvette for approximately 30 min before measurements were made. The response curves of the net CO_2 assimilation rate (A, μmol m^{-2} s^{-1}) to changing intercellular pCO_2 concentration (C_i, μmol CO_2 mol air^{-1}) were acquired by decreasing C_a (pCO_2 concentration in the cuvette) from 2000 down to 20 μmol CO_2 mol air^{-1} at a photosynthetic photon flux density (PPFD) of 1500 or 2000 μmol photon m^{-2} s^{-1}. The CO_2 compensation point (Γ) and maximum carboxylation efficiency (*CE*) were calculated from the intercept (Vogan et al., 2007) and slope (Wang et al., 2006) of the CO_2 response curves. Light response curves were acquired by increasing the PPFD from 0 to 2000 μmol photon m^{-2} s^{-1} at C_a 400 μbar. The quantum efficiency for CO_2 assimilation (φ) and respiration rates (R_d) were calculated from the slope and intercept of the light-response curves (PPFD < 100 μmol photons m^{-2} s^{-1}).

$^{13}CO_2$ Pulse-Labeling and Quenching Procedure

Carbon flux analysis was performed with a custom-built gas exchange freeze clamp apparatus (**Supplementary Figure 7**). Measurements were made from two youngest fully expanded leaves for each plant during the tillering stage. Two leaves of up to 22 cm in length were placed inside a gas exchange chamber (23.5 cm × 4.5 cm × 0.4 cm), the top was constructed from a piece of clear flexible plastic to allow light penetration and the bottom from a sheet of aluminum foil to accelerate cooling when freeze clamping. The foil and plastic were attached with foil tape to a three-sided aluminum frame to provide rigidity. Two holes were drilled through the side of the frame to accommodate the air inlet and outlet tubes. A third hole on the end enabled thermocouples to be threaded through the frame to measure leaf and air temperature inside the labeling chamber. The chamber was then placed in a mounting frame allowing the leaf to be inserted prior to sealing the chamber with a foam gasket secured with bulldog clips. The mounting frame was positioned horizontally between two LED banks capable of providing illumination to the upper leaf surface of up to 1,000 μmol photons m^2 s^{-1}.

Air was drawn from outside the laboratory through a compressor. The air stream passed through an oil water separator and flow control valve into a copper coil placed in an ice bath for cooling. The air stream could be directed into the leaf chamber or by-passed into a CO_2 conditioning unit. In the latter, CO_2 could be removed from the air with soda lime and then optionally enriched with $^{13}CO_2$ gas (300 ppm) prepared by mixing $NaH^{13}CO_3$ (Sigma-Aldrich, United States; see text footnote 2) with 2-hydroxypropanoic acid (lactic acid). The flow of air passing over the leaf (3 ml/min) was adjusted with a flow controller, and the CO_2 concentration of the incoming and outgoing air streams measured with two CO_2 analyzers (WMA-5, PP-Systems, United States)[8]. The humidity of the air inside the chamber was maintained at ∼60% with the addition of water to the soda lime chamber or reduced by passing air through a chamber of silica gel.

The leaf chamber was mounted on the stand in such a way that the plane of the leaf was halfway between two pneumatically operated aluminum bars. These were cooled with liquid nitrogen, and when released they clamped together fitting inside the aluminum chamber frame, very rapidly freezing the leaf. A fan was mounted horizontally to the bars to blow the fog from the liquid nitrogen away from the chamber to ensure there was minimal disruption to the environment before freezing occurred. The lower bar was positioned in such a way as to push the chamber up on closure. The cold temperatures and force of the bars closing meant the chamber disintegrated enabling the leaf to be removed with tweezers and placed in a liquid nitrogen bath for 10 s before subsequent storage at −80°C. To perform $^{13}CO_2$ pulse-labeling, leaves were acclimated at steady-state conditions prior to scrubbing CO_2 from the incoming air stream, then subjected to $^{13}CO_2$ enriched air for a duration of 0 and 60 s and metabolic activity was quenched at these time points by freeze clamping the leaves as above. Freeze-quenched tissue was ground into a fine powder by mortar and pestle in liquid nitrogen. Finely ground leaf tissues were freeze-dried for 3 days and placed into sealed tubes. The sealed tubes containing finely ground lyophilized leaf tissue samples were shipped to Max Planck Institute of Molecular Plant Physiology, Germany for the metabolite analyses.

Metabolite Analyses and Calculation of Total Pool Size, Enrichment and Isotopomer Distribution

Aliquots (3 or 5 mg) of finely ground lyophilized rice leaf tissue were extracted with chloroform-methanol as described in Arrivault et al. (2017), and the lyophilized extracts were resuspended in 300 μL or 600 μL purified (Millipore, United States) water, respectively. Isotopomers were measured by reverse-phase LC-MS/MS (malate, aspartate, 3PGA, PEP, citrate + isocitrate; Arrivault et al., 2017; n.b. citrate and isocitrate were not resolved using this method) and anion-exchange LC-MS/MS (malate, PEP, citrate; Lunn et al., 2006 with modifications as described in Figueroa et al., 2016). Total amounts of malate, aspartate, citrate + isocitrate, and citrate were calculated by summing isotopomers. The total amounts of 3PGA and PEP were determined enzymatically in trichloroacetic acid extracts using a Sigma-22 dual-wavelength photometer (Merlo et al., 1993), with PEP being measured in freshly prepared extracts. ^{13}C enrichment and relative isotopomer distribution were calculated as in Szecowka et al. (2013).

Statistical Analysis

Statistical analysis for all experiments was performed in R version 3.0.0 (The R Foundation for Statistical Computing, Vienna, Austria) using a one-way analysis of variance (ANOVA)

[8]https://ppsystems.com/

and a Tukey *post hoc* test or a Student's *t*-test with a *p*-value of <0.05.

RESULTS

Overexpression of C_4 Cycle Genes in *Oryza sativa*

Immunoblotting of F_2 generation transgenic plants showed that the protein of the correct size for *Zm*PEPC, *Zm*NADP-MDH, *Zm*NADP-ME, and *Zm*PPDK was stably expressed in both the quadruple and quintuple cross rice lines, with *Os*GDCH protein almost undetectable in the quintuple cross (**Figure 1**). In all lines, protein abundance was lower than that of wild-type maize plants. We next sought to determine whether these proteins were localized to the correct cell type and subcellular compartment. Immunolocalization analysis of the quadruple cross line revealed that *Zm*PEPC was localized to the cytosol of MCs similar to the single *Zm*PEPC transgenic line (Giuliani et al., 2019b) and the triple cross line (**Supplementary Figure 3**). *Zm*PPDK, was localized to the chloroplast in both BSCs and MCs (**Supplementary Figure 8**). The *Zm*NADP-MDH and *Zm*NADP-ME antisera cross-reacted with native protein in wild-type rice and so it was not possible to distinguish protein encoded by the endogenous rice gene from that encoded by the maize transgene. Overexpression of these enzymes conferred enhanced enzyme activity (**Supplementary Table 1**). In the quadruple cross line, PEPC activity was 18.4-fold higher compared to wild-type rice plants, PPDK 5.8-fold higher, NADP-MDH 9.9-fold, and NADP-ME 4.1-fold (**Table 1**).

Phenotypic and Photosynthetic Perturbations Associated With Overexpression of C_4 Cycle Genes

Given that protein levels and activities of all four introduced C_4 enzymes were enhanced compared to wild-type plants, we investigated whether this affected growth and photosynthesis. None of the crossed lines consistently showed altered chlorophyll content (**Figure 2A**). Tiller number in the quintuple cross lines (**Figure 2B**) and plant height in both quadruple and quintuple crosses (**Figure 2C**) were significantly reduced. Phenotypic perturbations were most marked in the quintuple cross lines (**Supplementary Figure 9**), although the plants still developed and flowered at the same time with wild-type plants. These growth perturbations were not observed in the single C_4 gene transgenic lines of *Zm*PEPC, *Zm*PPDK, *Zm*NADP-MDH, and *Zm*NADP-ME (Giuliani et al., 2019b) but were observed in the single *Osgdch* knockdown line (Lin et al., 2016).

To investigate whether overexpression of C_4 genes impacted photosynthesis, the response of net CO_2 assimilation rate (A) to CO_2 concentration under photorespiratory conditions (21% O_2) was measured. In the quadruple cross line there were no differences in CO_2 assimilation (**Figures 3A,C**), Γ, CE, R_d or Φ compared to wild-type plants (**Table 2**). These results suggested

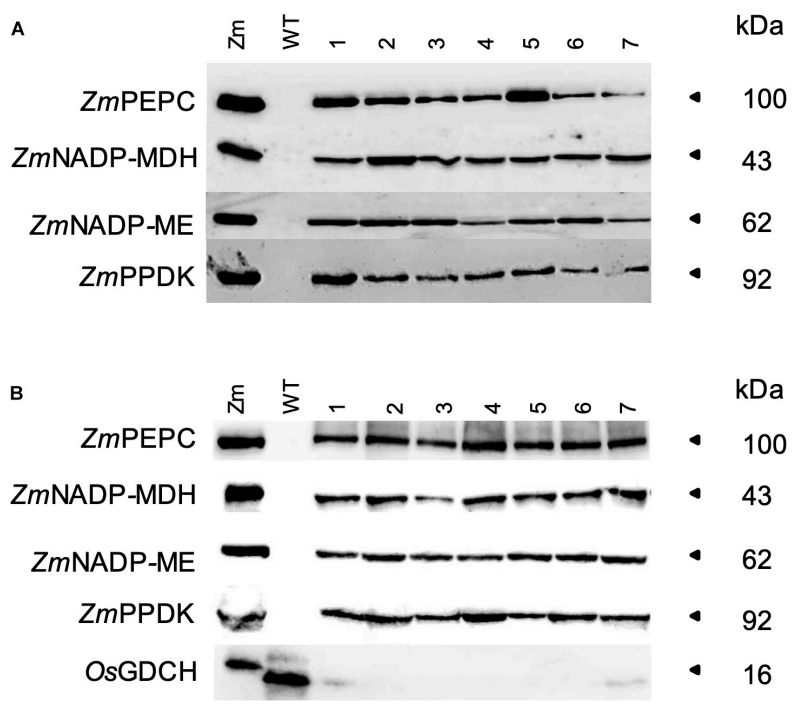

FIGURE 1 | Soluble leaf protein. Immunoblots for **(A)** Quadruple and **(B)** Quintuple crosses. Maize (Zm), wild-type rice (WT), and plants of F_2 crosses (numbers). Protein was extracted from the youngest fully expanded leaf at mid-tillering stage. Samples were loaded on an equal leaf area basis (0.2364 mm^2 for *Zm*PEPC and *Zm*PPDK, and 2.364 mm^2 for *Zm*NADP-MDH, *Zm*NADP-ME, and *Os*GDCH).

TABLE 1 | The enzyme activities of PEPC, NADP-MDH, NADP-ME, and PPDK.

	PEPC	NADP-MDH	NADP-ME	PPDK
	$\mu mol\ s^{-1}\ m^{-2}$	$\mu mol\ s^{-1}\ m^{-2}$	$\mu mol\ s^{-1}\ m^{-2}$	$\mu mol\ s^{-1}\ m^{-2}$
Maize	42.54 ± 1.09^a	12.81 ± 0.39^b	24.21 ± 2.38^a	9.44 ± 1.34^a
WT	1.03 ± 0.01^c	3.50 ± 0.03^c	0.56 ± 0.07^c	0.30 ± 0.03^c
Quadruple	18.96 ± 2.56^b	34.98 ± 3.26^a	2.34 ± 0.10^b	1.76 ± 0.14^b

Values are average ± SE of 2–3 plants of maize, wild-type rice (WT), and F_2 quadruple crosses. Different letters within groups indicated those values that are statistically different based on a one-way ANOVA, P-value < 0.05.

that accumulation of active maize C_4 cycle enzymes in rice does not significantly affect leaf level photosynthetic gas-exchange and that the phenotypic perturbations are not associated with reduced CO_2 assimilation.

For the quintuple cross, the CO_2 assimilation rate in response to increasing intercellular CO_2 concentrations (*Ci*) was reduced, most notably at lower CO_2 concentrations (<700 μmol CO_2 mol^{-1}, **Figure 3B**). Under non-photorespiratory conditions (2% O_2, **Figure 3B**) CO_2 assimilation rates were similar to wild-type plants. Consistent with this, the quintuple cross had a significantly higher Γ under high photorespiratory conditions but not under low photorespiratory conditions (**Table 2**). In response to changes in photon flux density, photosynthesis was saturated at lower light levels than in wild-type plants (400 μmol photons m^{-2} s^{-1} versus 1750 μmol photons m^{-2} s^{-1}, respectively, **Figure 3D**) with significantly lower Φ and higher R_d than wild-type plants (**Table 2**).

To investigate these photosynthetic responses for the quintuple cross line in more detail, CO_2 responses were measured under conditions conducive to low and high rates of photorespiration. Under low light (400 μmol photons m^{-2} s^{-1}) and high CO_2 (2000 μmol CO_2 mol air^{-1}), conditions conducive to low photorespiration, CO_2 responses of the quintuple cross line were similar to wild-type plants (**Supplementary Figures 10A,D**). Under high light (1500 μmol photons m^{-2} s^{-1}) and low CO_2 (400 μmol CO_2 mol air^{-1}), conditions conducive to high photorespiration, CO_2 assimilation was lower in the quintuple cross (**Supplementary Figures 10B,C**). Correspondingly, Γ was higher, and carboxylation efficiency (*CE*) lower under conditions conductive to high rates of photorespiration but not under non-photorespiratory conditions (**Supplementary Table 1**). This is consistent with the photorespiratory-deficient phenotype observed in the single *Osgdch* knockdown lines (Lin et al., 2016).

Increased Incorporation of ^{13}C Into C_4 Acids Associated With Overexpression of C_4 Cycle Enzymes

We performed experiments to measure the flux of $^{13}CO_2$ through photosynthetic metabolism to investigate whether there was partial functionality of a C_4 pathway in these plants. There was significantly more incorporation of ^{13}C into malate in

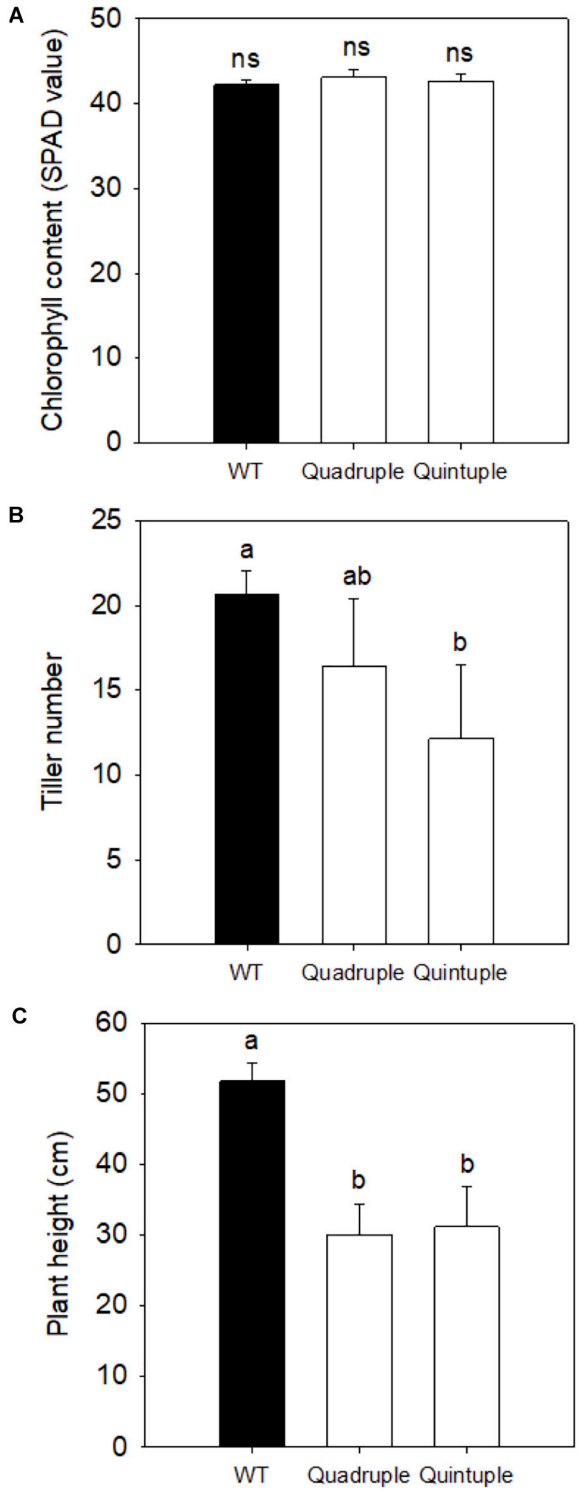

FIGURE 2 | (A) Leaf Chlorophyll content, **(B)** tiller number, and **(C)** plant height of wild-type (WT), Quadruple, and Quintuple lines. Chlorophyll SPAD values, tiller number, and plant height are means ± SE of eight individual F_2 plants and eight WT plants, 90 days post germination. Different letters within groups indicated those values that are statistically different based on a one-way ANOVA with a Tukey multiple comparison test for *post hoc* pairwise comparison, *P*-value < 0.05. ns indicates non-significant.

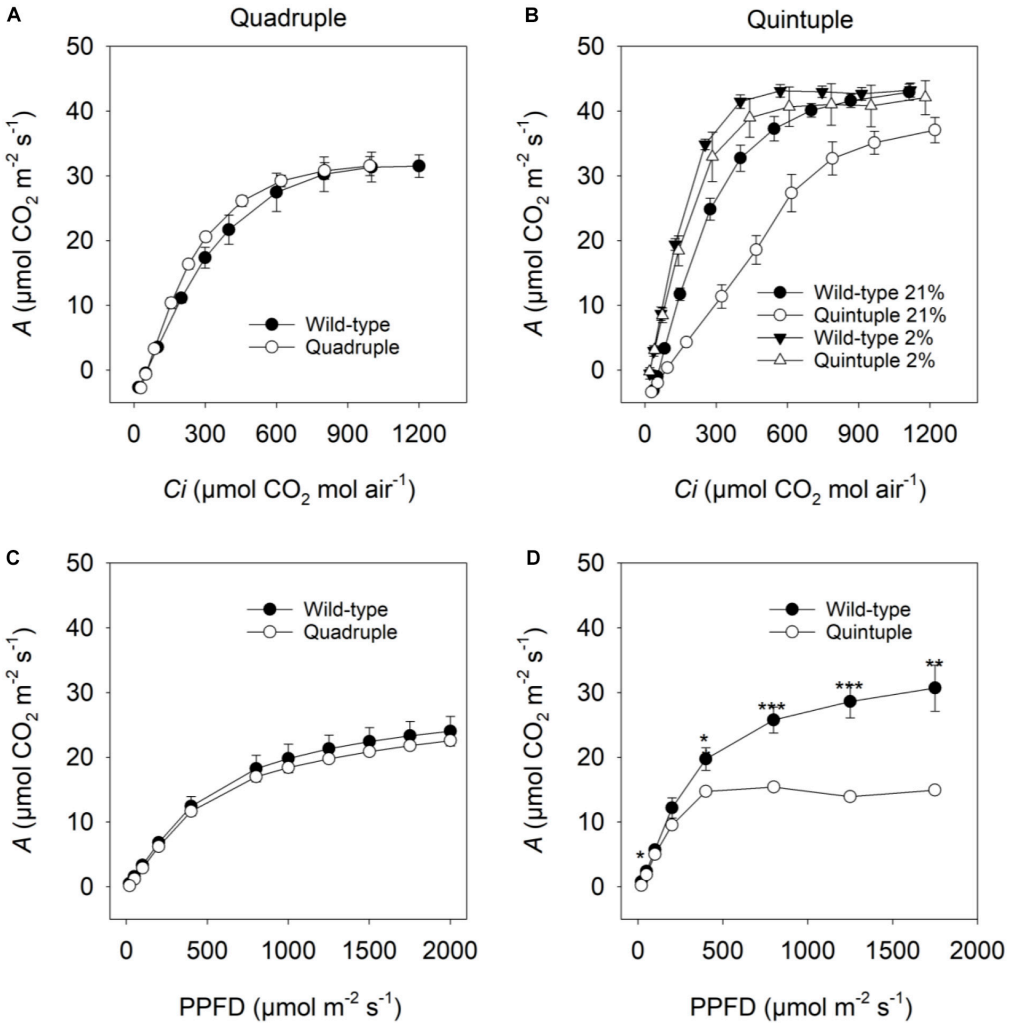

FIGURE 3 | (A,B) Net CO_2 assimilation rate (*A*) in response to intercellular pCO_2 (*Ci*) and **(C,D)** photosynthetic photon flux density (PPFD). Measurements **(A,B)** were made at 2000 μmol photons m^{-2} s^{-1} under 21% O_2 **(A,B)** and 2% O_2 **(B)**. Measurements **(C,D)** were made at a pCO_2 (C_a) of 400 μmol CO_2 mol air^{-1}. Values are means ± SE of three F_2 cross plants and three wild-type (WT) plants for quadruple comparisons and seven F_2 cross plants and four WT plants for quintuple comparisons. A Student's *t*-test was performed for **(C,D)**. Significant differences between WT and quintuple within PARi level are indicated by **P*-value < 0.05, ***P*-value < 0.01, ****P*-value < 0.001.

TABLE 2 | Comparison of photosynthetic parameters.

	Γ	*CE*	R_d	Φ
	μmol CO_2 m^{-2} s^{-1}	μmol CO_2 m^{-2} s^{-1} μmol CO_2 mol^{-1}	μmol CO_2 m^{-2} s^{-1}	mol CO_2 mol^{-1} quanta
WT 21% O_2	53.44 ± 0.40	0.11 ± 0.01	0.23 ± 0.04	0.04 ± 0.00
Quadruple 21% O_2	56.11 ± 5.24	0.10 ± 0.01	0.55 ± 0.25	0.03 ± 0.00
WT 21% O_2	59.48 ± 1.70^b	0.13 ± 0.01^a	0.60 ± 0.20^b	0.06 ± 0.01^a
Quintuple 21% O_2	88.76 ± 4.13^a	0.05 ± 0.00^b	1.09 ± 0.23^a	0.05 ± 0.00^b
WT 2% O_2	23.17 ± 5.07	0.19 ± 0.01^a	–	–
Quintuple 2% O_2	21.34 ± 1.74	0.15 ± 0.01^b		–

CO_2 compensation point (Γ), carboxylation efficiency (CE), respiration rates (R_d), and quantum yield for CO_2 assimilation (φ). Measurements of Γ and CE were made at a PPFD of 2000 μmol photons m^{-2} s^{-1}. Φ at a pCO_2 (C_a) of 400 μmol CO_2 mol air^{-1} and a leaf temperature of 25°C. Values are means ± SE of three F_2 cross plants and three wild-type (WT) plants for quadruple comparisons and seven F_2 cross plants and four WT plants for quintuple comparisons. Different letters within groups indicated those values that are statistically different based on a Student's t-test, P-value < 0.05.

the quadruple and quintuple cross lines than in wild-type plants (**Figure 4** and **Supplementary Figure 11**), with the m_1 isotopomer being more abundant than the other ^{13}C-labeled isotopomers (**Table 3**), consistent with increased fixation of $^{13}CO_2$ via PEPC in the transgenic lines expressing the maize PEPC enzyme. Labeling of aspartate was also significantly higher than wild-type in the quintuple line (**Figure 4** and **Supplementary Figure 11**), with the m_1 isotopomer being more abundant that m_2-m_4 isotopomers (**Table 3**), indicating $^{13}CO_2$ fixation into oxaloacetate by maize PEPC and conversion to aspartate by endogenous rice aspartate aminotransferase activity. There was no evidence of significant difference in labeling of aspartate in the quadruple line. There was almost complete labeling of the 3PGA pool in wild-type plants after a 60 s pulse, and similarly high levels of labeling of 3PGA were observed in the quadruple and quintuple lines (**Figure 4** and **Supplementary Figure 11**). There was also substantial labeling of PEP after a 60 s pulse, with no significant differences between the transgenic lines and wild-type plants (**Figure 4** and **Supplementary Figure 11**). There was almost no labeling of citrate and isocitrate in wild-type plants and the quadruple line (**Figure 4**). In contrast, the enrichment of ^{13}C in citrate and isocitrate in the quintuple line was 10-fold higher than in wild-type plants.

DISCUSSION

We have previously shown that overproduction of individual C_4 enzymes in rice has no consistent effect on CO_2 assimilation or plant growth (Giuliani et al., 2019b; Karki et al., 2020).

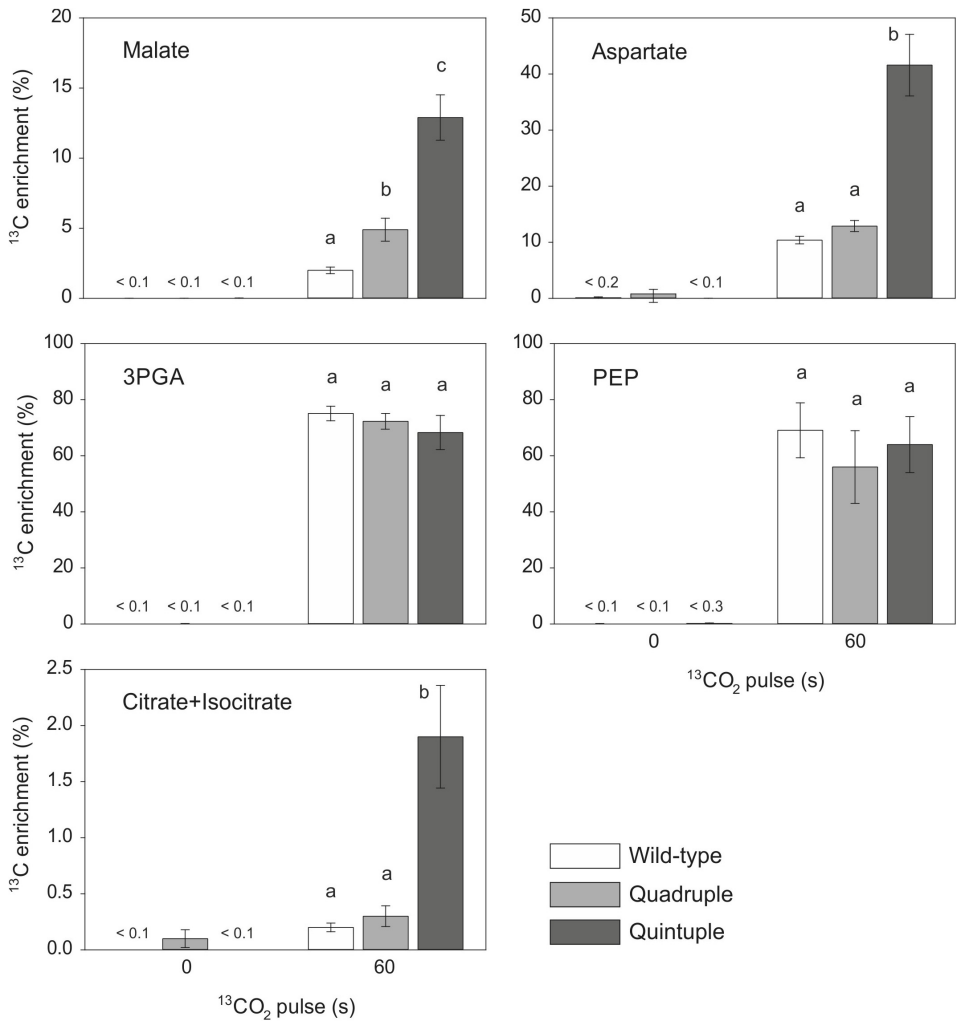

FIGURE 4 | $^{13}CO_2$ pulse-labeling of wild-type, quadruple and quintuple rice cross lines. Fully expanded leaves at mid-tillering stage were pulse-labeled with $^{13}CO_2$ (300 ppm) for 60 s under steady state photosynthetic conditions. Isotopomers of malate, aspartate, 3PGA, PEP, and citrate + isocitrate were measured in extracts from pulse-labeled (60 s) and non-labeled (0 s) leaves by LC-MS/MS, and ^{13}C enrichment (%) was calculated after correction for natural abundance. Values are means ± SE (n = 2–4). The original data are presented in **Supplementary Dataset A**. Different letters within groups indicated those values that are statistically different based on a one-way ANOVA, P-value < 0.05.

TABLE 3 | Relative isotopomer distribution (%) of malate and aspartate in wild-type, quadruple and quintuple rice lines after pulse-labeling with $^{13}CO_2$ for 60 s.

		m_0	m_1	m_2	m_3	m_4
Malate	Wild-type	96.4 ± 0.35	1.2 ± 0.17	1.0 ± 0.08	0.6 ± 0.07	0.7 ± 0.03
	Quadruple	90.5 ± 1.09	4.1 ± 0.64	2.4 ± 0.27	1.3 ± 0.12	1.6 ± 0.2
	Quintuple	79.3 ± 1.15	7.0 ± 0.02	3.3 ± 0.11	3.6 ± 0.19	6.8 ± 1.06
Aspartate	Wild-type	77.2 ± 0.19	11.6 ± 0.83	5.5 ± 0.29	3.8 ± 0.72	1.9 ± 0.01
	Quadruple	71.5 ± 1.67	14.7 ± 2.49	6.7 ± 1.15	4.7 ± 0.76	2.3 ± 1.22
	Quintuple	28.1 ± 7.37	26.6 ± 2.93	12.5 ± 2.13	16.4 ± 0.89	16.4 ± 1.41

The relative abundance of each isotopomer (m_n) for a given metabolite is represented; n is the number of ^{13}C atoms incorporated. Values are means ± SE of 2–4 plants of wild-type rice and F_2 quadruple crosses. The original data are presented in ***Supplementary Dataset A.***

The exception to this was the transgenic line overexpressing *Zm*NADP-ME which exhibited a small decrease in plant height and reduced maximal photosynthetic rate at high CO_2. Previous attempts to overproduce *Zm*NADP-ME in rice have led to increased photoinhibition of photosynthesis, leaf chlorophyll bleaching and serious stunting attributed to an increase in the NADPH/$NADP^+$ ratio in the chloroplast stroma due to the exchange with 2-oxoglutarate involved in photorespiration (Tsuchida et al., 2001). Severe phenotypic effects were not observed in our *Zm*NADP-ME line; however, we were only able to advance a single line containing 6 copies of the construct in which protein accumulation was higher than in our rice control, but still only around 10% of the activity found in maize. In contrast, the experiments of Tsuchida et al. (2001) used a rice chlorophyll a/b binging protein (*cab*) promoter allowed for high level but not cell specific expression, and activities of up to 60% of maize levels were achieved, leading to a much more severe phenotype. Overproduction of all four targeted C_4 enzymes in a single plant led to a slight decrease in tiller number and plant height but otherwise growth and photosynthesis were unaffected. These results are consistent with previous reports of engineering a single-cell C_4 pathway in rice (Taniguchi et al., 2008). A quintuple cross that combined overexpression of the four C_4 enzymes, *Zm*PEPC, *Zm*NADP-MDH, *Zm*NADP-ME, and *Zm*PPDK with knockdown of the native rice *Os*GDCH, thereby compromising the photorespiratory pathway, led to further reductions in tiller number and plant height. A strong negative effect on photosynthesis was also observed in the quintuple cross consistent with the photorespiratory-deficient phenotype of the single *Osgdch* knockdown line (Lin et al., 2016).

Our results show that *Zm*PEPC is catalytically active *in vivo* when expressed in combination with other C_4 enzymes in rice, and substantially increases the fixation of CO_2 into C_4 acids. This is in contrast to published radiolabeling studies of rice expressing *Zm*PEPC alone (Fukayama et al., 2003; Miyao et al., 2011) in which there was no increase in incorporation of labeled carbon into C_4 acids, despite the extractable activity of PEPC in these plants approaching or exceeding maize levels. Despite strong evidence for operation of a partial C_4 pathway in our transgenic lines up to the point of malate production, there was no evidence for the regeneration of PEP via the rest of the C_4 cycle. The labeling of PEP at a similar level in all three genotypes, wild-type, quadruple and quintuple crosses, suggests that rather than being produced by a functional C_4 cycle, PEP is being produced from 3PGA via 2PGA catalyzed by phosphoglyceromutase and enolase (Furbank and Leegood, 1984), consistent with the majority of CO_2 still being fixed via Rubisco in C_3 photosynthesis rather than through the operation of a complete C_4 cycle.

The very low incorporation of ^{13}C into citrate and isocitrate in wild-type rice plants is consistent with previous studies (Tcherkez et al., 2009; Szecowka et al., 2013) indicating little flux of carbon into the tricarboxylic acid (TCA) cycle via mitochondrial pyruvate dehydrogenase (mPDH) in the light, due to deactivation of the mPDH by phosphorylation (Randall et al., 1996; Tovar-Méndez et al., 2003). The increased labeling of citrate in the quintuple cross line suggests that the mPDH is more active in the light in this line, potentially leading to respiration of C_4 acids via the TCA cycle, which would be deleterious for C_4 photosynthetic flux. This might be due to lower rates of photorespiration leading to less photophosphorylation of PDH (Tovar-Méndez et al., 2003). Further, increased levels of pyruvate in the mitochondria (from decarboxylation of malate by NAD-malic enzyme), can inhibit the mPDK kinase (Schuller and Randall, 1990). We propose that there is a modified regulation of the TCA cycle to avoid wasteful respiration of C_4 acids, and that such modification might have been needed for the evolution of an efficient C_4 photosynthetic pathway.

Evidence that *Zm*PEPC can be localized to the cytosol of MCs (Giuliani et al., 2019b) and is catalytically active, leading to the fixation of CO_2 into C_4 acids provides important evidence in support of installing a fully functional C_4 photosynthetic pathway into rice. However, absolute quantification of flux into and through C_4 acids would require further pulse-chase labeling studies and may prove difficult with the low rates of labeling relative to C_3 photosynthetic fixation obtained in the current transgenic lines.

Achieving the correct cellular localization of the C_4 enzymes introduced into the rice lines shown here remains an important and unresolved issue. We introduced intact maize C_4-specific genes containing the promoter into rice (Karki et al., 2020) based on the strategy by Miyao et al. (2011). This approach was originally adopted because it had been reported that the 5′- flanking region of the maize C_4 specific genes drove high-level MC-specific expression of a reporter β-glucuronidase (GUS) gene in rice leaves (Matsuoka et al., 1993, 1994). However, we knew at the time that this approach might not lead to cell specific expression of the enzymes (Sheen and Bogorad, 1987;

Nomura et al., 2005). Miyao et al. (2011) had reported that expression of the native promoter and full-length maize PPDK gene in rice led to the accumulation of protein in both the MCs and BSCs. At the same time they also raised the possibility that maize PEPC might also accumulate in both cell types, Therefore, it was not unexpected that in the lines reported here, or the single transgenic lines used as parents for the crosses in this study (Karki et al., 2020), PPDK and PEPC accumulated in both cell types, It remains unclear why in selected events, maize PEPC appears to confined the BS (Giuliani et al., 2019b). Owing to the absence antibodies specific for *Zm*NADP-ME and *Zm*NADP-MDH, we have been unable to establish the cell localization of these enzymes. However, given that the promoters used, cell-specific expression seems unlikely (Nomura et al., 2005).

To mitigate the risk associated with this approach we also generated constructs where each coding sequence (CDS) was fused to either the M promoter *ZmPEPC* (Matsuoka et al., 1993) or the BS promoter *OsPCK1* (Nomura et al., 2005) with the *nopaline synthase* (*nos*) terminator at the 3′ end. These lines were not used as neither approach led to the cell specific expression (unpublished, W. P. Quick, personal communication). As no alternative strategy for achieving cell-specific expression was available at the time, lines containing the full length genes and native promoters were used as these lead to an enrichment of maize PEPC and PDDK in the correct cell type. Incorrect or partial localization of enzymes would potentially limit the operation of a C_4 cycle (Miyao et al., 2011), with the potential to lead to deleterious phenotypes, i.e., (Tsuchida et al., 2001). Thus, the cell-specific expression of enzymes remains an active area of research for the consortium. It has been suggested that gene specificity may be generated by elements that are not present in the promoter (Hibberd and Covshoff, 2010).

In addition to high level, cell specific expression of C_4 cycle enzymes in rice, fully functional C_4 photosynthetic biochemistry requires appropriate enzyme regulation in the environment of a rice leaf cell (Burnell and Hatch, 1985; Chastain et al., 1997). For example, the activity of C_4 specific PPDK is regulated in the light through protein phosphorylation by the PPDK regulatory protein (Burnell and Hatch, 1985). Similarly, NADP-MDH is regulated by light through the thioredoxin cascade (Miginiac-Maslow et al., 2000). Both enzymes are regulated in the same manner even when expressed within C_3 leaves (Fukayama et al., 2001; Taniguchi et al., 2008). A recent study has shown that C_4 NADP-ME is also regulated in the light by reversible phosphorylation at Ser419 which is involved in the binding of NADP at the active site (Bovdilova et al., 2019). In contrast, PEPC is regulated by both metabolite effectors and reversible phosphorylation, but the mechanisms of regulation in C_3 and C_4 leaves are different (Vidal and Chollet, 1997). Indeed, Fukayama et al. (2003) observed inappropriate phosphorylation of PEPC in their transgenic rice lines and proposed this as a reason for lack of labeling of C_4 acids in the light. The regulatory mechanisms for other enzymes are less well understood. It is unclear at present whether enzyme levels *per se* or enzyme regulation in our rice transgenic lines, or both, is limiting C_4 flux.

In NADP-ME C_4 plants assimilation of a single CO_2 molecule requires at least 10 transport steps between cells and within subcellular compartments of the MC and BSC. The identity of most of the transporter proteins supporting is now known, although there remains some uncertainty about malate import to the BS chloroplast and the export of pyruvate following malate decarboxylation (Weber and von Caemmerer, 2010; Ermakova et al., 2019). The next logical step is to introduce these into the current prototype. This was initially planned and the started as part of a 6-year strategy to develop a prototype expressing all known genes required to support the C_4 biochemical pathway. Since then the emergence of Golden Gate cloning has enabled the consortium to reduce that strategy to 6-months by creating a large multigene overexpression construct (Ermakova et al., 2019). This makes the prototype development strategy adopted in this study obsolete.

A plethora of other changes are required to support a fully functional C_4 pathway in rice. This includes, but is not limited to, engineering the correct leaf anatomy (Hattersley and Watson, 1975; Dengler et al., 1994; Dengler and Taylor, 2000; Muhaidat et al., 2007) and morphological specializations such as increased vein density (Sedelnikova et al., 2018). In addition, thought must be given to the photosynthetic functionalization of the BSCs of rice, which contain a large central vacuole, with very few mitochondria, peroxisomes or chloroplasts (Sage and Sage, 2009; Ermakova et al., 2019). Where chloroplasts do occur, they are smaller than those in MCs. Increasing chloroplast number and volume in the BSCs will no doubt be important for achieving C_4 photosynthesis in rice (Chonan, 1970, 1978; Dengler et al., 1994; Ueno et al., 2006; Wang et al., 2017). Insufficient chloroplast volume in the BSCs of rice may have led to limitations in C_4 acid decarboxylation in the transgenic lines described here. In addition, MCs of rice are highly lobed to assist with photorespiratory CO_2 scavenging (Sage and Sage, 2009); whereas the MCs of C_4 species are not. It is possible that these features may hinder the transport of metabolites between cells (Sage and Sage, 2009). Other modifications such as cross-sectional area of the BSCs, modifying the cell wall properties for diffusion of CO_2 (von Caemmerer and Furbank, 2003) and increasing plasmodesmatal frequency at the BSC/MC interface to support metabolite diffusion may be necessary (Ermakova et al., 2019). The genetic regulators of many of these changes are not known, and so future goals include identification and incorporation of necessary genes for anatomical modifications into a version of the current biochemical prototype, with the ultimate goal of engineering an efficient C_4 pathway in rice. The research presented here represents a small step toward this goal.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**supplementary material.**

AUTHOR CONTRIBUTIONS

HL, SA, RC, JL, MS, RF, and WQ designed the experiments together. HL provided all the plant materials. HL and

EB performed enzyme activity assay, immunoblotting, immunolocalization, and gas exchange measurements. WQ, RF, MS, JL, and RC designed the gas exchange freeze clamp apparatus. SA performed metabolite analysis. HL, SA, RC, and WQ wrote the manuscript. SC and JH designed constructs. SK performed plant transformation. HL and RC performed the $^{13}CO_2$ labeling experiment. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the C_4 Rice Project grants from the Bill & Melinda Gates Foundation to IRRI (Grant ID#51586) and the University of Oxford (OPP1129902), and by the Max Planck Society (SA, JL, and MS).

ACKNOWLEDGMENTS

This article was dedicated to the memory of Dr. John Sheehy, who was the architect and inspiration of the C_4 Rice Consortium and its attempts to engineer a rice plant with C_4 or C_4-like photosynthesis. We wish to thank Florencia Montecillo, Juvy Reyes, and Irma Canicosa for their help with plant transformation, husbandry, and physiological measurements at IRRI C_4 Rice Centre as well as Regina Feil and Manuela Guenther for metabolite measurements at the MPI-MP.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.564463/full#supplementary-material

Supplementary Dataset A | Isotopomer and metabolite amounts, ^{13}C enrichments and relative isotopomer abundances of malate, aspartate, 3PGA, PEP and citrate+isocitrate in wild-type, quadruple and quintuple lines.

Supplementary Dataset B | Isotopomer and metabolite amounts, ^{13}C enrichments and relative isotopomer abundances of malate, aspartate, 3PGA, PEP, and citrate in wild-type and quintuple lines.

REFERENCES

Arrivault, S., Obata, T., Szecówka, M., Mengin, V., Guenther, M., Hoehne, M., et al. (2017). Metabolite pools and carbon flow during C4 photosynthesis in maize:13CO2 labeling kinetics and cell type fractionation. *J. Exp. Bot.* 68, 283–298. doi: 10.1093/jxb/erw414

Bovdilova, A., Alexandre, B. M., Höppner, A., Luís, I. M., Alvarez, C. E., Bickel, D., et al. (2019). Posttranslational Modification of the NADP-Malic Enzyme Involved in C4 Photosynthesis Modulates the Enzymatic Activity during the Day. *Plant Cell* 31, 2525–2539. doi: 10.1105/tpc.19.00406

Burnell, J. N., and Hatch, M. D. (1985). Light-dark modulation of leaf pyruvate, Pi dikinase. *Trends Biochem. Sci.* 10, 288–291. doi: 10.1016/0968-0004(85)90090-8

Chastain, C. J., Lee, M. E., Moorman, M. A., Shameekumar, P., and Chollet, R. (1997). Site-directed mutagenesis of maize recombinant C4-pyruvate, orthophosphate dikinase at the phosphorylatable target threonine residue. *FEBS Lett.* 413, 169–173. doi: 10.1016/S0014-5793(97)00884-3

Chonan, N. (1970). Studies on the photosynthetic tissues in the leaves of cereal crops. V. Comparison of the mesophyll structure among seedling leaves of cereal crops. *Proc. Crop Sci. Soc. Jpn.* 39, 418–425. doi: 10.1626/jcs.39.418

Chonan, N. (1978). Comparative anatomy of mesophyll among the leaves of gramineous crops. *Jpn. Agr. Res. Q.* 12, 128–131.

Dengler, N., and Taylor, W. C. (2000). "Developmental aspects of C4 photosynthesis," in *Photosynthesis: Physiology and Metabolism*, eds R. C. Leegood, T. D. Sharkey, and S. von Caemmerer (Dordrecht: Springer), 471–495. doi: 10.1007/0-306-48137-5_20

Dengler, N. G., Dengler, R. E., Donnelly, P. M., and Hattersley, P. W. (1994). Quantitative leaf anatomy of C3 and C4 grasses (Poaceae)—bundle- sheath and mesophyll surface-area relationships. *Ann. Bot.* 73, 241–255. doi: 10.1006/anbo.1994.1029

Ermakova, M., Danila, F. R., Furbank, R. T., and von Caemmerer, S. (2019). On the road to C4 rice: advances and perspectives. *Plant J.* 101, 940–950. doi: 10.1111/tpj.14562

Figueroa, C. M., Feil, R., Ishihara, H., Watanabe, M., Kölling, K., Krause, U., et al. (2016). Trehalose 6-phosphate coordinates organic and amino acid metabolism with carbon availability. *Plant J.* 85, 410–423. doi: 10.1111/tpj.13114

Fukayama, H., Hatch, M. D., Tamai, T., Tsuchida, H., Sudoh, S., Furbank, R. T., et al. (2003). Activity regulation and physiological impacts of maize C4-specific phosphoenolpyruvate carboxylase overproduced in transgenic rice plants. *Photosynth. Res.* 77, 227–239. doi: 10.1023/A:1025861431886

Fukayama, H., Tsuchida, H., Agarie, S., Nomure, M., Onodera, H., Ono, K., et al. (2001). Significant accumulation of C4-specific pyruvate, orthophosphate dikinase in a C3 plant, rice. *Plant Physiol.* 127, 1136–1146. doi: 10.1104/pp.010641

Furbank, R. T., and Leegood, R. C. (1984). Carbon metabolism and gas exchange in leaves of *Zea mays* L. *Planta* 162, 457–462. doi: 10.1007/BF00393458

Giuliani, R., Karki, S., Covshoff, S., Lin, H. C., Coe, R. A., Koteyeva, N. K., et al. (2019a). Knockdown of glycine decarboxylase complex alters photorespiratory carbon isotope fractionation in Oryza sativa leaves. *J. Exp. Bot.* 70, 27773–27786. doi: 10.1093/jxb/erz083

Giuliani, R., Karki, S., Covshoff, S., Lin, H. C., Coe, R. A., Koteyeva, N. K., et al. (2019b). Transgenic maize phosphoenolpyruvate carboxylase alters leaf-atmosphere CO2 and 13CO2 exchange in Oryza sativa. *Photosynth. Res.* 142, 153–167. doi: 10.1007/s11120-019-00655-4

Hatch, M. D. (1971). The C4 pathway of photosynthesis. *Biochem. J.* 125, 425–432. doi: 10.1042/bj1250425

Hatch, M. D. (1987). C4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochim. Biophys. Acta* 895, 81–106. doi: 10.1016/S0304-4173(87)80009-5

Hatch, M. D., Kagawa, T., and Craig, S. (1975). Subdivision of C4-pathway species based on differing C4 acid decarboxylating systems and ultrastructural features. *Aust. J. Plant Physiol.* 2, 111–128. doi: 10.1071/PP9750111

Hatch, M. D., Slack, C. R., and Johnson, H. S. (1967). Further studies on a new pathway of photosynthetic carbon dioxide fixation in sugar-cane and its occurrence in other plant species. *Biochem. J.* 102, 417–422. doi: 10.1042/bj1020417

Hattersley, P. W., and Watson, L. (1975). Anatomical parameters for predicting photosynthetic pathways of grass leaves: the 'maximum lateral cell count' and the 'maximum cells distant count'. *Phytomorph* 25, 325–333.

Hibberd, J. M., and Covshoff, S. (2010). The regulation of gene expression required for C4 photosynthesis. *Ann. Rev. Plant Biol.* 61, 181–207. doi: 10.1146/annurev-arplant-042809-112238

Hibberd, J. M., Sheehy, J. E., and Langdale, J. A. (2008). Using C4 photosynthesis to increase the yield of rice-rationale and feasibility. *Curr. Opin. Plant. Biol.* 11, 228–231. doi: 10.1016/j.pbi.2007.11.002

Kajala, K., Covshoff, S., Karki, S., Woodfield, H., Tolley, B. J., Dionora, M. J. A., et al. (2011). Strategies for engineering a two-celled C4 photosynthetic pathway into rice. *J. Exp. Bot.* 62, 3001–3010. doi: 10.1093/jxb/err022

Karki, S., Lin, H. C., Fanila, F. R., Abu-Jamous, B., Giuliani, R., Emms, D. M., et al. (2020). A role for neutral variation in the evolution of C4 photosynthesis. *bioRxiv [Preprint]* doi: 10.1101/2020.05.19.104299

Lin, H. C., Karki, S., Coe, R. A., Bagha, S., Khoshravesh, R., Balahadia, C. P., et al. (2016). Targeted knockdown of GDCH in rice leads to a photorespiratory deficient phenotype useful as building block for C4 rice. *Plant Cell Physiol.* 57, 919–932. doi: 10.1093/pcp/pcw033

Lunn, J. E., Feil, R., Hendriks, J. H., Gibon, Y., Morcuende, R., Osuna, D., et al. (2006). Sugar-induced increases in trehalose 6–phosphate are correlated with redox activation of ADP-glucose pyrophosphorylase and higher rates of starch synthesis in Arabidopsis thaliana. *Biochem. J.* 397, 139–148. doi: 10.1042/BJ20060083

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. doi: 10.1046/j.1365-313x.1994.06030311.x

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. doi: 10.1073/pnas.90.20.9586

Merlo, L., Geigenberger, P., Hajirezaei, M., and Stitt, M. (1993). Changes of carbohydrates, metabolites and enzyme activities in potato tubers during development, and within a single tuber along astolon-apexgradient. *J. Plant Physiol.* 142, 392–402. doi: 10.1016/S0176-1617(11)81243-5

Meyer, C. R., Rustin, P., and Wedding, R. T. (1988). A simple and accurate spectrophotometric assay for phosphoenolpyruvate carboxylase activity. *Plant Physiol.* 86, 325–328. doi: 10.1104/pp.86.2.325

Miginiac-Maslow, M., Johansson, K., Ruelland, E., Issakidis-Bourguet, E., Schepens, I., Goyer, A., et al. (2000). Light- activation of NADP-malate dehydrogenase: a highly controlled process for an optimized function. *Physiol. Plant.* 110, 322–329. doi: 10.1111/j.1399-3054.2000.1100306.x

Miyao, M. (2003). Molecular evolution and genetic engineering of C4 photosynthetic enzymes. *J. Exp. Bot.* 54, 179–189. doi: 10.1093/jxb/erg026

Miyao, M., Masumoto, C., Miyazawa, S. I., and Fukayama, H. (2011). Lessons from engineering a single-cell C4 photosynthetic pathway into rice. *J. Exp. Bot.* 62, 3021–3029. doi: 10.1093/jxb/err023

Muhaidat, R., Sage, R. F., and Dengler, N. G. (2007). Diversity of Kranz anatomy and biochemistry in C4 Eudicots. *Am. J. Bot.* 94, 362–381. doi: 10.3732/ajb.94.3.362

Nomura, M., Higuchi, T., Ishida, Y., Ohta, S., Komari, T., Imaizumi, N., et al. (2005). Differential expression pattern of C4 bundle sheath expression genes in rice, a C3 plant. *Plant and Cell Physiol.* 46, 754–761. doi: 10.1093/pcp/pci078

Randall, D. D., Miernyk, J. A., David, N. R., Gemel, J., and Luethy, M. H. (1996). "Regulation of leaf mitochondrial pyruvate dehydrogenase complex activity by reversible phosphorylation," in *Protein Phosphorylation in Plants*, eds P. R. Shewry, N. G. Halford, and R. Hooley (Clarendon: Oxford Press), 87–103.

Sage, T. L., and Sage, R. F. (2009). The functional anatomy of rice leaves: implications for refixation of photorespiratory CO2 and efforts to engineer C4 photosynthesis into rice. *Plant Cell Physiol.* 50, 756–772. doi: 10.1093/pcp/pcp033

Schuller, K. A., and Randall, D. D. (1990). Mechanism of pyruvate inhibition of plant pyruvate dehydrogenase kinase and synergism with ADP. *Arch. Biochem. Biophys.* 278, 211–216. doi: 10.1016/0003-9861(90)90250-3

Sedelnikova, O. V., Hughes, T. E., and Langdale, J. A. (2018). Understanding the Genetic Basis of C4 Kranz Anatomy with a View to Engineering C3 Crops. *Annu. Rev. Genet.* 52, 249–270. doi: 10.1146/annurev-genet-120417-031217

Sheen, J. Y., and Bogorad, L. (1987). Differential expression of C4 pathway genes in mesophyll and bundle sheath cells of greening maize leaves. *J. Biol. Chem.* 262, 11726–11730.

Szecowka, M., Heise, R., Tohge, T., Nunes-Nesi, A., Vosloh, D., Huege, J., et al. (2013). Metabolic fluxes in an illuminated *Arabidopsis rosette*. *Plant Cell* 25, 694–714. doi: 10.1105/tpc.112.106989

Taniguchi, Y., Ohkawa, H., Masumoto, C., Fukuda, T., Tesshu, T., Lee, K., et al. (2008). Overproduction of C4 photosynthetic enzymes in transgenic rice plants: an approach to introduce the C4-like photosynthetic pathway into rice. *J. Exp. Bot.* 59, 1799–1809. doi: 10.1093/jxb/ern016

Tcherkez, G., Mahé, A., Gauthier, P., Mauve, C., Gout, E., Bligny, R., et al. (2009). In folio respiratory fluxomics revealed by 13C isotopic labeling and H/D isotope effects highlight the noncyclic nature of the tricarboxylic acid 'cycle' in illuminated leaves. *Plant Physiol.* 151, 620–630. doi: 10.1104/pp.109.142976

Tovar-Méndez, A., Miernyk, J. A., and Randall, D. D. (2003). Regulation of pyruvate dehydrogenase complex activity in plant cells. *Eur. J. Biochem.* 270, 1043–1049. doi: 10.1046/j.1432-1033.2003.03469.x

Tsuchida, H., Tamai, T., Fukayama, H., Agarie, S., Nomura, M., Onodera, H., et al. (2001). High level expression of C4-specific NADP-malic enzyme in leaves and impairment of photoautotrophic growth of a C3 plant, rice. *Plant Cell Physiol.* 42, 138–145. doi: 10.1093/pcp/pce013

Ueno, O., Kawano, Y., Wakayama, M., and Takeda, T. (2006). Leaf vascular systems in C3 and C4 grasses: a two-dimensional analysis. *Annu. Bot.* 97, 611–621. doi: 10.1093/aob/mcl010

Ueno, Y., Hata, S., and Izui, K. (1997). Regulatory phosphorylation of plant phosphoenolpyruvate carboxylase: role of a conserved basic residue upstream of the phosphorylation site. *FEBS Lett.* 417, 57–60. doi: 10.1016/S0014-5793(97)01254-4

Vidal, J., and Chollet, R. (1997). Regulatory phosphorylation of C4 PEP carboxylase. *Trends Plant Sci.* 2, 230–237. doi: 10.1016/S1360-1385(97)89548-9

Vogan, P. J., Frohlich, M. W., and Sage, R. F. (2007). The functional significance of C3-C4 intermediate traits in Heliotropium L (Boraginaceae): gas exchange perspectives. *Plant Cell Environ.* 30, 1337–1345. doi: 10.1111/j.1365-3040.2007.01706.x

von Caemmerer, S., and Furbank, R. T. (2003). The C4 pathway: an efficient CO2 pump. *Photosynth. Res.* 77, 191–207. doi: 10.1023/A:1025830019591

von Caemmerer, S., Quick, W. P., and Furbank, R. T. (2012). The development of C4 rice: current progress and future challenges. *Science* 336, 1671–1672. doi: 10.1126/science.1220177

Wang, P., Khoshravesh, R., Karki, S., Tapia, R., Balahadia, C. P., Bandyopadhyay, A., et al. (2017). Re-creation of a key step in the evolutionary switch from C3 to C4 leaf anatomy. *Curr. Biol.* 27, 3278–3287. doi: 10.1016/j.cub.2017.09.040

Wang, Q., Zhang, Q., Fan, D., and Lu, C. (2006). Photosynthetic light and CO2 utilization and C4 traits of two novel super-rice hybrids. *J. Plant Physiol.* 163, 529–537. doi: 10.1016/j.jplph.2005.04.035

Weber, A. P., and von Caemmerer, S. (2010). Plastid transport and metabolism of C3 and C4 plants — comparative analysis and possible biotechnological exploitation. *Curr. Opin. Plant Biol.* 13, 256–264. doi: 10.1016/j.pbi.2010.01.007

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 16 October 2020
doi: 10.3389/fpls.2020.570436

Insights Into the Regulation of the Expression Pattern of Calvin-Benson-Bassham Cycle Enzymes in C_3 and C_4 Grasses

Chidi Afamefule and Christine A. Raines**

School of Life Sciences, University of Essex, Colchester, United Kingdom

C_4 photosynthesis is characterized by the compartmentalization of the processes of atmospheric uptake of CO_2 and its conversion into carbohydrate between mesophyll and bundle-sheath cells. As a result, most of the enzymes participating in the Calvin-Benson-Bassham (CBB) cycle, including RubisCO, are highly expressed in bundle-sheath cells. There is evidence that changes in the regulatory sequences of *RubisCO* contribute to its bundle-sheath-specific expression, however, little is known about how the spatial-expression pattern of other CBB cycle enzymes is regulated. In this study, we use a computational approach to scan for transcription factor binding sites in the regulatory regions of the genes encoding CBB cycle enzymes, SBPase, FBPase, PRK, and GAPDH-B, of C_3 and C_4 grasses. We identified potential *cis*-regulatory elements present in each of the genes studied here, regardless of the photosynthetic path used by the plant. The trans-acting factors that bind these elements have been validated in *A. thaliana* and might regulate the expression of the genes encoding CBB cycle enzymes. In addition, we also found C_4-specific transcription factor binding sites in the genes encoding CBB cycle enzymes that could potentially contribute to the pathway-specific regulation of gene expression. These results provide a foundation for the functional analysis of the differences in regulation of genes encoding CBB cycle enzymes between C_3 and C_4 grasses.

Keywords: C_4 photosynthesis, gene expression regulation, *cis*-regulatory elements, transcription factor binding sites, Calvin-Benson-Bassham cycle

OPEN ACCESS

Edited by:
Martha Ludwig,
University of Western Australia, Australia

Reviewed by:
Nelson J. M. Saibo,
New University of Lisbon, Portugal
Thomas D. Sharkey,
Michigan State University, United States

****Correspondence:***
Chidi Afamefule
chidi.afamefule@essex.ac.uk
Christine A. Raines
rainc@essex.ac.uk

Specialty section:
This article was submitted to Plant Physiology, a section of the journal Frontiers in Plant Science

Received: *07 June 2020*
Accepted: *23 September 2020*
Published: *16 October 2020*

Citation:
Afamefule C and Raines CA (2020) Insights Into the Regulation of the Expression Pattern of Calvin-Benson-Bassham Cycle Enzymes in C_3 and C_4 Grasses. Front. Plant Sci. 11:570436. doi: 10.3389/fpls.2020.570436

INTRODUCTION

C_4 plants achieve higher photosynthetic efficiency by concentrating CO_2 around RubisCO. In contrast with enzymes participating in C_3 photosynthesis, C_4-enzymes are compartmentalized to specific cell types, namely mesophyll (M) and bundle-sheath (BS) cells. Enzymes enriched in M cells include phosphoenolpyruvate carboxylase (PEPC) and pyruvate orthophosphate dikinase (Ppdk), whereas decarboxylating malic enzymes (NAD or NADP-Me) and RubisCO are enriched in the BS cells (Sheen and Bogorad, 1987; Hibberd and Covshoff, 2010; Berry et al., 2011).

During C_4 evolution a change in localization of the enzymes involved in CO_2 assimilation resulted in the compartmentalization of these reactions in either the M or BS cell types. A number of regulatory elements conferring a M or BS specific expression pattern have been identified in the

regulatory sequences of the genes encoding PEPC, Ppdk; or NADP-ME, NAD-ME, and RubisCO (Nomura et al., 2000; Berry et al., 2011; Williams et al., 2016; Reyna-Llorens et al., 2018). To further interrogate those regulatory elements, a combination of comparative transcriptomics to identify differential expression of genes (Bräutigam et al., 2011; Aubry et al., 2014; Xu et al., 2016) and DNAse-seq to map differences in open chromatin regions between M and BS cells (Burgess et al., 2019) have been used. These studies have led to the identification of putative *cis*-regulatory elements and the trans-acting transcription factors binding to those elements, and have shown that the motifs conferring differences in expression in the C_4 species have been recruited from pre-existing sequences in C_3 species, rather than being generated *de novo* during the evolution of the C_4 condition (Niklaus and Kelly, 2019).

Calvin Benson-Bassham (CBB) cycle enzymes, including RubisCO, are expressed in both C_3 and C_4 species. Similar to RubisCO, most of the CBB cycle enzymes are enriched in BS cells in C_4 species (Sheen and Bogorad, 1987; John et al., 2014; Rao et al., 2016). Unlike RubisCO, little is known about the changes in the regulatory sequences of the other 10 genes encoding CCB cycle enzymes that enable such compartmentalization, limiting our ability to develop strategies to manipulate this pathway to improve photosynthetic efficiency. Here, we present a bioinformatics analysis of the regulatory sequences of genes encoding CBB cycle enzymes with the aim of identifying regulatory elements that are common to C_3 and C_4 species, or C_4-specific regulatory elements that control photosynthesis and contribute to C_4 compartmentalization. We selected four of the CBB cycle enzymes known to be redox-regulated by the ferredoxin/thioredoxin (Fd/TRX) system (Michelet et al., 2013) and that function exclusively in the CBB cycle: SBPase, FBPase (chloroplastic variant), PRK and GAPDH-B. Given the numerous independent origins of C_4 photosynthesis that might have led to parallel evolution of *cis*-regulatory elements (Sage et al., 2012), in this paper we focus on a small subset of eight grasses from the Poaceae family whose genomes have been sequenced and annotated.

In this study we have identified putative regulatory elements that are common in both C_3 and C_4 species as well as C_4-specific elements. We have also used existing data to explore the expression patterns of the trans-acting factors that have been shown or proposed to bind to these elements, suggesting a possible role in the compartmentalization of CBB cycle enzymes in C_4 plants. The results presented here provide the basis for future functional studies.

MATERIALS AND METHODS

DNA Sequences

Genomic sequences encoding CBB cycle enzymes of *Oryza sativa* (Ouyang et al., 2007), *Hordeum vulgare* (Beier et al., 2017; Mascher et al., 2017), *Brachypodium distachyon* (International Brachypodium Initiative, 2010), *Zea mays* (Schnable et al., 2009; Hirsch et al., 2016), *Sorghum bicolor* (McCormick et al., 2018), *Setaria viridis* (v2.1, DOE-JGI)[1] and *Panicum virgatum* (v1.0, DOE-JGI, see footnote) were obtained from Phytozome12 (Goodstein et al., 2011). *Arabidopsis thaliana* genes (AT3G55800—*SBPase*, AT3G54050—*chlFBPase*, AT1G32060—*PRK*, and AT1G42970—*GAPDHB*) were used to identify orthologs in every species. For the genomic sequences encoding CBB cycle enzymes of *Dichanthelium oligosanthes* (Studer et al., 2016), the *A. thaliana* coding sequences were aligned against the *D. oligosanthes* genome using BLAST (Altschul et al., 1990) to find orthologous genes. Sequences used are included in **Supplementary Material**.

Motif Prediction in Conserved Non-coding Sequences (CNS)

Genomic sequences were aligned using mVISTA (Frazer et al., 2004) and aligned CNSs were used as input for motif prediction using MEME (v5.1.1; Bailey et al., 2009). Motif site distribution was set to zoops and maximum motif width to the size of the shorter CNS. Predicted motifs were used as input in FIMO (Grant et al., 2011) to scan the regulatory sequences of orthologous genes in other species.

Motif Scanning of Genomic Sequences

A collection of 529 plant transcription factor motifs validated in *A. thaliana* (O'Malley et al., 2016) were used to scan for motifs using FIMO (Grant et al., 2011) with default parameters.

Data Processing and Visualization

Data processing and visualization were performed using R 3.6.0 (R Core Team, 2019). The dplyr package (Wickham et al., 2019) was used to filter the identified motifs by $q < 0.05$, genomic feature, and by species. The UpSetR package (Gehlenborg, 2019) was used to generate **Figure 2A** showing all possible interactions; and the ggplot2 (Wickham, 2016) and the gggenes packages (Wilkins, 2019) were used to generate **Figure 2B**.

Transcriptomics Analysis

Transcriptomic data from RNAseq experiments in which mesophyll and bundle sheath cells were separated in *P. virgatum* (Rao et al., 2016), *S. viridis* (John et al., 2014), *Panicum hallii* (Washburn et al., 2017), and *Setaria italica* (Washburn et al., 2017) were obtained from NCBI (BioProject accession numbers: PRJNA293441, PRJEB5074, PRJNA475365). A classification-based quantification was performed using kallisto (Bray et al., 2016) with the transcriptomes and genome annotation obtained from Phytozome 12 (Goodstein et al., 2011; Bennetzen et al., 2012). In short, a kallisto index was built with the reference transcriptome of each species, and kallisto quant was used to quantify abundance of pair-end reads with default parameters. Differential expression analysis was performed with R packages DESeq (Anders and Huber, 2010) using estimateSizeFactors, estimateispersions and nbinomTest functions; DESeq2 (Love et al., 2014) using DESeq function, and edgeR (Robinson et al., 2009; McCarthy et al., 2012) using estimateCommonDisp,

[1] http://phytozome.jgi.doe.gov/

estimateTagwiseDisp and exactTest functions. *P*-values were adjusted with the Hochberg method in the three analyses, and only genes with adjusted $p < 0.05$ in at least one of the analyses were included in **Table 1**. Parallel (Tange, 2018) was used at every step to run jobs in parallel.

To construct **Supplementary Table 1**, we used the 57 *A. thaliana* transcription factors that have been shown to bind the identified transcription factor binding sites (TFBS, 50 shared by different orthologous genes, **Figure 2**; plus 7 absent from C_3 or C_4 species, **Supplementary Figure S5**). We identify the orthologous genes in grass species and evaluate their enrichment in M or BS cells using publicly available transcriptomic data for *S. viridis* (John et al., 2014), *S. italica* (Washburn et al., 2017), *P. virgatum* (Rao et al., 2016), *P. halli* (Washburn et al., 2017), *Z. mays* (Chang et al., 2012), and *S. bicolor* (Döring et al., 2016). All these databases separate M and BS cells from whole leaves. We identified 10 orthologous genes significantly enriched (adj. $p < 0.05$) in *P. virgatum*, which corresponded to 8 genes in *A. thaliana*. For *P. halli* we identified 21 orthologs corresponding to 11 *A. thaliana* genes. For *S. viridis* we identified 53 orthologs corresponding to 26 *A. thaliana* genes. For *S. italica* we identified 46 orthologs corresponding to 22 *A. thaliana* genes. For *Z. mays* we identified 10 orthologs corresponding to 4 *A. thaliana* genes. For *S. bicolor* we identified 2 orthologs corresponding to 2 *A. thaliana* genes. In **Table 1**, we only included the *A. thaliana* genes for which the log2 fold was at least 1, and with consistent data from at least two species. We also removed *Z. mays* and *S. bicolor* orthologous genes as their transcriptomic data did not add any information on the *A. thaliana* genes included on **Table 1**.

RESULTS

To account for the numerous independent origins of C_4 photosynthesis, we focus on a small subset of eight grasses: *Oryza sativa*, *Hordeum vulgare*, *Brachypodium distachyon*, *Dichanthelium oligosanthes*, *Zea mays*, *Sorghum bicolor*, *Panicum virgatum*, and *Setaria viridis*. All of these plant species belong to the Poaceae family and shared a common ancestor around 50 million years ago. *O. sativa*, *H. vulgare*, *B. distachyon,* and *D. oligosanthes* perform C_3 photosynthesis, whereas *Z. mays, S. bicolor, P. virgatum,* and *S. viridis* perform C_4 photosynthesis. Notably, *D. oligosanthes* belongs to the PACMAD clade (**Figure 1**), to which all selected C_4 species belong, and shares a common ancestor with them around 15 million years ago (Studer et al., 2016). To identify conserved regulatory regions in genes encoding CBB cycle enzymes of C_3 and C_4 grasses, we aligned each gene against its orthologous gene in a representative C_3 species (*B. distachyon*; **Figure 1A** and **Supplementary Figures S1A**, **S2A**, **S3A**) and against its orthologous gene in a representative C_4 species (*S. bicolor*; **Figure 1C** and **Supplementary Figures S1C**, **S2C**, **S3B**). The genomic sequence including potential promoters [2000 base pair (bp)] upstream from the annotated transcription start site (or start codon otherwise) and potential terminators (1,000 bp downstream from the end of 3′UTR or stop codon) was used to allow for the identification of putative regulatory regions outside coding sequences. Regions showing between 50 and 100% identity were plotted and conserved regions with over 70% identity were colored depending on the genomic feature (**Figures 1A,C**; coding sequences in purple, untranslated regions [UTRs] in cyan, and intergenic regions and introns in pink) As expected, most of the coding sequences were conserved among all orthologous genes, whereas only parts of the introns and intergenic sequences showed over 70% identity. We defined those regions as conserved non-coding sequences (CNS). For *SBPase*, we identified one CNS located at the last intron of most orthologs (**Figures 1A,C**), and two CNSs found only in *SBPase* orthologous genes from PACMAD species (C_4 species + *D. oligosanthes*; **Figure 1C**). In addition, we found one CNS located at the 5′ intergenic region of all *PRK* genes (**Supplementary Figure S1**), and two CNSs located at the 5′ intergenic region of *FBPase* genes from PACMAD species (**Supplementary Figure S2**). To further characterize these CNSs, they were subjected to motif prediction using MEME (Bailey et al., 2009), which generated a position weight matrix for the predicted motifs (**Figure 1B** and **Supplementary Figures S1B**, **S2B**). We used these motifs to scan the orthologous genes of other species, and identified the *PRK* CNS in the intergenic regions of *PRK* orthologs in non-grasses species (**Supplementary Dataset 1**). These results indicate that there are conserved potential *cis*-regulatory sequences shared between C_3 and C_4 species. However, this alignment approach is based on sequence identity over at least 50 bp; so it was possible that smaller motifs, such as transcription factor binding sites (TFBS) could have been disregarded.

To evaluate the presence of TFBS in the regulatory regions of genes encoding CBB cycle enzymes, a dataset containing validated TFBS in *Arabidopsis thaliana* (O'Malley et al., 2016) was used to scan the putative regulatory sequences (intergenic regions, untranslated regions, and introns) of orthologous genes, i.e., *SBPase* orthologs across the subset of eight grass species were scanned at the same time. We first determined the *A. thaliana* TFBS shared between orthologous genes, and used those to compare between the genes encoding the selected four CBB cycle enzymes (**Figure 2A**). This way, we identified one TFBS present in all of the potential regulatory sequences (common_CBB, in **Figure 2A**) that was bound by VRN1 in *A. thaliana*. This TFBS was also identified it in the putative regulatory regions of genes encoding photorespiratory (*GDCH*) and housekeeping proteins (*CBP20*) (**Supplementary Dataset 2**), suggesting that it might play a regulatory role not limited to photosynthetic genes. We also identified 13 putative TFBS shared between *SBPase, FBPase,* and *GAPDHB* orthologous genes (common_SFG), 9 TFBS shared between *GAPDHB* and *SBPase* orthologous genes (common_GS), one shared between *GAPDHB* and *PRK* orthologous genes (common_GP), and one TFBS shared between *GAPDHB* and *FBPase* orthologous genes (common_GF). In addition, 17, 2, and 6 putative TFBS were shared between *GAPDHB* orthologs (common_GAPDHB), *FBPase* orthologs (common_FBP), and *PRK* orthologs (common_PRK); but not between any other group of orthologous genes. Notably, these common sequences can be found in potential regulatory sequences of other

TABLE 1 | Differential expression of trans-acting factors binding putative TFBS in bundle sheath and mesophyll cells of C_4 species.

Transcription factor	*A. thaliana* name	*Panicum virgatum* name	log2 FC	*Setaria viridis* name	log2 FC	*Panicum hallii* name	log2 FC	*Setaria italica* name	log2 FC	Group
VRN1	AT3G18990	Pavir.8NG077400.2	3.6	Sevir.8G068300.1	**1.4**					cCBB
LOB	AT5G63090					Pahal.5G488600.2	3.2	Seita.5G119400.1	4.6	c34G
OBP3	AT3G55370			Sevir.3G064900.1	**6.4**	Pahal.3G092800.1	3.5	Seita.3G064100.1	**5.4**	c34P
				Sevir.3G064900.2	**6.6**	Pahal.3G092800.2	**7.9**	Seita.9G033400.1	**5.8**	
				Sevir.3G064900.3	**4.9**	Pahal.3G092800.3	**3.6**	Seita.9G452000.1	**7.2**	
				Sevir.9G032600.1	**6.5**	Pahal.9G030900.1	**5.4**			
				Sevir.9G455900.1	**7.9**	Pahal.9G513900.1	**3.1**			
				Sevir.9G455900.2	**7.0**	Pahal.9G513900.2	**9.8**			
At5g66940	At5g66940			Sevir.3G015900.1	2.2	Pahal.7G338900.1	2.8	Seita.3G014900.1	**5.1**	C3AS
AREB3	AT3G56850	Pavir.5KG593700.1	**3.3**	Sevir.9G425100.2	**5.1**					C4AP
AT3G12130	AT3G12130			Sevir.4G224600.1	–0.3	Pahal.1G071400.1	–0.9	Seita.3G029300.2	–0.9	C3AP
				Sevir.4G224600.2	–1.7	Pahal.1G071400.3	–7.3	Seita.4G214800.1	–0.5	
ERF5	AT5G47230			Sevir.1G261900.1	–1.1	Pahal.9G383200.1	**–3.0**	Seita.1G257600.1	**–1.7**	cSFG
AS2	AT1G65620			Sevir.3G246200.2	–5.4			Seita.5G408700.1	**–3.4**	c34G
					–3.4					
ERF1	AT3G23240			Sevir.8G100900.1	**–5.3**	Pahal.2G139200.1	**–3.4**	Seita.2G138400.1	**–2.9**	c34G
				Sevir.9G504700.1	**–3.7**	Pahal.8G262800.1	1.9	Seita.9G500100.1	–1.8	
ERF9	AT5G44210	Pavir.5NG539500.1	–2.4	Sevir.3G196300.1	1.2			Seita.5G348000.1	**–1.4**	c34G
				Sevir.5G352700.1	**–2.5**					
ERF15	AT2G31230			Sevir.2G118100.1	**–4.5**	Pahal.8G107700.1	–1.5	Seita.2G112200.1	**–2.1**	c34G
				Sevir.8G182200.1	5.7			Seita.8G173100.1	**7.6**	
								Seita.8G237900.1	**4.3**	
TCX2	AT4G14770			Sevir.2G055300.3	–6.2	Pahal.3G490800.1	0.7	Seita.2G050700.1	**–3.6**	C3AF
				Sevir.3G398500.1	0.7	Pahal.9G158300.1	**7.0**	Seita.3G382000.1	–0.7	
				Sevir.9G159400.4	**5.0**			Seita.9G161100.3	5.1	
				Sevir.9G159400.6	**4.8**			Seita.9G161100.2	7.5	
				Sevir.9G159400.1	2.7					
				Sevir.9G159400.7	**4.1**					
ERF73	AT1G72360	Pavir.9NG798900.1	5.5	Sevir.2G400300.5	**2.9**	Pahal.2G447100.1	**1.8**	Seita.2G390000.1	**3.6**	cSFG
				Sevir.9G520900.1	**–1.4**	Pahal.2G447100.2	**2.2**	Seita.2G390000.3	**1.2**	

(Continued)

TABLE 1 | Continued

Transcription factor	*A. thaliana* name	*Panicum virgatum* name	log2 FC	*Setaria viridis* name	log2 FC	*Panicum hallii* name	log2 FC	*Setaria italica* name	log2 FC	Group
				Sevir.9G521200.2	**–4.1**	Pahal.9G580000.1	**–1.4**	Seita.9G516500.1	**–1.9**	
						Pahal.9G580100.1	**–2.1**	Seita.9G516600.1	3.0	
						Pahal.9G580200.1	**–2.1**	Seita.9G516700.1	**–4.2**	
						Pahal.9G580200.2	**–2.1**	Seita.9G516700.2	**–4.1**	
								Seita.9G516800.1	**–3.6**	
								Seita.9G516800.2	**3.4**	
bZIP16	AT2G35530			Sevir.2G084800.1	0.8	Pahal.1G023000.4	**1.1**	Seita.3G090500.2	**1.2**	C4AP
				Sevir.3G092500.1	–0.7	Pahal.1G023000.5	–0.7	Seita.3G090500.3	–9.7	
						Pahal.1G023000.6	**1.6**	Seita.9G474400.1	0.3	
						Pahal.3G063900.2	0.6			
						Pahal.9G536900.1	0.5			

The expression level of each Arabidopsis thaliana transcription factor ortholog was obtained from RNAseq data of Panicum virgatum, Setaria viridis, Panicum hallii and Setaria italica. Every A. thaliana transcription factor had one or more orthologs in the target species. The log2 fold change between bundle sheath and mesophyll cells was calculated using three different methods to evaluate the difference in expression level between cell types (see "Materials and Methods" section), and only genes with adjusted $p < 0.05$ in at least one of the methods were included. log2 fold values in bold indicate that they are significant using the three methods. Transcription factors were classified as putative activators (in green) if they were enriched in bundle sheath cells, or putative repressors (in gold) if they were enriched in mesophyll cells. Transcription factors in blue showed enrichment in both bundle sheath and mesophyll cells among different orthologs. Last column indicates the group to which each transcription factor belongs (see ***Figure 2****): common_CBB (cCBB), common_SFG (cSFG), common_GAPDHB (c34G), common_PRK (c34P), C3-Absent_SBPase (C3AS), C3-Absent_PRK (C3AP), and C4-Absent_PRK (C4AP).*

orthologous genes from some but not in all of the species in the study. The fact that all the *A. thaliana* TFBS found in *SBPase* orthologs are shared with other genes (common_SFG, common_GS) whereas most of the TFBS found in *PRK* orthologs are not shared (common_PRK) suggests different mechanisms for the regulation of the expression of the genes encoding CBB cycle enzymes. Most of the trans-acting factors binding to the identified TFBS belong to the Apetala2/Ethylene-Response-Factor (AP2/ERF) family, and often recognize similar binding sites. A comparison between the location of the *A. thaliana* TFBS at the putative regulatory regions of the orthologs in the selected species (**Figure 2B** and **Supplementary Figure S4**) revealed that the TFBS tend to cluster together in discrete regions of the putative regulatory sequences, although the genomic coordinates of these clusters change between species.

We used a similar approach to identify C_4-specific TFBS contributing to the difference in expression pattern between C_3 and C_4 species. After scanning together orthologous genes (i.e., all *SBPase* orthologous genes with the *A. thaliana* validated TFBS), we selected TFBS absent from the putative regulatory regions of genes encoding C_3-enzymes (**Supplementary Figure S5**). Using this approach (choosing absent motifs from genes encoding C_3 enzymes rather than present motifs in all genes encoding C_4 enzymes), it was possible to account for the multiple independent origins of C_4 photosynthesis. Three TFBS were found absent from *SBPase* C_3-genes, bound by At5g66940, BZR1, and CEJ1 in *A. thaliana*; one absent from *FBPase* C_3-genes (bound by TCX2) and one absent from *PRK* C_3-genes (bound by At3g12130). In addition, using the same approach to identify TFBS absent from genes encoding C_4 enzymes revealed two C_3-specific motifs in the 5′ intergenic region of *PRK* (bound by AREB3 and bZIP16). These results suggest that there are C_3- and C_4-specific TFBS that might contribute to the compartmentalization of C_4 CBB cycle enzymes.

To further understand how the identified TFBS might regulate the expression pattern of CBB cycle enzymes, we obtained the transcriptomic data from a collection of RNA-seq experiments on C_4 species where samples were taken separately from mesophyll and bundle sheath cells (John et al., 2014; Rao et al., 2016; Washburn et al., 2017), and assessed the expression pattern of the trans-acting factors. Despite the complexities of using data from different experiments, and the limited validation of the interaction between trans-acting factors and the identified TFBS (i.e., only validated in the C_3 species *A. thaliana*), we identified ten trans-acting factors differentially enriched in M and BS cell types (**Table 1**). These trans-acting factors were the orthologs of the validated trans-acting factors in *A. thaliana*, and could be classified into three categories in regard to BS-specific enrichment (**Table 1**): (1) putative activators, if all the orthologs were consistently enriched in BS over M cells, such as the orthologs of *At5g66940*, whose TFBS are only found in *SBPase* orthologs of C_4 species; (2) putative repressors, if all the orthologs were consistently enriched in M over BS cells, for example the orthologs of *At3g12130*, whose TFBS are only found in *PRK* orthologs of C_4 species; (3) broad regulators, if enrichment of orthologous genes was inconsistent within species (some were enriched in BS over M cells while others

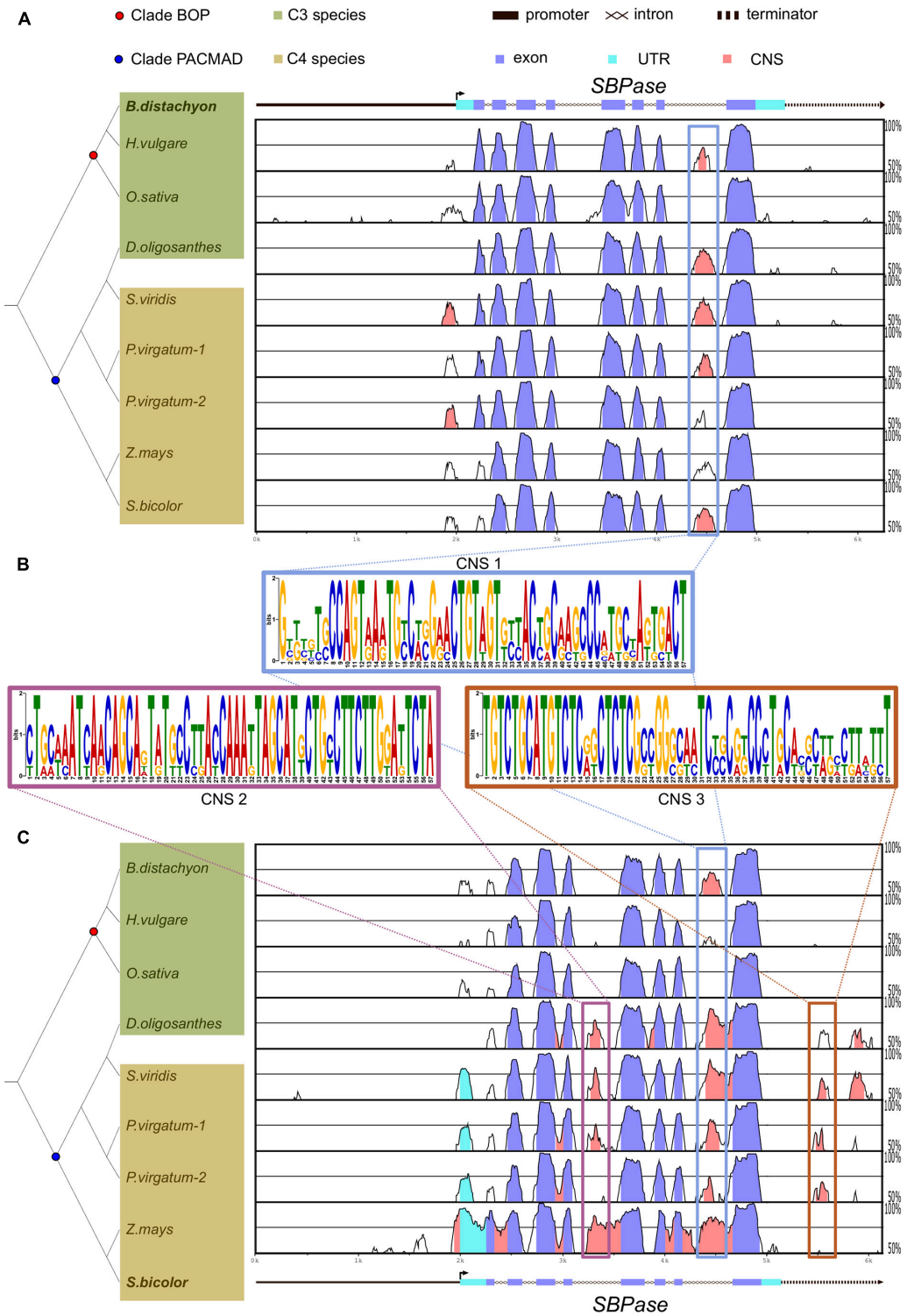

FIGURE 1 | *SBPase* coding sequence is highly conserved among C_3 and C_4 grasses in comparison to putative regulatory regions. **(A,C)** mVISTA plots of *Brachypodium distachyon* **(A)** and *Sorghum bicolor* **(C)** *SBPase* aligned to *SBPase* orthologs in C_3 and C_4 grasses. Genomic region includes approximately 2 kb upstream from the transcription start site and 1 kb after the end of the 3′ untranslated region (UTR). UTRs, exons, and introns are annotated. The vertical line with a small perpendicular arrow indicates the transcription start site and the arrowhead the orientation of the gene. The graph shows sequences with 50–100% identity and regions with > 70% identity within 50 base pairs are highlighted in purple if they are located in exons, in cyan if they are located in UTRs, or in pink if they are located outside exons or UTRs. Boxes highlight conserved non-coding sequences (CNSs), and the predicted position weight matrix for each conserved sequence is included **(B)**. On the left side, the phylogenetic relationship between C_3 (in green) and C_4 (in brown) grasses is shown. Common ancestor of BOP clade and PACMAD clade species are shown as a red and as a blue dot, respectively. Note that *Dichanthelium oligosanthes* is a C_3 species within the PACMAD.

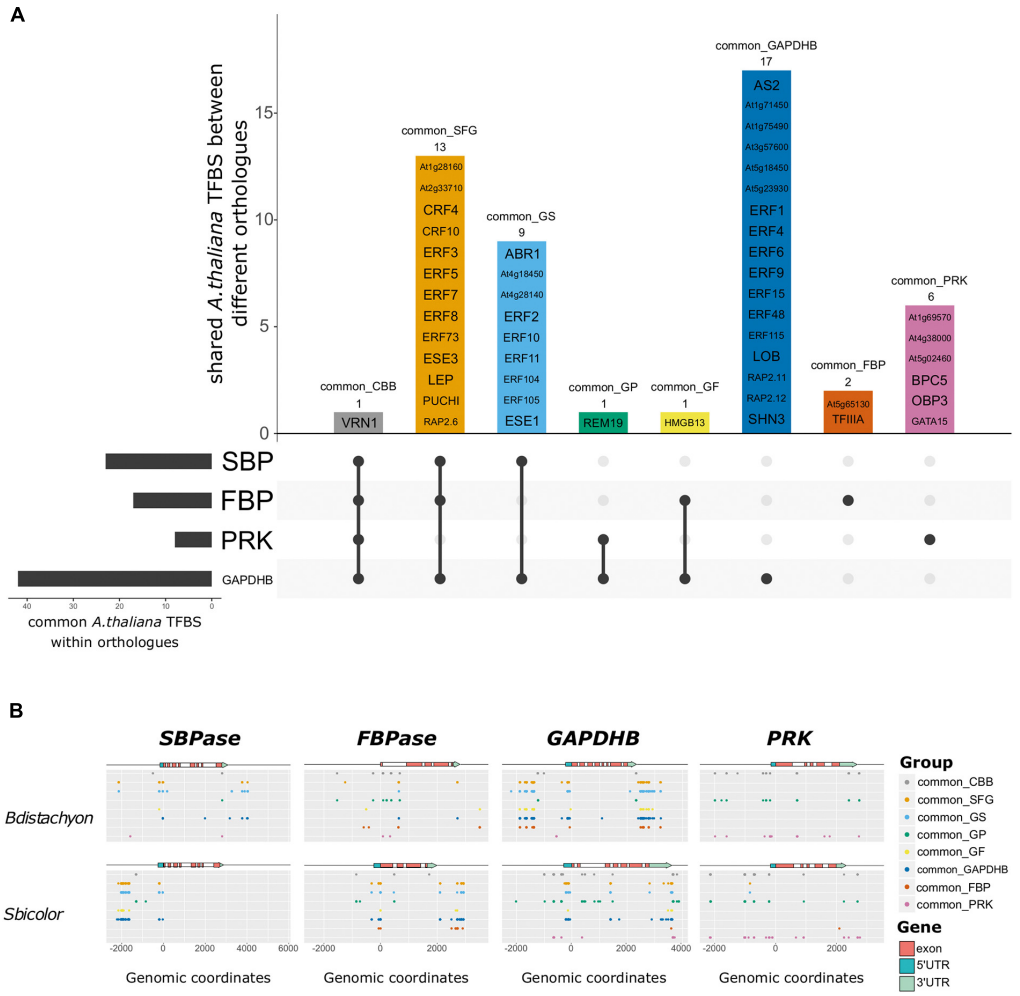

FIGURE 2 | *Arabidopsis thaliana* transcription factor binding sites (TFBS) identified in the potential regulatory sequences of genes encoding C_3 and C_4 Calvin-Benson-Bassham (CBB) cycle enzymes. **(A)** Upset plot showing the identified *A. thaliana* TFBS and in which orthologous genes they are found. Horizontal bars represent the number of common motifs identified within orthologs, vertical bars represent the motifs shared between different orthologous genes, as indicated by the dots below. The name of the *A. thaliana* TFBS is included inside the vertical bars, and the number of motifs as well as the name of the gene group are indicated above. There is only one common motif shared across all orthologs (common_CBB: VRN1, gray bar), and most of the common motifs identified in *GAPDHB* are not common in the genes encoding other enzymes (common_GADPHB). In addition, many motifs are shared between *GAPDHB*, *FBPase*, and *SBPase* (common_SFG), whereas most *PRK* motifs are not shared with the genes encoding other enzymes (common_PRK). **(B)** Localization of each gene group in the genomic region around genes encoding CBB cycle enzymes in *Brachypodium distachyon* and *Sorghum bicolor*. The *x*-axis corresponds to the genomic coordinates with the start codon corresponding to the + 1 position. The colored arrow represents the gene structure with UTRs in blue, exons in red, and potential promoter and terminator as a black line. The dots represent the genomic coordinates of each of the motifs within each gene group. Different gene groups are separated along the *y*-axis. Despite being comprise by the same TFBS, the distribution of the dots changes between species. Notably, TFBS can be found at multiple coordinates in the same gene. Most of the trans-acting factors binding to the identified *A. thaliana* TFBS belong to the same family, and often bind to similar sequences. In fact, TFBS tend to cluster in discrete regions that might play a role in the regulation of the expression of the corresponding gene.

were enriched in M over BS cells), such as the orthologs of *TCX2*, which are found enriched in both M and BS cells, and whose TFBS are only found in *FBPase* orthologs of C_4 species.

DISCUSSION

In this study, we have used publicly available data to analyze putative regulatory regions of genes encoding a selected subset of CBB cycle enzymes (SBPase, FBPase, PRK, and GAPDHB) in C_3 and C_4 species. We used two different approaches to identify potential regulatory elements that might contribute to the compartmentalization of CBB enzymes in C_4 species. The alignment of the genomic regions of the orthologs encoding the selected CBB cycle enzymes allowed us to identify conserved non-coding sequences (CNSs) shared by C_3 and C_4 orthologous genes, whereas the scanning of putative regulatory regions with TFBS validated in *A. thaliana*, allowed us to identify putative C_4-specific regulatory elements. The results presented here provide

new information on putative regulatory elements of the genes encoding SBPase, FBPase, PRK, and GAPDHB in both C_3 and C_4 species and although we do not provide experimental evidence in this paper the results form the basis for future functional studies.

The alignment of the genomic regions of orthologous CBB genes revealed a number of CNSs shared between C_3 and C_4 species. We identified a highly conserved sequence in the 5′ intergenic region of every *PRK* gene (**Supplementary Figure S1**). This CNS stands out because of its length (113 bp) and the level of conservation, as it can be found in C_3 species even outside of Poaceae (**Supplementary Dataset S1**). These attributes suggest that this region could have contributed to the regulation of *PRK* expression throughout evolutionary history. In contrast, the CNS identified in the last intron of *SBPase* orthologous genes (**Figure 1**) was only found in species belonging to Poaceae, suggesting that the possible contribution to the regulation of *SBPase* genes is limited to Poaceae species (**Supplementary Dataset S1**). Nevertheless, the location of this conserved region highlights the relevance of searching for regulatory elements outside of the up- and down-stream non-coding sequences of genes (Rose, 2019). Additionally, we also identified CNSs conserved only within the more closely related species of the PACMAD clade (*D. oligosanthes*, *Z. mays, S. bicolor, P. virgatum*, and *S. viridis*) but not within the more distant related species of the BOP clade (*O. sativa*, *H. vulgare*, and *B. distachyon*; **Figure 1** and **Supplementary Figure S2**). The fact that these CNSs are only shared between the species of the PACMAD clade, including *D. oligosanthes* which performs C_3 photosynthesis, suggests that these CNSs do not play a role in C_4 compartmentalization and instead they are a result of shared evolutionary history. However, the significance of the contribution of these conserved regions to the levels or patterns of expression of these genes remains to be elucidated experimentally.

Using a different approach based on validated TFBS and their trans-acting factors in *A. thaliana*, we identified putative (i.e., non-validated in grasses) TFBS shared by the genes encoding CBB cycle enzymes in both C_3 and C_4 species (**Figure 2A**), as well as C_4-specific (C_3-absent) putative TFBS (**Supplementary Figure S5**). We found three putative TFBS absent from *SBPase* C_3-genes, one absent from *FBPase* C_3-genes, and one absent from *PRK* C_3-genes. The identification of *A. thaliana* TFBS in genes encoding C_4 CBB cycle enzymes supports the hypothesis that C_4 genes co-opted regulatory elements of C_3 genes to establish their restricted expression pattern (Brown et al., 2011; Xu et al., 2016; Borba et al., 2018; Reyna-Llorens et al., 2018). Notably, we did not identify any C_3- or C_4-specific putative TFBS in the regulatory regions of *GAPDHB* orthologs (**Supplementary Figure S5**) but found more shared TFBS between C_3 and C_4 *GAPDHB* orthologs (**Figure 2A**). Despite being expressed in BS cells, which should allow for CBB cycle function in those cells, *GAPDHB* is enriched in M cells (Majeran et al., 2005; Rao et al., 2016). The lack of C_4-specific putative TFBS in *GAPDHB* regulatory regions suggests that its expression might be regulated similarly in both C_3 and C_4 plants. Most of the identified TFBS are recognized by members of the AP2/ERF family in *A. thaliana*, which supports the results of a recent study in which this family of TFBS was enriched in the regulatory regions of C_4 photosynthetic genes (Burgess et al., 2019). We realized that these putative TFBS were often quite similar and cluster together at specific locations in the genome and this warrants further investigation to explore the functional significance. Despite the similarities, we only identified one TFBS, bound by VRN1 in *A. thaliana*, in the putative regulatory regions of every gene selected for this study, but its presence in other non-photosynthetic genes indicates that VRN1 is unlikely to be exclusive to the regulation of the expression of genes encoding CBB cycle enzymes. Furthermore, the variety in the putative TFBS identified in different sets of orthologous genes indicates differences in the regulatory networks controlling their expression. These results suggest that there is no "master" transcriptional regulator coordinating the expression of the genes encoding CBB cycle enzymes, in contrast to what has been reported in other metabolic pathways (Okada et al., 2009; Nützmann et al., 2018). In addition, the lack of a unique, "master" regulator would emphasize the importance of the simultaneous manipulation of multiple targets to increase CBB cycle efficiency (Simkin et al., 2015, 2017; López-Calcagno et al., 2020).

Based on data validated in the model plant *A. thaliana*, we used a computational approach to identify *cis*-regulatory elements whose putative trans-acting factors might play a role in C_4 compartmentalization. These data have been used to investigate the putative role of orthologous genes in other crops (Capote et al., 2018; Moon et al., 2018; Burgess et al., 2019; Zeng et al., 2019; DeMers et al., 2020; Elzanati et al., 2020; Gray et al., 2020; Zhou et al., 2020), and allow us to generate a compelling hypothesis, as it is expected that similar DNA-binding domains of trans-acting factors would have similar DNA sequence preferences (Lambert et al., 2019). However, several complementary experimental approaches will be needed to provide evidence of functional significance in C_4 plants. To confirm the TFBS in different species, transcription factor binding assays such as DAP-seq (O'Malley et al., 2016) could be developed in some of the grass species examined in this study. To assess the chromatin accessibility of potential regulatory regions, experiments such as DNAse-seq (Zhang and Jiang, 2015) or ATAC-seq (Buenrostro et al., 2015; Bajic et al., 2018; Maher et al., 2018), could be implemented. To enhance our ability to detect regulatory elements within coding sequences (Reyna-Llorens et al., 2018), functional assays that discriminate between conserved sites with a regulatory role and conserved sites with a coding sequence role could be developed. Finally, the generation of transcriptomic data from different species using a comparable sampling process, should allow us to unveil consistent pattern of expression among different species.

Taking all of our results together, we propose that the compartmentalization of the CBB cycle enzymes investigated in this study has occurred through the recruitment of TFBS whose trans-acting factors are enriched in either one of the C_4 cell types. The expression pattern of any gene is determined by a combination of the TFBS present and the corresponding trans-acting factors binding to these regulatory regions at any given

time. It then follows that the expression pattern of any gene can be changed either by recruiting new TFBS or by altering the expression pattern of the trans-acting factors. Thus, to enrich the expression of C_4 enzymes in BS cells, new TFBS could be recruited into gene regulatory regions of C_4 species to confer BS-specific expression. Alternatively, trans-acting factors could become enriched in BS cells to promote the expression of C_4 enzymes in BS cells (as the predicted putative activators), or these factors could become enriched in M cells to repress the expression of C_4 enzymes in M cells (predicted putative repressors). This transcriptional regulation would likely be complemented by regulation at post-transcriptional and/or post-translational level to achieve a precise regulation of the expression pattern of CBB cycle enzymes.

To our knowledge, and excluding the extensive work on *RubisCO* (discussed in Hibberd and Covshoff, 2010; Berry et al., 2011; Schlüter and Weber, 2020), this is the first study to focus specifically on the differences in the regulatory sequences of CBB cycle genes between C_3 and C_4 species. These results provide a hypothetical foundation for future functional analysis. Future experiments should include the *in vivo* validation of the trans-acting factors binding to *cis*-regulatory elements, and the resultant regulation of CBB cycle genes; the transfer of C_4-specific transcription factors into C_3 species to establish a C_4-like expression pattern; or the precise genome editing of the *cis*-elements to evaluate their contribution to compartmentalization in C_4 plants.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

CAR conceived the study and supervised the research with input from CA. CA performed the analysis and wrote the manuscript with input from CAR. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Realising Improved Photosynthetic Efficiency (RIPE) initiative awarded to CAR by the University of Illinois, United States. RIPE was possible through support from the Bill & Melinda Gates Foundation, FCDO and FFAR, grant no. OPP1172157.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.570436/full#supplementary-material

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

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

Aubry, S., Kelly, S., Kümpers, B. M. C., 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. doi: 10.1371/journal.pgen.1004365

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208.

Bajic, M., Maher, K. A., and Deal, R. B. (2018). Identification of open chromatin regions in plant genomes using ATAC-Seq. *Methods Mol. Biol.* 1675, 183–201. doi: 10.1007/978-1-4939-7318-7_12

Beier, S., Himmelbach, A., Colmsee, C., Zhang, X. Q., Barrero, R. A., Zhang, Q., et al. (2017). Construction of a map-based reference genome sequence for barley, *Hordeum vulgare* L. *Sci. Data* 4:170044.

Bennetzen, J. L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A. C., et al. (2012). Reference genome sequence of the model plant Setaria. *Nat. Biotechnol.* 30, 555–561.

Berry, J. O., Patel, M., and Zielinski, A. (2011). "Chapter 12 C4 gene expression in mesophyll and bundle sheath cells," in *C4 Photosynthesis and Related CO_2 Concentrating Mechanisms*, eds A. S. Raghavendra and R. F. Sage (Dordrecht: Springer), 221–256. doi: 10.1007/978-90-481-9407-0_12

Borba, A. R., Serra, T. S., Górska, A., Gouveia, P., Cordeiro, A. M., Reyna-Llorens, I., et al. (2018). Synergistic binding of bHLH transcription factors to the promoter of the maize NADP-ME gene used in C4 photosynthesis is based on an ancient code found in the ancestral C3 state. *Mol. Biol. Evol.* 35, 1690–1705. doi: 10.1093/molbev/msy060

Bräutigam, A., Kajala, K., Wullenweber, J., Sommer, M., Gagneul, D., Weber, K. L., et al. (2011). An mRNA blueprint for C4 photosynthesis derived from comparative transcriptomics of closely related C3 and C4 species. *Plant Physiol.* 155:142. doi: 10.1104/pp.110.159442

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519

Brown, N. J., Newell, C. A., Stanley, S., Chen, J. E., Perrin, A. J., Kajala, K., et al. (2011). Independent and parallel recruitment of preexisting mechanisms underlying C(4) photosynthesis. *Science* 331, 1436–1439. doi: 10.1126/science.1201248

Buenrostro, J. D., Wu, B., Chang, H. Y., and Greenleaf, W. J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* 109, 21.29.21–21.29.29.

Burgess, S. J., Reyna-Llorens, I., Stevenson, S. R., Singh, P., Jaeger, K., and Hibberd, J. M. (2019). Genome-wide transcription factor binding in leaves from C3 and C4 grasses. *Plant Cell* 31, 2297–2314. doi: 10.1105/tpc.19.00078

Capote, T., Barbosa, P., Usie, A., Ramos, A. M., Inacio, V., Ordas, R., et al. (2018). ChIP-Seq reveals that QsMYB1 directly targets genes involved in lignin and suberin biosynthesis pathways in cork oak (*Quercus suber*). *BMC Plant Biol.* 18:198. doi: 10.1186/s12870-018-1403-5

Chang, Y.-M., Liu, W.-Y., Shih, A. C.-C., Shen, M.-N., Lu, C.-H., Lu, M.-Y. J., et al. (2012). Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. *Plant Physiol.* 160:165. doi: 10.1104/pp.112.203810

DeMers, L. C., Redekar, N. R., Kachroo, A., Tolin, S. A., Li, S., and Saghai Maroof, M. A. (2020). A transcriptional regulatory network of Rsv3-mediated extreme resistance against Soybean mosaic virus. *PLoS One* 15:e0231658. doi: 10.1371/journal.pgen.0231658

Döring, F., Streubel, M., Bräutigam, A., and Gowik, U. (2016). Most photorespiratory genes are preferentially expressed in the bundle sheath cells

of the C4 grass Sorghum bicolor. *J. Exper. Bot.* 67, 3053–3064. doi: 10.1093/jxb/erw041

Elzanati, O., Mouzeyar, S., and Roche, J. (2020). Dynamics of the transcriptome response to heat in the moss, *Physcomitrella* patens. *Int. J. Mol. Sci.* 21:1512. doi: 10.3390/ijms21041512

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32, W273–W279.

Gehlenborg, N. (2019). UpSetR: a More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets. Available online at: https://rdrr.io/cran/UpSetR/ (accessed May 21, 2020).

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2011). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186.

Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018. doi: 10.1093/bioinformatics/btr064

Gray, S. B., Rodriguez-Medina, J., Rusoff, S., Toal, T. W., Kajala, K., Runcie, D. E., et al. (2020). Translational regulation contributes to the elevated CO_2 response in two *Solanum species*. *Plant J.* 102, 383–397. doi: 10.1111/tpj.14632

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

Hirsch, C. N., Hirsch, C. D., Brohammer, A. B., Bowman, M. J., Soifer, I., Barad, O., et al. (2016). Draft assembly of elite inbred line PH207 provides insights into genomic and transcriptome diversity in maize. *Plant Cell* 28, 2700–2714. doi: 10.1105/tpc.16.00353

International Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463, 763–768. doi: 10.1038/nature08747

John, C. R., Smith-Unna, R. D., Woodfield, H., Covshoff, S., and Hibberd, J. M. (2014). Evolutionary convergence of cell-specific gene expression in independent lineages of C4 grasses. *Plant Physiol.* 165, 62–75. doi: 10.1104/pp.114.238667

Lambert, S. A., Yang, A. W. H., Sasse, A., Cowley, G., Albu, M., Caddick, M. X., et al. (2019). Similarity regression predicts evolution of transcription factor sequence specificity. *Nat. Genet.* 51, 981–989. doi: 10.1038/s41588-019-0411-1

López-Calcagno, P. E., Brown, K. L., Simkin, A. J., Fisk, S. J., Vialet-Chabrand, S., Lawson, T., et al. (2020). Stimulating photosynthetic processes increases productivity and water-use efficiency in the field. *Nat. Plants* 6, 1054–1063. doi: 10.1038/s41477-020-0740-1

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.

Maher, K. A., Bajic, M., Kajala, K., Reynoso, M., Pauluzzi, G., West, D. A., et al. (2018). Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. *Plant Cell* 30, 15–36. doi: 10.1105/tpc.17.00581

Majeran, W., Cai, Y., Sun, Q., and Van Wijk, K. J. (2005). Functional differentiation of bundle sheath and mesophyll maize chloroplasts determined by comparative proteomics. *Plant Cell* 17:3111. doi: 10.1105/tpc.105.035519

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., et al. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544, 427–433.

McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 40, 4288–4297. doi: 10.1093/nar/gks042

McCormick, R. F., Truong, S. K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., et al. (2018). The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* 93, 338–354. doi: 10.1111/tpj.13781

Michelet, L., Zaffagnini, M., Morisse, S., Sparla, F., Perez-Perez, M. E., Francia, F., et al. (2013). Redox regulation of the Calvin-benson cycle: something old, something new. *Front. Plant Sci.* 4:470. doi: 10.3389/fpls.2013.00470

Moon, S., Chandran, A. K. N., An, G., Lee, C., and Jung, K. H. (2018). Genome-wide analysis of root hair-preferential genes in rice. *Rice* 11:48.

Niklaus, M., and Kelly, S. (2019). The molecular evolution of C4 photosynthesis: opportunities for understanding and improving the world's most productive plants. *J. Exp. Bot.* 70, 795–804. doi: 10.1093/jxb/ery416

Nomura, M., Katayama, K., Nishimura, A., Ishida, Y., Ohta, S., Komari, T., et al. (2000). The promoter of rbcS in a C3 plant (rice) directs organ-specific, light-dependent expression in a C4 plant (maize), but does not confer bundle sheath cell-specific expression. *Plant Mol. Biol.* 44, 99–106.

Nützmann, H.-W., Scazzocchio, C., and Osbourn, A. (2018). Metabolic gene clusters in eukaryotes. *Annu. Rev. Genet.* 52, 159–183. doi: 10.1146/annurev-genet-120417-031237

Okada, A., Okada, K., Miyamoto, K., Koga, J., Shibuya, N., Nojiri, H., et al. (2009). OsTGAP1, a bZIP transcription factor, coordinately regulates the inductive production of Diterpenoid Phytoalexins in rice. *J. Biol. Chem.* 284, 26510–26518. doi: 10.1074/jbc.m109.036871

O'Malley, R. C., Huang, S.-C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., et al. (2016). Cistrome and Epicistrome features shape the regulatory DNA landscape. *Cell* 165, 1280–1292. doi: 10.1016/j.cell.2016.04.038

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., et al. (2007). The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res.* 35, D883–D887.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.

Rao, X., Lu, N., Li, G., Nakashima, J., Tang, Y., and Dixon, R. A. (2016). Comparative cell-specific transcriptomics reveals differentiation of C4 photosynthesis pathways in switchgrass and other C4 lineages. *J. Exper. Bot.* 67, 1649–1662. doi: 10.1093/jxb/erv553

Reyna-Llorens, I., Burgess, S. J., Reeves, G., Singh, P., Stevenson, S. R., Williams, B. P., et al. (2018). Ancient duons may underpin spatial patterning of gene expression in C4 leaves. *Proc. Natl. Acad. Sci. U.S.A.* 115:1931. doi: 10.1073/pnas.1720576115

Robinson, M. D., Mccarthy, D. J., and Smyth, G. K. (2009). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Rose, A. B. (2019). Introns as gene regulators: a brick on the accelerator. *Front. Genet.* 9:672. doi: 10.3389/fgene.2018.00672

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

Schlüter, U., and Weber, A. P. M. (2020). Regulation and evolution of C4 photosynthesis. *Annu. Rev. Plant Biol.* 71, 183–215.

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115.

Sheen, J. Y., and Bogorad, L. (1987). Regulation of levels of nuclear transcripts for C4 photosynthesis in bundle sheath and mesophyll cells of maize leaves. *Plant Mol. Biol.* 8, 227–238. doi: 10.1007/bf00015031

Simkin, A. J., Lopez-Calcagno, P. E., Davey, P. A., Headland, L. R., Lawson, T., Timm, S., et al. (2017). Simultaneous stimulation of sedoheptulose 1,7-bisphosphatase, fructose 1,6-bisphophate aldolase and the photorespiratory glycine decarboxylase-H protein increases CO_2 assimilation, vegetative biomass and seed yield in *Arabidopsis*. *Plant Biotechnol. J.* 15, 805–816. doi: 10.1111/pbi.12676

Simkin, A. J., Mcausland, L., Headland, L. R., Lawson, T., and Raines, C. A. (2015). Multigene manipulation of photosynthetic carbon assimilation increases CO_2 fixation and biomass yield in tobacco. *J. Exp. Bot.* 66, 4075–4090. doi: 10.1093/jxb/erv204

Studer, A. J., Schnable, J. C., Weissmann, S., Kolbe, A. R., Mckain, M. R., Shao, Y., et al. (2016). The draft genome of the C3 panicoid grass species *Dichanthelium oligosanthes*. *Genome Biol.* 17:223.

Tange, O. (2018). *GNU Parallel 2018*. Morrisville: Lulu.

Washburn, J. D., Kothapalli, S. S., Brose, J. M., Covshoff, S., Hibberd, J. M., Conant, G. C., et al. (2017). Ancestral reconstruction and C3 bundle sheath transcript abundance in the paniceae grasses indicate the foundations for all three biochemical c4 sub-types were likely present in the most recent ancestor. *bioRxiv* [Preprint]. doi: 10.1101/162644

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.

Wickham, H., François, R., Henry, L., and Müller, K. (2019). *dplyr: a Grammar of Data Manipulation*. Available online at: https://CRAN.R-project.org/package=dplyr (accessed May 21, 2020).

Wilkins, D. (2019). *gggenes: Draw Gene Arrow Maps in 'ggplot2'*. Available online at: https://rdrr.io/cran/gggenes/ (accessed May 17, 2020).

Williams, B. P., Burgess, S. J., Reyna-Llorens, I., Knerova, J., Aubry, S., Stanley, S., et al. (2016). An untranslated *cis*-element regulates the accumulation of multiple C4 enzymes in *Gynandropsis gynandra* Mesophyll cells. *Plant Cell* 28:454. doi: 10.1105/tpc.15.00570

Xu, J., Bräutigam, A., Weber, A. P. M., and Zhu, X.-G. (2016). Systems analysis of cis-regulatory motifs in C4 photosynthesis genes using maize and rice leaf transcriptomic data during a process of de-etiolation. *J. Exper. Bot.* 67, 5105–5117. doi: 10.1093/jxb/erw275

Zeng, Z., Zhang, W., Marand, A. P., Zhu, B., Buell, C. R., and Jiang, J. (2019). Cold stress induces enhanced chromatin accessibility and bivalent histone modifications H3K4me3 and H3K27me3 of active genes in potato. *Genome Biol.* 20:123.

Zhang, W., and Jiang, J. (2015). Genome-wide mapping of DNase I hypersensitive sites in plants. *Methods Mol. Biol.* 1284, 71–89. doi: 10.1007/978-1-4939-2444-8_4

Zhou, P., Li, Z., Magnusson, E., Gomez Cano, F., Crisp, P. A., Noshay, J. M., et al. (2020). Meta gene regulatory networks in maize highlight functionally relevant regulatory interactions. *Plant Cell* 32, 1377–1396. doi: 10.1105/tpc.20.00080

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

HYPOTHESIS AND THEORY
published: 02 November 2020
doi: 10.3389/fpls.2020.578739

Evolutionary Convergence of C_4 Photosynthesis: A Case Study in the Nyctaginaceae

Roxana Khoshravesh[1,2†], Matt Stata[1†], Shunsuke Adachi[1,3†], Tammy L. Sage[1†] and Rowan F. Sage[1†]*

[1] Department of Ecology and Evolutionary Biology, The University of Toronto, Toronto, ON, Canada, [2] Department of Biology, The University of New Mexico, Albuquerque, NM, United States, [3] Institute of Global Innovation Research, Tokyo University of Agriculture and Technology, Fuchu, Japan

OPEN ACCESS

Edited by:
Tingshuang Yi,
Kunming Institute of Botany, Chinese Academy of Sciences, China

Reviewed by:
Isabel Larridon,
Royal Botanic Gardens, Kew, United Kingdom
Sidonie Bellot,
Royal Botanic Gardens, Kew, United Kingdom

***Correspondence:**
Rowan F. Sage
r.sage@utoronto.ca

†**ORCID:**
Roxana Khoshravesh
orcid.org/0000-0002-1766-8993
Matt Stata
orcid.org/0000-0002-5744-4898
Shunsuke Adachi
orcid.org/0000-0003-0471-3369
Tammy L. Sage
orcid.org/0000-0002-7061-832X
Rowan F. Sage
orcid.org/0000-0001-6183-9246

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: 01 July 2020
Accepted: 06 October 2020
Published: 02 November 2020

Citation:
Khoshravesh R, Stata M, Adachi S, Sage TL and Sage RF (2020) Evolutionary Convergence of C_4 Photosynthesis: A Case Study in the Nyctaginaceae. Front. Plant Sci. 11:578739. doi: 10.3389/fpls.2020.578739

C_4 photosynthesis evolved over 65 times, with around 24 origins in the eudicot order Caryophyllales. In the Caryophyllales family Nyctaginaceae, the C_4 pathway is known in three genera of the tribe Nyctagineae: *Allionia*, *Okenia* and *Boerhavia.* Phylogenetically, *Allionia* and *Boerhavia/Okenia* are separated by three genera whose photosynthetic pathway is uncertain. To clarify the distribution of photosynthetic pathways in the Nyctaginaceae, we surveyed carbon isotope ratios of 159 species of the Nyctaginaceae, along with bundle sheath (BS) cell ultrastructure, leaf gas exchange, and C_4 pathway biochemistry in five species from the two C_4 clades and closely related C_3 genera. All species in *Allionia, Okenia* and *Boerhavia* are C_4, while no C_4 species occur in any other genera of the family, including three that branch between *Allionia* and *Boerhavia*. This demonstrates that C_4 photosynthesis evolved twice in Nyctaginaceae. *Boerhavia* species use the NADP-malic enzyme (NADP-ME) subtype of C_4 photosynthesis, while *Allionia* species use the NAD-malic enzyme (NAD-ME) subtype. The BS cells of *Allionia* have many more mitochondria than the BS of *Boerhavia*. Bundle sheath mitochondria are closely associated with chloroplasts in *Allionia* which facilitates CO_2 refixation following decarboxylation by mitochondrial NAD-ME. The close relationship between *Allionia* and *Boerhavia* could provide insights into why NADP-ME versus NAD-ME subtypes evolve, particularly when coupled to analysis of their respective genomes. As such, the group is an excellent system to dissect the organizational hierarchy of convergent versus divergent traits produced by C_4 evolution, enabling us to understand when convergence is favored versus when divergent modifications can result in a common phenotype.

Keywords: *Allionia*, *Boerhavia*, C_4 photosynthesis, convergent evolution, Nyctaginaceae phylogeny, PEP carboxylase

INTRODUCTION

C_4 photosynthesis is a complex trait that arises following modifications to hundreds if not thousands of individual genes within a genome (Gowik et al., 2011). Despite this, it is one of the most convergent of evolutionary phenomena in the biosphere, with over 65 independent origins (Conway-Morris, 2003; Sage, 2016; Heyduk et al., 2019). Evolutionary convergence, however, does not necessarily reflect convergence throughout the hierarchy of traits that give rise to a complex

phenotype, because multiple mechanisms can support a common function (Losos, 2011). This is well illustrated in the case of C_4 photosynthesis and crassulacean acid metabolism (CAM), each of which have been repeatedly assembled using disparate enzymes and structural modifications (Sage et al., 2012; Christin and Osborne, 2013; Edwards, 2019). While examples of evolutionary convergence are many, the mechanisms of convergence remain a major question in the life sciences, particularly in the cases where complex traits such as C_4 photosynthesis repeatedly evolve (Blount et al., 2018). Because the complexity of the C_4 system is well-understood, as well as the phylogenetic distribution of the many C_4 clades, C_4 photosynthesis represents an excellent system to understand the mechanics of convergent evolution, and its implication for the rise of C_4-dominated biomes over the past 30 million years (Christin and Osborne, 2013; Heyduk et al., 2019).

C_4 photosynthesis first captures CO_2 at low concentration in an outer mesophyll (M) compartment via the activity of phosphoenolpyruvate (PEP) carboxylase (PEPCase), and then concentrates it into an internal compartment, typically a layer of cells around the leaf vasculature termed the bundle sheath (BS; Edwards and Walker, 1983; Hatch, 1987).[1] The C_4 pathway begins with the conversion of CO_2 to bicarbonate (HCO_3^-) by carbonic anhydrase (CA) in M cells, followed by PEP carboxylation (**Figure 1**). These two steps occur in all C_4 plants, and thus are universally convergent traits in C_4 photosynthesis; however, different paralogs have been recruited to carry out the PEPCase and CA functions, demonstrating evolutionary flexibility at a lower level of organization (Christin et al., 2010b, 2013a; Ludwig, 2016a; Heyduk et al., 2019). The product of PEP carboxylation, oxaloacetate (OAA), is too labile to safely move between M and BS cells, so it must be converted to a stable metabolite (Edwards and Walker, 1983). Metabolite transport between M and BS cells is by diffusion, which necessitates that metabolites form steep concentration gradients to support rapid flux, and thus must be stable at high concentration (Bräutigam and Weber, 2011). The solution to the challenge presented by OAA instability is to convert it to the stable metabolites malate or aspartate. This highlights another fundamental feature of evolutionary convergence, in that it occurs where there are strict physiochemical constraints such as OAA lability. A common step (OAA conversion) is accomplished via divergent metabolic solutions (formation of malate versus aspartate). The selected transport metabolite, as it turns out, reflects the enzyme that catalyzes the decarboxylation step.

There are three major decarboxylating enzymes co-opted for the decarboxylation step in C_4 photosynthesis: NADP-malic enzyme (NADP-ME), NAD-malic enzyme (NAD-ME) and PEP carboxykinase (PCK; Hatch, 1987). Most C_4 lineages use NADP-ME as the primary decarboxylating enzyme, about 1/3 use NAD-ME and five lineages predominantly use PCK (Sage R. F. et al., 2011). The consequence of the decarboxylase selection affects many aspects of C_4 photosynthesis, to include the pathway biochemistry, BS and M ultrastructure, M to BS transport processes, leaf energetics, and photosynthetic efficiency (Hatch, 1987; Drincovich et al., 2011; Ghannoum et al., 2011). To recognize the suite of traits associated with each decarboxylation mode, three subtypes of C_4 photosynthesis have been delineated – the NADP-ME subtype, the NAD-ME subtype, and the PCK subtype (**Figure 1**). Although each subtype conducts C_4 photosynthesis, they represent three evolutionary mechanisms to concentrate CO_2 that are derived from a distinct set of biochemical, structural and transport traits (Rao and Dixon, 2016). While each subtype represents a divergent means of concentrating CO_2 around Rubisco, traits associated within each subtype reflect strong convergence in response to constraints imposed by the decarboxylating enzyme. The NADP-ME enzyme co-opted by the C_4 cycle is located in the BS chloroplasts, where it uses malate and $NADP^+$ to produce CO_2, NADPH and pyruvate, with the NADPH directly supporting reduction of PGA produced by Rubisco. NAD-ME is mitochondrial, and uses malate and NAD^+ to produce CO_2, NADH and pyruvate. Because it is active in the BS mitochondria, substantial mitochondrial volume is needed to meet the metabolic requirements of the C_4 pathway, and consistently, species of the NAD-ME subtype often have more and/or larger mitochondria in BS cells than NADP-ME species (Dengler and Nelson, 1999; Edwards and Voznesenskaya, 2011). The PCK enzyme is located in the cytosol, and uses ATP and OAA to produce PEP, CO_2, and ADP + P_i. Large amounts of ATP are needed for the PCK reaction, and hence there is a large investment in BS mitochondrial volume to meet the ATP requirement (Yoshimura et al., 2004; Voznesenskaya et al., 2006). The decarboxylation type also constrains the selection of the transport metabolite. In NADP-ME species, NADPH produced by malate oxidation in the BS chloroplast is rapidly consumed in the metabolism of PGA generated by Rubisco, so there is no feedback onto malate flux. In NAD-ME species, the use of malate as a transport molecule would be problematic. NADH cannot readily exit the mitochondria, and there is an insufficient energy sink in the mitochondria to use the large amount of NADH that would be generated if NAD-ME oxidized malate imported from the M cells. The utilization of aspartate avoids these issues because there is no net import of reducing power into the BS mitochondria (Kanai and Edwards, 1999). Once in the BS mitochondria, aspartate forms OAA, which is reduced to malate via malate dehydrogenase using the NADH generated by NAD-ME (**Figure 1**).

PEP carboxykinase directly uses OAA, and hence to avoid flooding the BS cytosol with reducing power, aspartate is also imported from M cells. However, the PCK reaction requires large amounts of ATP to support rapid photosynthesis, and this can be generated in the mitochondria by oxidizing NADH created by NAD-ME using malate directly imported into the BS mitochondria from the M cells (Leegood and Walker, 1999). PEP carboxykinase species are thought to use NAD-ME at about a fourth to a third of the rate of the PCK reaction, in order to supply sufficient NADH for ATP production (Leegood and Walker,

[1] The tissue layer where CO_2 is concentrated is often comprised of parenchymatous bundle sheath cells; however, in an example of the non-convergence, other cell layers have also been co-opted as the site of CO_2 concentration. Despite this variation, bundle sheath (BS) is the term generally used to refer to the high CO_2 compartment. We generally follow this convention here, although we do use the anatomically correct name when discussing a specific Kranz anatomy cell type that is not technically BS tissue.

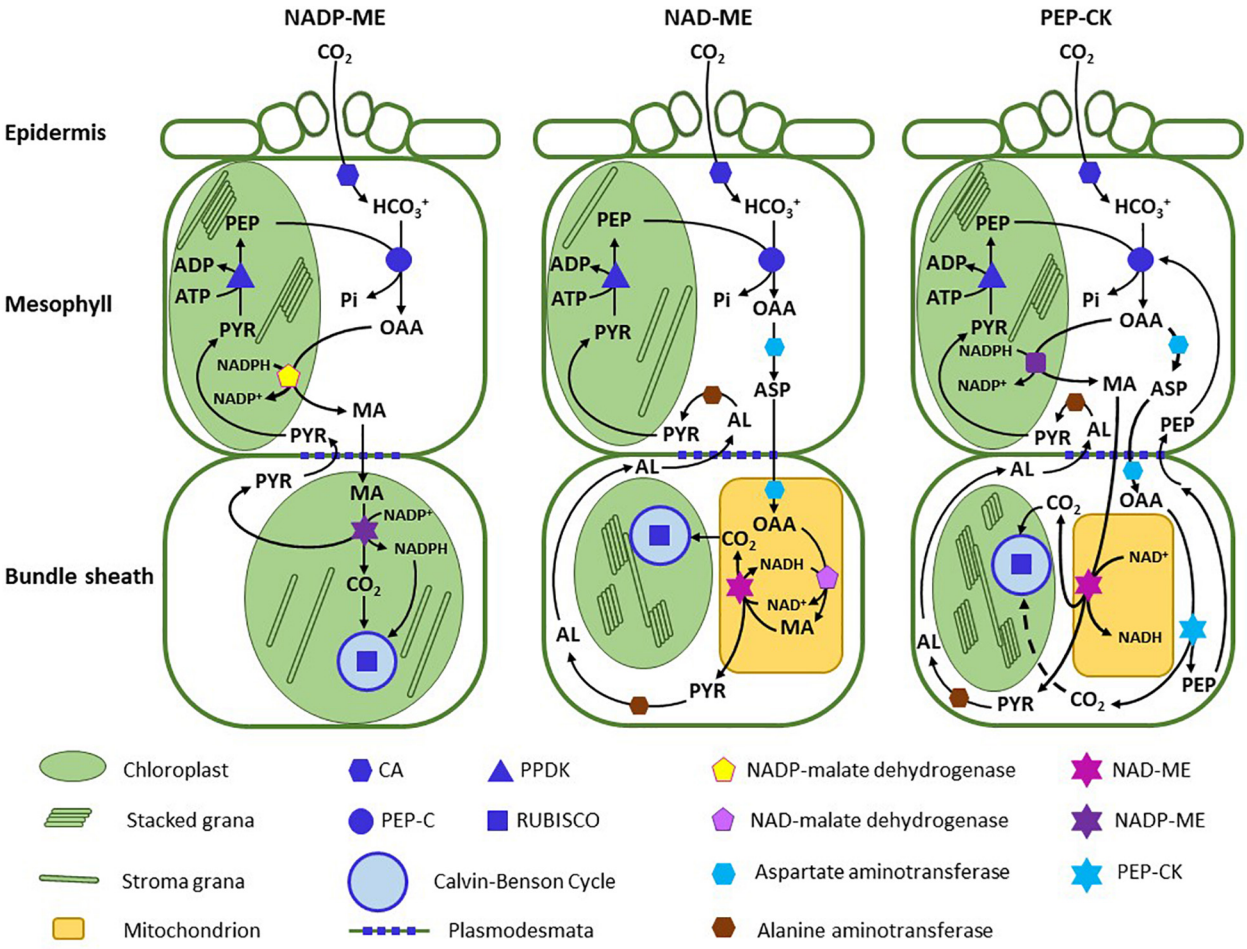

FIGURE 1 | A schematic outlining the C_4 pathway in NADP-ME type species, NAD-ME species, and PEP carboxykinase species. The diagram represents the classical depiction of the pathway (as per Hatch, 1987 and Furbank, 2011) assuming malate or aspartate predominate in their respective pathways. For clarity, nitrogen exchange reactions between aspartate and pyruvate are not shown in the NAD-ME and PCK subtypes. Thylakoid stacking is indicated by multi-layered rectangles in the chloroplasts. AL, alanine; ASP, aspartate; CA, carbonic anhydrase; MA, malate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PEP-C, PEP carboxylase; PEP-CK, PEP carboxykinase; PPDK, pyruvate, phosphate dikinase; PVA, pyruvate.

1999). Consistently, PCK species have many mitochondria in the BS (Dengler and Nelson, 1999).

To meet the energy requirements imposed by the decarboxylating enzymes and associated transport systems, a distinct arrangement of chloroplast membranes occurs in each of the three subtypes (Hatch, 1987; Edwards and Voznesenskaya, 2011; Rao and Dixon, 2016). NADP-ME species have low grana stacking in the BS chloroplasts, yet high stacking in the M chloroplasts. More grana stacking in M cells increases the PSII content in the thylakoids, and hence the potential to generate NADPH by linear electron transport. Because NADP-ME species import reducing power into the BS with malate, the demand for NADPH to reduce PGA in the BS chloroplasts of NADP-ME plants is halved.; hence less PSII and granal stacking are required (Kanai and Edwards, 1999). Chloroplasts in the M cells of NADP-ME species are well stacked to support malate reduction following PEP carboxylation. NAD-ME species, by contrast, have low grana stacking in M chloroplasts because their primary energetic function is to produce ATP for PEP regeneration. The thylakoids of the BS exhibit pronounced stacking in NAD-ME species, to supply sufficient NADPH to potentially reduce all the PGA generated by the C_3 cycle. However, there can be variation on these general chloroplast phenotypes, depending upon the degree to which multiple decarboxylases are employed, whether PGA is exported to the M tissue for reduction, or if aspartate is imported into the chloroplast in lieu of malate, as suggested to occur in NADP-ME *Flaveria* species (Leegood and Walker, 1999; Furbank, 2011).

The theoretical pattern that emerges in C_4 evolution is thus one of universal convergence in terms of overall outcome (CO_2 concentration into a BS-like compartment), with some steps being universal (CO_2 conversion to bicarbonate, PEP carboxylation), and others being divergent (decarboxylation, transport, chloroplast energetics, and possibly how PEP is regenerated). Convergence occurs within the subtype pathways where specific biophysical constraints are present, for example, in

transport metabolites associated with the distinct decarboxylating enzymes. An organizational hierarchy between convergence and divergence can thus be envisioned for complex traits that scales from the level of the genes up through to the composite phenotype. The challenge for research on convergent evolution is to describe this hierarchy and to understand why and under what constraints convergent patterns emerge (Losos, 2011). To do this, one needs effective research systems, such as fast cycling microbes, or in terrestrial plants, multiple lineages where convergence has occurred, such as the dozens of clades that have independently evolved C_4 photosynthesis (Blount et al., 2018; Heyduk et al., 2019). However, most closely related C_4 clades are of the same subtype (Sage R. F. et al., 2011), which restricts the ability to examine sub-type impacts on convergent versus divergent solutions to creating a C_4 pathway. One potentially strong study system occurs in *Portulaca*, where both NADP-ME and NAD-ME species are present (Voznesenskaya et al., 2010; Ocampo et al., 2013). Another possibility occurs in the Nyctagineae tribe of the Nyctaginaceae. Here, three related genera of the tribe Nyctagineae - *Allionia*, *Boerhavia* and *Okenia* - contain C_4 species while their closest relatives in the genera *Anulocaulis*, *Cyphomeris, Commicarpus*, and *Nytctaginia* are not known to have any C_4 species (Sage R. F. et al., 2011). In the phylogeny of Douglas and Manos (2007), *Okenia* and *Boerhavia* form a common C_4 clade, while *Allionia* forms a distinct C_4 clade that is separated from the *Okenia/Borehavia* lineage by the *Anulocaulis/Nyctaginia* complex. However, there has been no systematic survey of the occurrence of C_3 and C_4 photosynthesis in the Nyctaginaceae, so it is unclear whether C_4 may exist in *Anulocaulis, Commicarpus, Cyphomeris,* and *Nyctaginia*, or even whether *Allionia*, *Okenia*, and *Boerhavia* are completely C_4, as opposed to also containing C_3 and C_3–C_4 intermediate species. *Boerhavia* and *Allionia* are listed as being NADP-ME, although this is based on enzyme assays for *Boerhavia* only (Muhaidat et al., 2007). Intriguingly, our preliminary TEM images show many mitochondria in the BS ultrastructure of *Allionia*, suggesting it is NAD-ME. If so, then the *Allionia* and *Okenia/Boerhavia* clade could join *Portulaca* in forming a robust study system for addressing evolutionary convergence as influenced by C_4 subtype.

To evaluate the potential of the Nyctaginaceae to become a model for studying evolutionary convergence, we present here a detailed study of the distribution of the C_3 and C_4 pathways in the Nyctaginaceae. We present a survey of carbon isotope ratios from 560 herbarium specimens, in addition to an anatomical/ultrastructural study using five representative species of *Allionia*, *Anulocaulis*, *Boerhavia*, *Commicarpus*, and *Nyctaginia*. We also examine climate data and geographic distributions of species in *Allionia*, *Anulocaulis*, *Boerhavia*, *Commicarpus*, *Cyphomeris*, and *Nyctaginia* to evaluate ecological factors contributing to C_4 origins in the Nyctaginaceae. The biochemical sub-types of C_4 species in *Allionia* and *Boerhavia* were determined, and we present a transcriptome-based phylogeny that updates the phylogenies of Douglas and Manos (2007) and Douglas and Spellenberg (2010). We also use transcriptome data to evaluate whether there has been convergence in the gene sequences of the major C_4 pathway enzymes. For comparative purposes, we also present TEM images of leaf tissues from two *Portulaca* (Portulaceae) species previously shown to be NADP-ME (*Portulaca pilosa*) or NAD-ME (*Portulaca oleracea*). Through these efforts, we present an initial hierarchical assessment of how divergence occurs during the convergent evolution of C_4 photosynthesis.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Seeds of *Allionia incarnata*, *Anulocaulis gypsogenus*, *Boerhavia burbigeana*, *Boerhavia coccinea*, *Commicarpus scandens, and Nyctaginia capitata* were collected from naturally occurring field populations in Southwestern North America and Australia between 2000 and 2016 (see **Supplementary Table 1** for collection and voucher information). *Portulaca pilosa* seeds were a gift from Gerry Edwards and Elena Vosnesenskaya (Washington State University), while *P. oleracea* grew as a weed in our greenhouse. Seeds were sown directly into 10 or 20 L pots containing a sandy-loam mixture and grown in the University of Toronto greenhouse complex housed on the roof of the Earth's Sciences Centre. Plants were watered as needed to avoid soil drying (daily in high summer, two to three times weekly in cooler weather), and fertilized bimonthly with a Miracle-Grow commercial fertilizer mix (All-Purpose Brand, 24-8-16) amended with 4 mM calcium nitrate and 1 mM magnesium sulfate. Greenhouse conditions were 27 to 33°C daytime temperature, 20–24°C night temperature, and a peak photon flux density (PFD) of 1500 μmol photons m^{-2} s^{-1} on clear days. We used 400 W sodium vapor lamps to supplement natural daylight to maintain a minimum PFD on cloudy days of 250 μmol m^{-2} s^{-1} over a >13 h photoperiod. Unless otherwise indicated, three to five plants were sampled for gas exchange, biochemical assay, leaf structural properties and transcriptomics between May and October of 2010 to 2017. Care was taken to sample tissues under the same environmental conditions in the greenhouse (full sun exposure, a leaf temperature near 27°C, at a time between 9 am and 2 pm) to minimize year to year and month to month variation. For the characteristics examined here, subtle variation that may exist between sampling dates is not known to affect results pertinent to our hypotheses. For example, C_3 versus C_4 expression patterns are largely constitutive, and sun to shade variation in phenotype would not be present at the bright light intensities present in our greenhouses during summer.

Carbon Isotope Ratio

To determine the distribution of the C_3 and C_4 pathways in species of the Nyctaginaceae, 560 herbarium specimens representing 23 genera and 159 species were sampled from the herbaria at Kew gardens (Richmond, London, United Kingdom), Missouri Botanical Gardens (St Louis, MO United States), and New York Botanical Gardens (The Bronx, New York, NY, United States; **Supplementary Tables 2**, 3). Species were sampled from all genera in the Nyctagineae with the exception of the monospecific genus *Cuscatlainia*. Approximately 50% (*Mirabilis*) to 100% (*Okenia, Allionia*) of the known species from the

sampled Nyctagineae genera are present in the survey. We also sampled species from 12 genera occurring in each of the other six tribes of the Nyctaginaceae (**Table 1** and **Supplementary Table 2**). To measure the carbon isotope ratio ($\delta^{13}C$) of each herbarium specimen, 2–4 mg of leaf or stem material were sampled from herbarium sheets and assayed for $\delta^{13}C$ by the Washington State University Stable Isotope Core.[2] $\delta^{13}C$ was determined for at least two distinct plant specimens if available in the herbaria. Carbon isotope ratios between −10 and −16‰ correspond to C_4 values, while carbon isotope ratios between −23 and −32‰ correspond to C_3 values. Species that are evolutionary intermediates between C_3 and C_4 photosynthesis exhibit $\delta^{13}C$ ratios that are similar to C_3 values except where a strong C_4 metabolic cycle has been engaged; such C_4-like plants typically exhibit $\delta^{13}C$ values between −16 and −22‰ (Monson et al., 1988; Von Caemmerer, 1992).

Leaf Gas Exchange and Biochemical Assay

The response of net carbon assimilation rate (A) to intercellular CO_2 concentration (C_i) was measured for *Allionia incarnata*, *B. coccinea*, and *N. capitata* using a Li-COR 6400 photosynthetic gas analyzer at 33°C, a vapor pressure differences of 2.0–2.5 kPa and a photosynthetic PFD of 1800 μmol m^{-2} s^{-1}. The most recent, fully mature leaves were first equilibrated in the leaf cuvette at an ambient CO_2 concentration of 400 μmol mol^{-1} and 1800 μmol photons m^{-2} s^{-1}. After equilibration and measurement, the ambient CO_2 was increased to 1200 μmol mol^{-1} and then gas exchange parameters were measured after equilibration. The CO_2 concentration was then reduced to near the CO_2 compensation point in approximately 13 steps, with steady-state gas exchange values determined at each step. The initial slope of the A *versus* C_i response was estimated as the linear slope of the measurements below a C_i of 100 μmol mol^{-1}. Intrinsic water use efficiency (WUE) was determined as the ratio of A to stomatal conductance (g_s) at an ambient CO_2 of 400 ppm, and the CO_2-saturated rate of A (A_{1200}) was measured at 1200 μmol mol^{-1} CO_2.

The activities of PEPC, NADP-malic enzyme, and NAD-malic enzyme were assayed at 30°C using a coupled-enzyme assay that measured oxidation/reduction rate of NADP(H) or NAD(H) at a wavelength of 340 nm using a Hewitt-Packard 8230 spectrophotometer following procedures in Ashton et al. (1990) as modified by Sage T. L. et al. (2011). Two to three cm^2 of recent, fully-mature leaves of *A. incarnata, B. coccinea,* and *N. capitata* were sampled under full illumination in the greenhouse and then rapidly ground using a glass tissue homogenizer in an extraction buffer (100 mM HEPES – pH 7.6, 5 mM $MgCl_2$, 10 mM $KHCO_3$, 2 mM EDTA, 10 mM 6-aminocaproic acid, 2 mM benzamide, 1 mM phenylmethylsulfonyl fluoride, 1% (w/v) PVPP, 2% (w/v) PVP, 0.5% Triton X-100, 2% (w/v) BSA, 5 mM DTT, 1% (w/v) casein). After removing two aliquots for chlorophyll assay (in 80% acetone at 645 and 663 nm; Arnon, 1949), the extract was centrifuged 30 s and divided into three aliquots and put on ice. The aliquot used for NAD-ME assay was immediately treated with sufficient $MnCl_2$ to give a 2 mM solution. PEPC was assayed by coupling the production of OAA to NADH oxidation via malate dehydrogenase in an assay buffer containing 50 mM Bicine (pH 8.0), 1 mM EDTA, 5 mM $MgCl_2$, 2 mM $NaHCO_3$, 2 mM DTT, 1 mM glucose 6-phosphate, 2 units ml^{-1} malate dehydrogenase, and 0.2 mM NADH. The reaction was initiated by the addition of PEP to give 5mM. NADP-ME was assayed by following $NADP^+$ reduction at 340 nm. The assay buffer contained 25 mM Tricine (pH 8.2), 1 mM EDTA, 20 mM $MgCl_2$, 2 mM DTT, 0.5 mM NADP. The reaction was initiated by the addition of malic acid to give 5 mM. NAD-ME was assayed by measuring NAD^+ reduction at 340 nm. The reaction mixture contained 25 mM Hepes (pH 7.2), 0.2 mM EDTA, 8 mM ammonium sulfate, 1 unit ml^{-1} malate dehydrogenase, 5 mM malic acid, 0.025 mM NADH, and 2 mM NAD^+. The reaction was initiated by adding $MnCl_2$ to give 5 mM.

Imaging and Quantification of Leaf Structure

For all anatomical and ultrastructural imaging, the middle portion of a leaf blade equidistant between the mid-rib and leaf margin were sampled from recent, fully expanded leaves. Leaf pieces approximately 2 mm^2 were collected between 9:00–11:00 am and were prepared for light and transmission electron microscopy as described previously (Khoshravesh et al., 2017). Cell and organelle features of BS cells were quantified using Image J software (Schneider et al., 2012). One leaf per plant was sampled from three to five plants. Each mean value per plant was compiled from measurements of 5–10 imaged cells from the adaxial region of the leaf. All measurements were conducted on planar images of cross (=tranverse) sections. Cells measured were randomly selected from the pool of cells where the plane of section passed through the central region of a BS cell, rather than the periphery. Parameters measured include BS cell area; number of chloroplasts per BS cell; number of mitochondria per BS cell; area of individual chloroplasts and mitochondria; and% BS cell area covered by all chloroplasts and mitochondria in a BS cell.

Phylotranscriptomic Analysis

To prepare a phylotranscriptome of the Nyctaginaceae, we used publicly availably RNA-sequence data from the NCBI Short Read Archive (**Supplementary Table 4**).[3] FASTQ read data was trimmed using Trimmomatic (Bolger et al., 2014) with minimum leading and trailing quality cutoffs at 30, a 5-base sliding window cutoff of 30, and 70bp minimum trimmed length. Transcriptomes were *de novo* assembled using Trinity (Grabherr et al., 2011) with default settings including *in silico* read normalization. Open reading frames were predicted and translated using the getorf program from the EMBOSS package (Rice et al., 2000), and the longest open reading frame for each putative locus identified by Trinity was selected using a Python script. OrthoFinder (Emms and Kelly, 2015) was used to predict groups of orthologous genes using 10 reference proteomes with MapMan annotations (Thimm et al., 2004) and 10 with all other annotation types

[2] www.isotopes.wsu.edu

[3] http://ncbi.nlm.nih.gov/sra

TABLE 1 | $\delta^{13}C$ data for sampled species from the Nyctaginaceae showing number of accepted species for each genus and the number of species sampled, the range of $\delta^{13}C$ for species means and number of C_4 species we identified.

Tribe and genus	Accepted species number/sampled species number	$\delta^{13}C$ range of species means	Number of C_4 species in Supplementary Table 2
Boldoeae			
Salpianthus Humb. & Bonpl.	5/3	−28.8 to −27.4	0
Bougainvilleeae			
Belemia Pires	1/1	non-Kranz anatomy	0
Phaeoptilum Radlk.	1/1	−24.4	0
Caribeeae			
Cryptocarpus H.B.K.	1/1	−27.2	0
Colignonieae			
Colignonia Endl.	6/6	−28.7 to −25.3	0
Leucastereae			
Andradea Allemão	1/1	−26.0	0
Leucaster Choisy	1/1	−28.8	0
Ramisia Glaz. ex Baillon	1/1	−24.3	0
Nyctagineae			
Abronia Juss.	24/20	−29.5 to −24.5	0
Acleisanthes A. Gray	17/14	−28.0 to −23.9	0
***Allionia* L.**	**2/2**	**−13.3 to −13.2**	**2**
Anulocaulis Standl.	5/5	−28.0 to −24.2	0
***Boerhavia* L.**	**40/43**	**−14.9 to −9.4**	**43**
Commicarpus Standl.	25/23	−28.7 to −25.2	0
Cyphomeris Standl.	2/2	−28.1 to −27.2	0
Mirabilis L.	54–60/23	−29.8 to −22.6	0
Nyctaginia Choisy	1/1	−26.2	0
***Okenia* Schldl. & Cham.**	**1/4**	**−14.0 to −12.3**	**4**
Tripterocalyx	4	27.8 to 24.9	0
Pisonieae			
Pisoniella (Heimerl) Standl.	1/1	−25.8	0
Grajalesia Miranda	1/1	−25.7	0
Cephalotomandra Karst. & Triana	1	−28.8	0

*C_4 genera are highlighted in bold. See **Supplementary Tables 2,3** for herbarium specimens, collection information and individual $\delta^{13}C$ values. Accepted species numbers follows Spellenberg, 2003, Tropicos (2020), and the International Plant Names Index (IPNI, 2020). Sampled species number includes species listed in **Supplementary Tables 2,3** whose acceptability is uncertain. Belemia fucsiodes is a type specimen examined with a dissecting scope for vein density only.*

available from the Phytozome v12 (**Supplementary Table 5**).[4] Single- or low-copy orthogroups were selected with a Python script based on the following criteria: a minimum of 40 of 53 species present; no more than 10% of the species having multiple sequences; and a minimum alignment length of 100 amino acids. For species with multiple sequences, a Python script was used to concatenate all orthogroup sequences from a given species if none overlapped by more than 10% of the shortest sequence length, or remove them all if they did. Sequences were assumed to represent fragmented assemblies in the former case and paralogs in the latter. This selection process resulted in 1084 genes for phylogenetic analysis. Initial protein alignments were produced with mafft (Katoh and Standley, 2013) and DNA codon alignments were generated from these using pal2nal (Suyama et al., 2006). The codon alignments were trimmed using trimAl (Capella-Gutiérrez et al., 2009) with a gap threshold of 0.5, then concatenated into a partitioned super-matrix using a Python script written by co-author M. Stata. Phylogenetic inference was conducted using RaxML (Stamatakis, 2014). This was conducted using separate evolutionary models for all partitions, rapid bootstrap analysis with a search for the best tree in one run (command line option -f a), and bootstrap convergence testing (autoMRE). Convergence testing showed that 50 bootstrap replicates were sufficient, but in order to be certain of support values we conducted 200. The species tree was visualized using FigTree.[5] All Python scripts are available at github.com/MattStata/Nyctaginaceae_Scripts.

C_4 Gene Phylogenies and Analysis

All homologs of PEPC, NADPME, NADME, and PPDK were identified in the Nyctaginaceae transcriptomes in the NCBI Short Read Archive (**Supplementary Table 4**) using BLASP with *Arabidopsis* sequences as queries. Matches were aligned using mafft (Katoh and Standley, 2013) and preliminary trees

[4]http://phytozome.jgi.doe.gov/

[5]http://tree.bio.ed.ac.uk/software/figtree

were inferred with FastTree (Price et al., 2010) in order to identify the number of copies present in Nyctaginaceae. For each copy, sequences from the Nyctagineae clade were extracted from the alignment and a consensus sequence was generated using Geneious (www.geneious.com) to create full-length consensus references for each copy onto which RNA-seq data for all Nyctagineae species in the NCBI SRA were mapped using HiSat2 (Kim et al., 2019) with the scoring argument – score-min L,0,-1.4. Read mappings were scrutinized manually using the software Geneious to be certain there were no additional paralogs beyond those we had detected, which would be evident as paralogous reads mapped onto the closest reference. Consensus sequences based on mapped reads were generated in Geneious. Gene trees were inferred using MrBayes (Huelsenbeck and Ronquist, 2001) with two runs, 40 chains, 2 million generations, and a heating factor of 0.05. About 10,000 trees from the end of each run where the average SD of split frequencies remained flat at or below 0.01, were used to generate the final tree and infer posterior probabilities using the consense program included with ExaBayes (Aberer et al., 2014). Trees were visualized using FigTree.[5] Gene expression values were calculated as reads per kb of transcript per million reads (RPKM) using the SAM files produced by HiSat and a Python script. Only species from the 1KP project[6] submitted by our lab were used for gene expression as these represent both C_4 lineages and a closely-related C_3 outgroup, and were grown and sampled identically. For these, the newest fully expanded leaves were sampled during a sunny day from plants grown in a glass house at the University of Toronto between 9 am and 1 pm. Because complete mappings to all transcripts were not conducted, numbers of total sequenced reads were used in RPKM calculations rather than total mapped reads. Scripts written by M Stata are available at github.com/MattStata/Nyctaginaceae_Scripts.

Species Distribution Data

Biogeographic distributions for species of the tribe Nyctagineae were obtained from the Global Biodiversity Information Facility website (GBIF).[7] Duplicate data points and those lacking herbarium records or corresponding to marine coordinates were removed (maptools, Bivand and Lewin-Koh, 2013). The remaining 15,870 observations represented 75 species within *Allionia* (two species), *Boerhavia* (40 species), *Anulocaulis* (five species), *Commicarpus* (25 species), *Cyphomeris* (two species), and *Nyctaginia* (one species). Bioclimate and monthly minimum and maximum temperature parameters (2.5 min resolution) were downloaded from the WorldClim 2.0 dataset[8] (Fick and Hijmans, 2017). Monthly potential evapotranspiration and Global Aridity Indexes (AI) were downloaded from CGIAR-CSI GeoPortal.[9] Values per observation for 19 bioclimatic variables in the Worldclim dataset, plus minimum and maximum temperatures and AI were then extracted using the extract function in the R raster package (Hijmans and van Etten, 2012). Median values per variable per species were calculated and normalized to Z-scores for use in subsequent analyses. Climate variables that significantly predicted the occurrence of photosynthetic type at $p < 0.05$ were selected using stepwise regression. Mixed-effect models (R package lme4, Bates et al., 2015) were built using the selected bioclimatic variables and photosynthetic subtypes (NAD-ME or NADP-ME) as the main effect and genus and species as random effects. These models were compared by ANOVA and Akaike's Information Criteria (AIC). The best model was selected based on the lowest AIC and p values <0.05. A principal component analysis was performed by R package FactoMineR (Lê et al., 2008) to evaluate species distribution across the multivariate predictors. A subset of the data for which we had phylogenetic data was also used to run a phylogenetically corrected ANOVA using phytools in R (Revell, 2012).

[6]www.onekp.com

[7]https://www.gbif.org/occurrence/

[8]https://www.worldclim.org/

[9]https://cgiarcsi.community

RESULTS

Carbon Isotope Ratios

For the $\delta^{13}C$ survey, we sampled herbarium specimens from all genera of the Tribe Nyctagineae, except for one monospecific genus from El Salvador (*Cuscatlainia vulcanicola*; **Table 1**). All sampled species from the genera *Allionia, Boerhavia,* and *Okenia* exhibited C_4 $\delta^{13}C$ values (−9 to −15‰) (**Table 1** and **Supplementary Tables 2,3**). All other species exhibited C_3 $\delta^{13}C$ values (−22 to −30‰), including all assayed species in the Nyctagineae genera *Abronia, Acleisanthes, Anulocaulis, Commicarpus, Cyphomeris, Mirabilis*, and *Tripterocalyx.* The $\delta^{13}C$ survey provides little evidence for C_3–C_4 intermediate species in the Nyctagineae. Values from all species were clearly C_4 or within the more negative range of $\delta^{13}C$ values typical of C_3 plants, with two possible exceptions – the arid zone species *Acleisanthes angustifolia* ($\delta^{13}C = -23.9$) and *Mirabilis polyphylla* ($\delta^{13}C = -22.6$). While high for a typical C_3 $\delta^{13}C$ value, these exceptions are within the range of values observed in arid-zone C_3 species with high WUE (Farquhar et al., 1989).

Gas Exchange and Biochemistry

Both *Allionia incarnata* and *B. coccinea* exhibited typical C_4 photosynthetic parameters. The CO_2 compensation point of photosynthesis (Γ) was below 5 μmol mol^{-1} in both species, the ratio of intercellular to ambient CO_2 concentration (C_i/C_a) ratio was 0.31–0.35, and their carboxylation efficiency of photosynthesis, measured as the initial slope of the photosynthetic response to intercellular CO_2 concentration (A/C_i response), was 5–7 times greater than the carboxylation efficiency in the C_3 *N. capitata* (**Figure 2** and **Table 2**). *Allionia* exhibited a steeper initial slope of the A/C_i response than *Boerhavia.* When relativized by dividing by the maximum net CO_2 assimilation rate at high CO_2 (A_{1200}) to correct for variation in photosynthetic capacity, the normalized carboxylation efficiency in *Allionia* was 50% greater than in *Boerhavia* (**Table 2**).

The *in vitro* activity of PEPC was high in the C_4 species relative to the C_3 *N. capitata*, and similar on a leaf area basis in the two C_4 species; however, on a chlorophyll basis, the PEPC activity is 63% higher in *Allionia* than *Boerhavia* (**Table 2**).

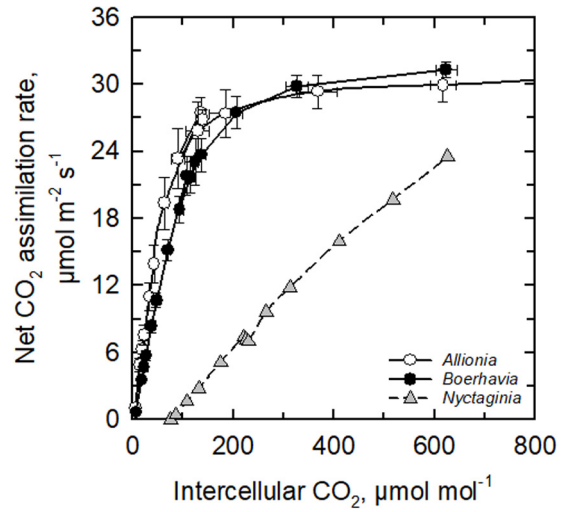

FIGURE 2 | The response of the net CO_2 assimilation rate to intercellular CO_2 concentration in *Allionia incarnata* (C_4), *Boerhavia coccinea* (C_4) and *Nyctaginia capitata* (C_3) at 30°C and a light intensity of 1800 µmol photons m^{-2} s^{-1}. Means ± SE, N = 4 (for the C_4 species) or 1 (for the C_3 species).

NADP-ME activity was high in *Boerhavia* and not detected in *Allionia*, while NAD-ME activity was negligible in *Boerhavia* and high in *Allionia* (**Table 2**). Chlorophyll content was 43% higher in *Boerhavia* than *Allionia*, and chlorophyll *a/b* ratio was 21% higher in *Boerhavia*. Observed enzyme activities are considered robust because they are equivalent to or greater than the observed *A* values in each species. From these results, we conclude that *Boerhavia* belongs to the NADP-ME subtype, while *Allionia* is of the NAD-ME subtype.

Leaf Structure and Ultrastructure in *Allionia*, *Boerhavia* and Two *Portulaca* Species

Allionia incarnata and *B. coccinea* exhibited typical C_4-Atriplicoid Kranz anatomy, which is characterized by large BS cells in planar cross sections with centripetal organelle arrangements (**Figures 3**, **4**; Edwards and Voznesenskaya, 2011). Notably, enlarged BS cells in both *Allionia* and *Boerhavia* do not wrap around the entire vascular bundle, mimicking patterns observed in *Atriplex* but not many other C_4 eudicots classified as having Atriplicoid Kranz anatomy (Muhaidat et al., 2007). *A. gypsogenus*, *C. scandens*, *N. capitata*, and *Mirabilis jalapa* have characteristic C_3 leaf anatomy (**Figures 3**, **4** and **Supplementary Figure 1**). Planar areas of the BS cells are comparable between the C_3 and C_4 species, indicating similar dimensions of the BS tissue in a radial direction with respect to the vasculature (**Table 3**). Both C_4 species exhibit greater coverage of the BS cell by chloroplasts than the C_3 species, because there are more chloroplasts per BS cell and greater mean area per chloroplast in the C_4 species (**Figure 4**; **Table 3**). In *A. incarnata*, the planar area of BS cells covered by mitochondria is significantly greater than in C_3 species and the C_4 *B. coccinea*, due to the presence of larger mitochondria in planar section, and more mitochondria per cell area (**Table 3**; **Figures 4A,C**). The percent mitochondrial

TABLE 2 | Summary of gas exchange and biochemical data for three Nyctaginaceae species.

Parameter	*Nyctaginia capitata*	*Boerhavia coccinea*	*Allionia incarnata*	C_4 *p* value (one-tailed)
Net CO_2 assimilation rate at 400 µmol CO_2 mol^{-1} air, in µmol m^{-2} s^{-1}	7.0	23.0 ± 2.1	26.8 ± 0.9	0.08
Net CO_2 assimilation rate at 1200 µmol CO_2 mol^{-1} air, in µmol m^{-2} s^{-1}	23.5	31.3 ± 0.7	30.5 ± 1.5	0.32
A_{400}/A_{1200}	0.30	0.74 ± 0.06	0.88 ± 0.03	**0.04**
A/g_s, mmol CO_2 mol^{-1} H_2O	100	158 ± 3	149 ± 5	0.10
C_i/C_a	0.58	0.31 ± 0.01	0.35 ± 0.02	0.10
Initial slope of the *A vs* C_i curve, mol m^{-2} s^{-1}	0.05	0.25 ± 0.03	0.36 ± 0.04	**0.04**
Initial slope/A_{1200}	0.002	0.008 ± 0.001	0.012 ± 0.001	**0.03**
CO_2 compensation point of *A*, µmol mol^{-1}	75	4.5 ± 0.5	2.5 ± 1.0	0.06
PEP carboxylase activity, µmol m^{-2} s^{-1}	15.1 ± 1.9	111.8 ± 7.9	131.7 ± 22.2	0.22
PEP carboxylase activity, mmol mol^{-1} CHL s^{-1}	27.0 ± 2.7	261.9 ± 12.4	429.7 ± 53.2	**0.01**
NADP-ME activity, µmol m^{-2} s^{-1}	0 ± 0	50.1 ± 5.6	0 ± 0	**< 0.001**
NADP-ME activity, mmol mol^{-1} CHL s^{-1}	0 ± 0	116.7 ± 8.9	0 ± 0	**< 0.001**
NAD-ME activity, µmol m^{-2} s^{-1}	0 ± 0	2.1 ± 0.6	27.7 ± 2.8	**< 0.001**
NAD-ME activity, mmol mol^{-1} CHL s^{-1}	0 ± 0	5.1 ± 1.4	91.5 ± 6.4	**< 0.001**
Chlorophyll, mmol m^{-2}	0.56 ± 0.02	0.43 ± 0.02	0.30 ± 0.01	**< 0.001**
Chlorophyll a/b ratio	3.46 ± 0.07	4.45 ± 0.09	3.67 ± 0.08	**< 0.001**

Means ± SE, N = 4 for the C_4 species, 1 for N. capitata gas exchange data, and 3 for N. capitata biochemical data. The p-values of one-tailed Students T-tests between the two C_4 species are indicated, with significant differences highlighted in bold. Gas exchange measurements were conducted at 33°C and a light intensity of 1800 µmol m^{-2} s^{-1}. Biochemical assays were conducted at 30°C. A, net CO_2 assimilation rate; C_i, intercellular CO_2 concentration; C_a, ambient CO_2 concentration in the leaf chamber; A_{400}, A at a CO_2 concentration of 400 µmol CO_2 mol^{-1} air; A_{1200}, A at a CO_2 concentration of 1200 µmol mol^{-1}; CHL, leaf chlorophyll content; g_s, leaf conductance to water vapor; ME, malic enzyme.

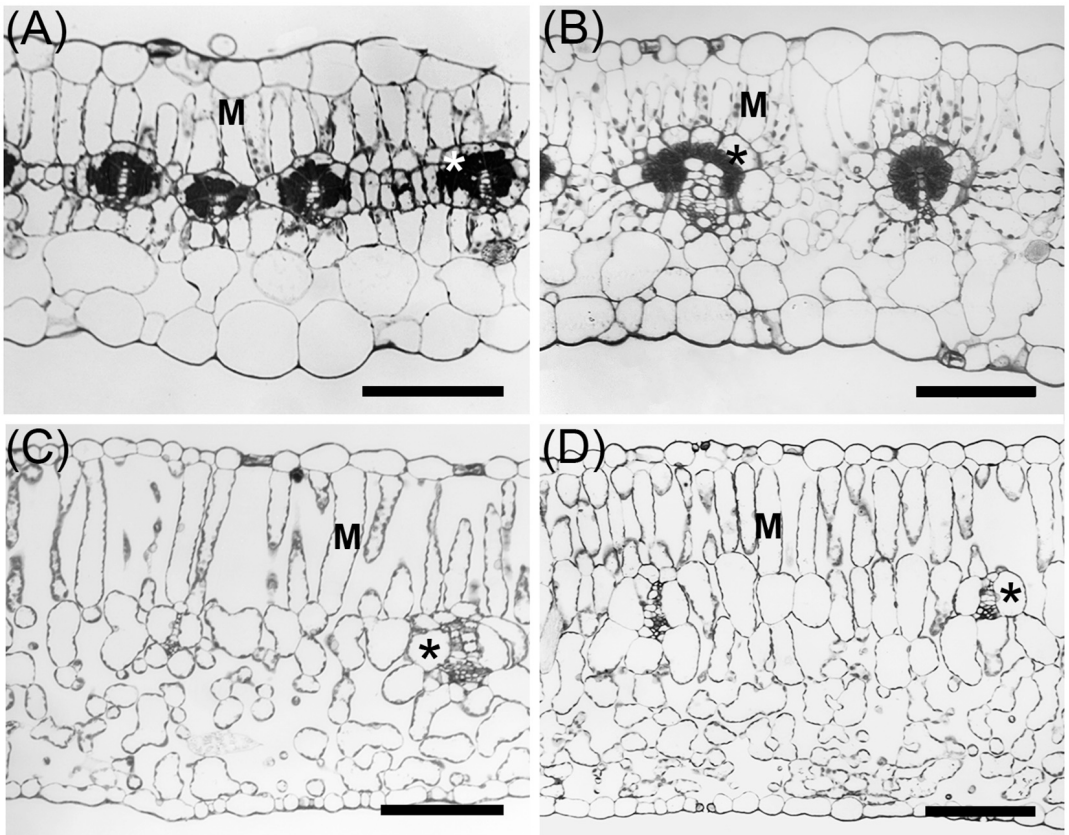

FIGURE 3 | Light microscope images of leaves from C_3 and C_4 species of Nyctaginaceae, tribe Nyctagineae. **(A)** *Allionia incarnata*, **(B)** *Boerhavia coccinea*, **(C)** *Commicarpus scandens*, **(D)** *Nyctaginia capitata.* M, mesophyll cells; *Asterisks mark bundle sheath cells. Scale bars = 100 μm.

coverage of the BS cells is similar in *B. coccinea* and the C_3 species (**Table 3**; **Figures 4B,D–F**).

The BS cells in species with C_3 $\delta^{13}C$ values have smaller and fewer chloroplasts than the C_4 species of the Nyctagineae, and these are positioned mostly along the outer BS wall exposed to intercellular airspace (**Figures 4E,F** and **Supplementary Figure 1**). None of the three C_3 species examined exhibit traits associated with C_3–C_4 intermediacy or even the incipient intermediate state termed "proto-Kranz"; there was not a greater mitochondrial number, nor a repositioning of mitochondria and chloroplasts toward the inner side of the BS cells. Chloroplasts in the BS of *Boerhavia* and the M tissue of *Allionia* are largely agranal, whereas distinct levels of grana stacking are apparent in the thylakoids of the BS chloroplasts in *Allionia*, and the M chloroplasts of *Boerhavia* (**Figure 5**)

In *P. oleracea* (NAD-ME), chloroplasts with pronounced grana stacks cluster in the inner BS (**Figure 6A**). Many mitochondria are interspersed between the chloroplasts, and exhibit distinct structural connections with nearby chloroplasts (**Figure 6B**). In *P. pilosa* (NADP-ME), chloroplasts also occur in the inner BS where they form clusters with no discernable thylakoid stacks (**Figures 6C,D**). Mitochondria are largely absent between chloroplasts, although some mitochondria occur along the inner wall of the BS in a pattern that is commonly observed in C_3–C_4 intermediate species (**Figure 6C**).

Species Phylogeny

We update the molecular phylogeny of the Nyctagineae using transcriptome sequence data available at the NCBI short read archive (**Figure 7**). The tree largely replicates the phylogenies of Douglas and Manos (2007) and Douglas and Spellenberg (2010) showing the C_4 and C_3 clades of the Nyctagineae correspond to a clade of xerophytic herbs and shrubs that Douglas and Spellenberg term the North American Xeric (NAX) clade. C_4 *Allionia* branches with *Cyphomeris* at the base of a clade that includes the isotopically C_3 *Anulocaulis* and *Nyctaginia* species, and the C_4 *Okenia* and *Boerhavia* species. *Anulocaulis* and *Nyctaginia* form a clade that branches between *Allionia/Cyphomeris* and *Boerhavia/Okenia. C. scandens* branches at the base of the clade containing the C_4 species and their immediate non-C_4 sister clades, with *Mirabilis* species branching just below *C. scandens*. Unlike Douglas and Spellenberg (2010) who predicted Bougainvilleeae + Pisonieae to be sister to the Nyctagineae, we predict only Bougainvilleeae branches in a sister position, with maximal support. Also, the branching order of Phytolaccaceae, Sarcobataceae, Gisekiaceae, and Nyctaginaceae is unclear in the literature and in the APG

Caryophllalyes tree, where they form a polytomy.[10] Our tree resolves them with 100% support, improving the phylogenetic context for the family.

Gene Phylogenies

We examined the trees of four important C_4 cycle genes – PEPCase, NADP-ME, NAD-ME, and PPDK (**Figure 8** and **Supplementary Figures 2–5**). For each gene, we assumed that the functional copy in the C_4 pathway was the one with the highest expression in the transcriptome analysis. In this analysis, we numbered gene copies based on branching order from the base of the gene tree; the numbers assigned are not meant to imply direct orthology with gene copies from any other lineage. For *Allionia incarnata* and two *Boerhavia* species, the *PEPC1* gene is the copy used in the C_4 pathway because it shows an expression strength that is orders of magnitude greater than *PEPC2 and PEPC3* (**Table 4**). By similar logic, *NAD-ME3* is the main decarboxylase gene in the C_4 pathway of *Allionia*, and *NADP-ME2* is the main decarboxylase gene in *Boerhavia. Allionia* also exhibited significant expression of *NADP-ME1,* approaching that of *NAD-ME3* (**Table 4**). The expression of the PEP carboxykinase gene was minimally detectable in both C_4 clades (not shown). In the gene trees for *PEPC1*, *PPDK* and *NADP-ME1*, the respective orthologs from numerous C_3 species of the Nyctagineae branch between the two C_4 lineages, with strong support (**Figure 8** and **Supplementary Figures 2–5**). This indicates the C_4 genes arose independently from an ancestral C_3 copy, rather than by lateral transfer between clades. If one or more of the genes had moved laterally between the C_4 clades, the orthologs from both C_4 lineages would form a common clade.

[10]www.mobot.org/MOBOT/research/APweb

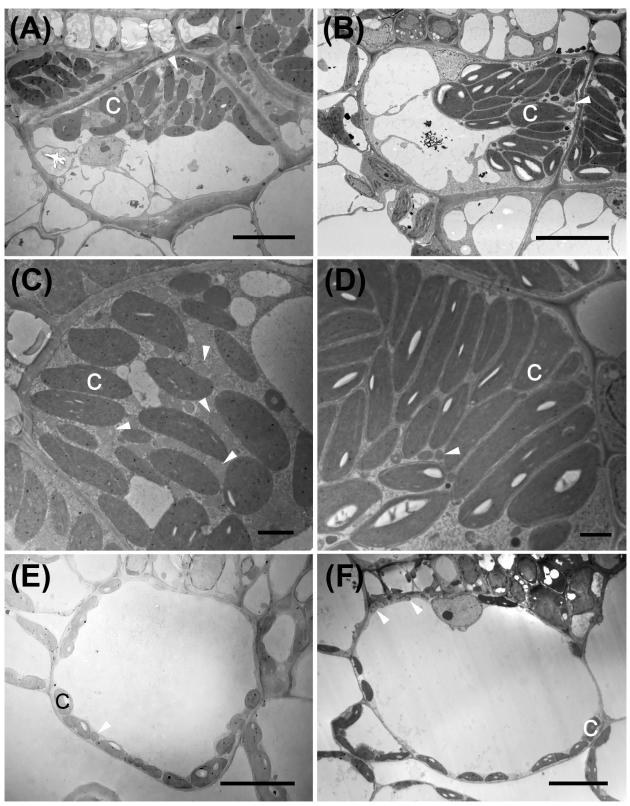

FIGURE 4 | Transmission electron micrographs illustrating bundle sheath cells of C_3 and C_4 Nyctaginaceae species from the tribe Nyctagineae. **(A–D)** C_4 species; **(A,C)** *Allionia incarnata,* **(B,D)**; *Boerhavia coccinea*; **(C,D)** high magnification images of the bundle sheath cells. Note high numbers of mitochondria in *Allionia incarnata* **(C)**. **(E,F)** The C_3 species *Commicarpus scandens* **(E)**, and *Nyctaginia capitata* **(F)**. C, chloroplast; arrowheads mark mitochondria. Scale bars = 10 μm.

Positively Selected Amino Acid Substitutions

We examined *PEPC1* sequences for evidence of convergence in the C_4 isoforms (**Table 5**). Christin et al. (2008) showed that 21 amino acid sites were under positive selection in the *PEPC1* of various C_4 grasses, with the best-known example of sequence convergence being a serine for alanine substitution near the maize 780 position in the PEPC1 sequence (position 774 in the *Flaveria* sequence; Gowik and Westhoff, 2011). We did not find evidence for convergence with other C_4 lineages at position 780 in *Allionia* nor *Boerhavia*, as they both exhibited an alanine, which is typical of C_3 isoforms (**Table 5** and **Supplementary Figure 6**). There was consistent convergence at sites 572, 761, and 807, where all C_4 species exhibited the same amino acids as those under positive selection in C_4 grasses. Convergence was observed in both C_4 lineages at site 813, but in only one of the six *Boerhavia/Okenia* species in the database (**Table 5**). *Boerhavia* also converged on the same amino acids as the C_4 grass species at site 733 and 863, and partially at position 502 (*B. purpurescens and B. torreyana* only). In total, of the 21 C_3 to C_4 amino acid switches repeatedly observed in the grass PEPCase sequences (Christin et al., 2007), about a third were replicated in the *Boerhavia* lineage and a fifth in the *Allionia* lineage (**Table 5**). At one site, position 577, the C_3 and C_4 Nyctaginaceae species share a serine with the C_4 grass *Zea mays*. In the grasses, most C_3 species examined by Christin et al. (2007, 2012b) exhibit an alanine at this site. It is possible that C_4 evolution in Nyctaginaceae did not require this substitution as the ancestral C_3 state was already a serine.

Ecological Distribution

Nyctagineae species from the genera *Allionia, Boerhavia, Anulocaulis, Commicarpus, Cyphomeris*, and *Nyctaginia* are distributed in tropical and subtropical regions of similar climate (**Supplementary Figures 7–9**). Six bioclimatic variables significantly predicted ($p < 0.05$) the C_3 and C_4 distribution using a stepwise regression model (**Supplementary Figure 8** and **Supplementary Table 6**). To address whether the variation in distribution of C_3 and C_4 species could be species or genera dependent, GLMM models were performed with and without considering taxa as random effects. The model was significantly improved ($p < 2e^{-16}$) when genus was added as the random effect (**Supplementary Table 6**); however, after this addition, none of the main effects significantly predicted the pattern of photosynthetic pathway distribution.

TABLE 3 | Ultrastructural parameters for the bundle sheath cells in cross section of two C_4 species and three C_3 species from the Nyctagineae tribe of the Nyctaginaceae.

	A. incarnata	*An. gypsogenus*	*B. coccinea*	*C. scandens*	*N. capitata*
	C_4	C_3	C_4	C_3	C_3
Cell area, μm^2	427 ± 169 a	727 ± 339 a	557 ± 183 a	505 ± 199 a	944 ± 389 a
% Cell covered by chloroplast,%	28.3 ± 5.9 b	2.3 ± 0.9 a	35.2 ± 11.5 b	13.8 ± 4.6 a	6.5 ± 4.2 a
% Cell area covered by mitochondria,%	3.6 ± 1.7 b	0.4 ± 0.2 a	1.2 ± 0.5 a	0.6 ± 0.3 a	0.5 ± 0.4 a
Number of chloroplasts per cell	16 (5–24) b	5 (2–11) a	18 (3–37) c	10 (7–15) b	10 (5–17) b
Number of mitochondria per cell	45 (12–111) c	14 (4–57) a	28 (4.8–75) b	13 (5–30) a	11 (7–52) a
Area per chloroplast, μm^2	8.3 ± 3 c	2.9 ± 1.1 ab	11.4 ± 3.8 c	6.1 ± 2.1 bc	5.3 ± 1.2 bc
Area per mitochondria, μm^2	0.7 ± 0.2 b	0.3 ± 0.1 a	0.4 ± 0.1 a	0.5 ± 0.2 ab	0.3 ± 0.2 a

Means ± SE, or median ± range for organelle numbers. N = 3–5. Letters besides values indicate statistically distinct groups by Kruskal Wallis test followed by a Dunn's test except for organelle numbers which are evaluated using a Poisson regression for count data (number of organelles per cell). Abbreviations: A, Allionia; An, Anulocaulis; B, Boerhavia, C, Commicarpus; N, Nyctaginia.

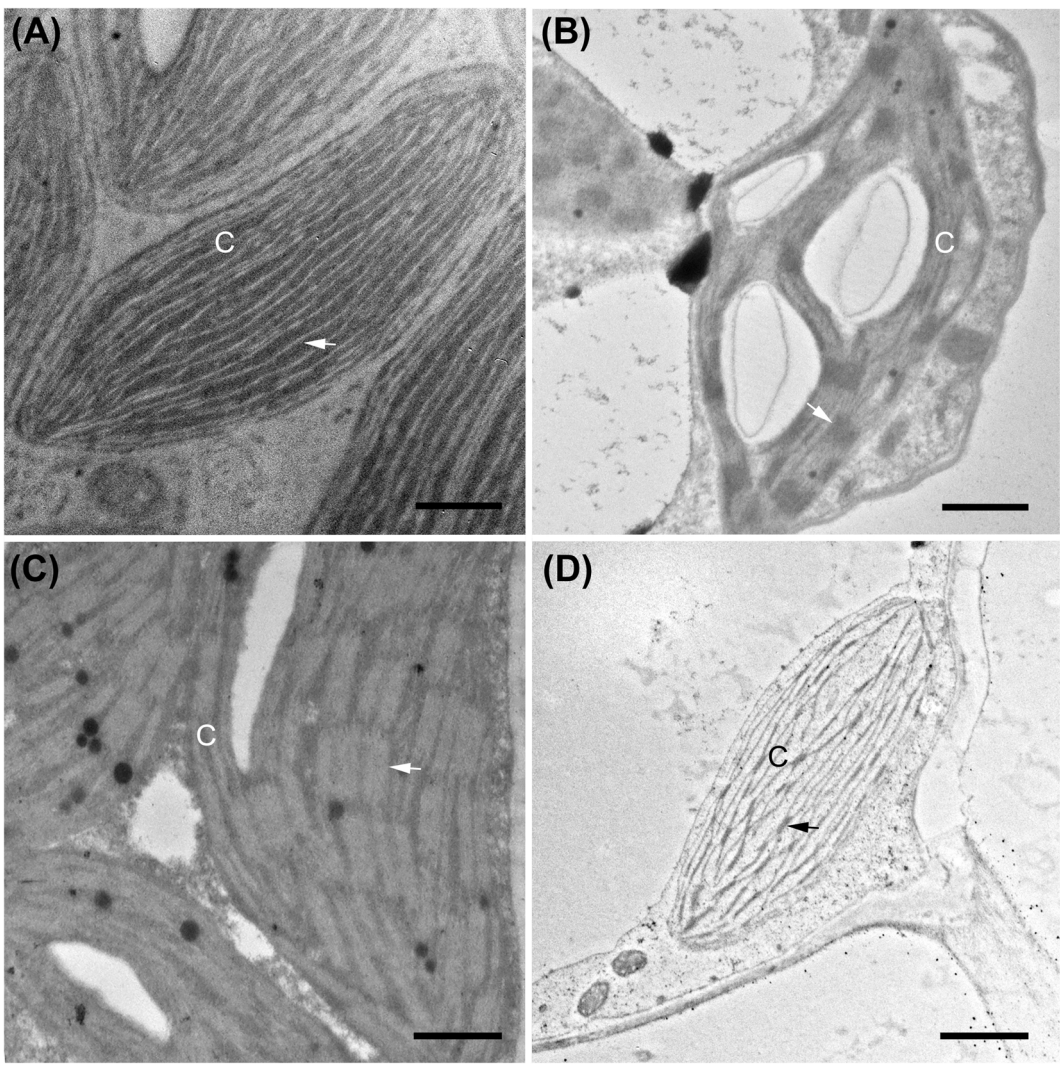

FIGURE 5 | Transmission electron micrographs illustrating chloroplast fine structure. **(A)** *Boerhavia burbidgeana* bundle sheath cell; **(B)** *B. burbidgeana* mesophyll cell; **(C)** *Allionia incarnata* bundle sheath cell; **(D)** *A. incarnata* mesophyll cell. In panels **(A,C)**, thylakoids are lighter stacks and striations against a dark stroma. In panels **(B,D)**, thylakoids are dark stacks and striations against a lighter-staining stroma. Scale bars, 0.5 μm. C, chloroplast; arrows mark stacked thylakoids. Scale bars = 0.5 μm.

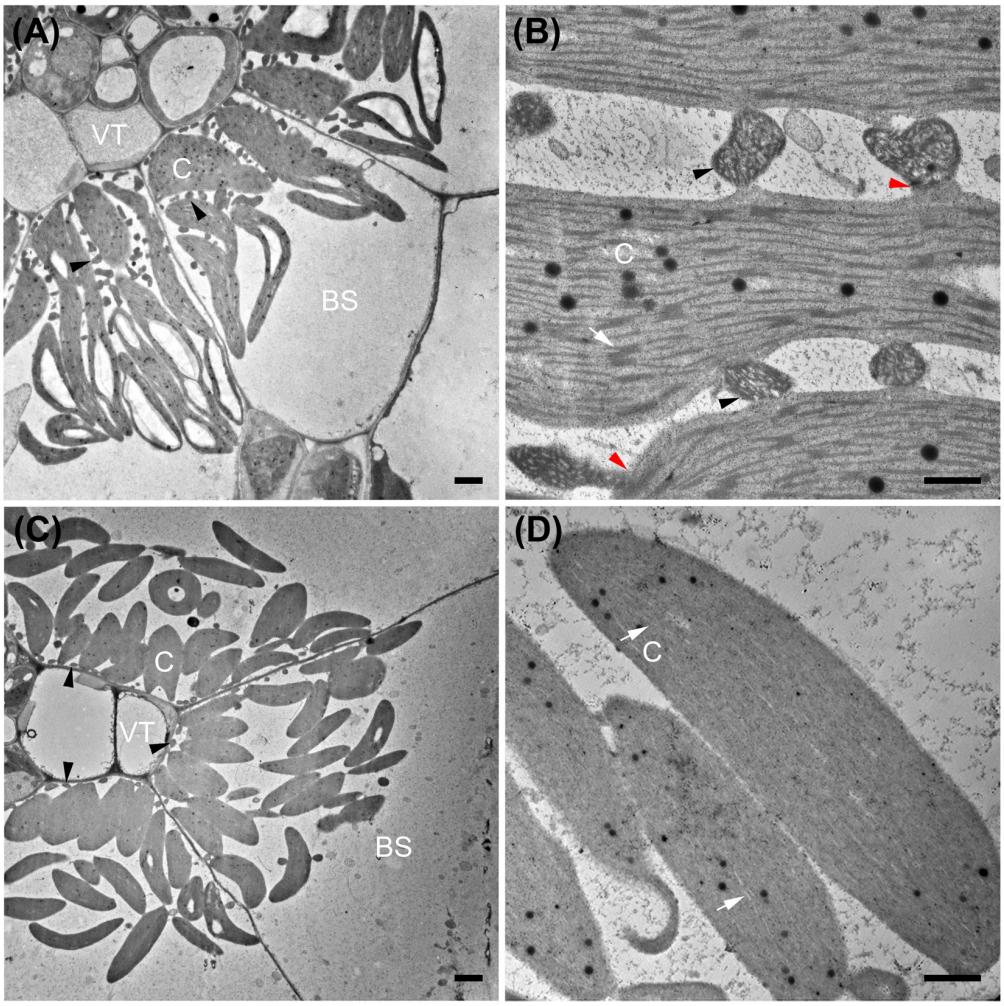

FIGURE 6 | Transmission electron micrographs of bundle sheath cells **(A,C)** and bundle sheath chloroplasts **(B,D)** from NAD-ME subtype *Portulaca oleracea* **(A,B)** and NADP-ME subtype *Portulaca pilosa* **(C,D)**. Red arrowheads mark mitochondria to chloroplasts connection structures in *P. oleracea* **(B)**. BS, bundle sheath; C, chloroplast; VT, vascular tissue; black arrowheads mark mitochondria; white arrows mark thylakoids. Bars = 2 μm **(A,C)** or 0.5 μm **(B,C)**.

TABLE 4 | RNA Transcript Expression of C_4 cycle enzymes from four selected species of the Nyctaginaceae.

	PEPC1	PEPC2	PEPC3	NAD1	NAD2	NAD3	NADP1	NADP2	NADP3
Mirabilis jalapa (C_3)	164	19	105	50	28	4	122	4	75
Allionia incarnata (C_4)	12069	33	88	80	25	2001	1751	114	13
Boerhavia burburgiana (C_4)	17742	1	144	56	6	0.4	8563	2	144
Boerhavia coccinea (C_4)	9331	0.1	128	85	27	0.2	3562	2	121

*Yellow shading indicates the expression values of gene copies that are predominant in the C_4 metabolic cycle. Green shading indicates enhanced expression of genes for which the activity in vitro was negligible. PEPC, PEP carboxylase; NADP, NADP malic enzyme; NAD, NAD-malic enzyme. Data were generated from transcriptomes in the 1KP database (www.onekp.com) and are available in the NCBI Short Read Archive (**Supplementary Table 4**). PEP carboxykinase transcript levels were found to be negligible in each species.*

A phylogenetically-corrected ANOVA also failed to detect a significant difference between the C_3 and C_4 species in response to six variables (**Supplementary Table 7**). These analyses show that the distribution of Nyctagineae species in climate and geographic space follows taxonomic and phylogenetic affinities rather than photosynthetic pathway.

DISCUSSION

Carbon isotope ratios demonstrate that all examined species in *Allionia*, *Boerhavia* and *Okenia* are C_4, while all examined species in other Nyctaginaceae genera are not, including species in *Anulocaulis*, *Commicarpus*, *Cyphomeris*, and *Nyctaginia* that

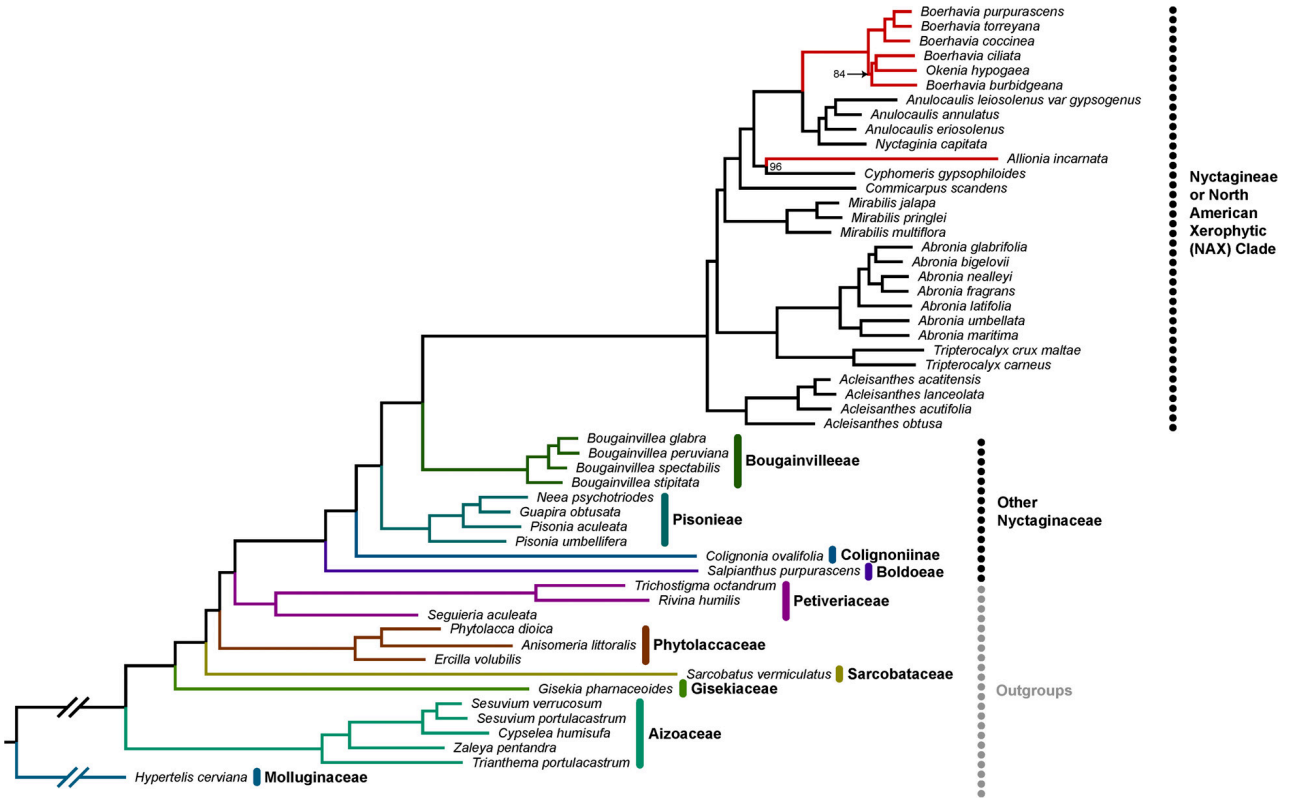

FIGURE 7 | Phylotranscriptomic reconstruction of Nyctaginaceae. A maximum likelihood phylogeny of the Nyctaginaceae family and six outgroup families with previously uncertain topology. This tree is based on a concatenated super-matrix of 847 codon-aligned single-copy orthologs totaling 834,018 sites. All bootstrap support values are 100% except for the two nodes with values indicated, based on 200 bootstrap replicates. Within the Nyctagineae (identified as the North American Xerophytic clade by Douglas and Spellenberg, 2010), C_4 lineages are denoted with red branches. The tree is rooted on *Hypertelis cerviana* and gapped branches here were shortened for visual convenience.

TABLE 5 | Comparison of amino acids in the C_4 PEP carboxylases of the *Allionia* and *Okenia/Boerhavia* C_4 clades with PEP carboxylase sites under positive selection in C_4 grasses.

	Sites with posterior probability > 0.999												Prob > 0.99			Prob > 0.95			
Clade	**466**	**517**	**531**	**560**	**577**	**579**	**625**	**637**	**761**	**780**	**794**	**807**	**572**	**599**	**813**	**665**	**733**	**863**	**866**
Boerhavia/Okenia	Met	Thr	Ala	Arg	Ser	Ala	Leu	Met	Ala	Ala	Phe	Lys	Gln	Ile	Arg, Gln*	His	Met	Lys	Glu
Allionia	Met	Thr	Ala	Arg	Ser	Ala	Val	Met	Ala	Ala	Phe	Lys	Gln	Ile	Gln	His	Phe	Asn	Glu
C_3 Nyctag	Met	Thr	Ala	Arg	Ser	Ala	Val	Met	Ser	Ala	Phe	Arg	Glu	Ile	Arg	His	Phe	Asn	Glu
C_4 grasses (Christin et al. (2007)	Ile, Val	Ala, Cys	Pro	Pro	Ser	Glu	Ala	Leu, Phe	Ala	Ser	Val	Lys	Gln	Val	Gln	Asn	Met, Val	Lys	Glu

Site numbers correspond to the maize PEP carboxylase gene (PEPC). Sites under positive selection were identified by a screen of grass lineages examined by Christin et al. (2007). Posterior probabilities noted with blue headers are from that study. Sites showing an identical amino acid substitution in either or both C_4 Nyctaginaceae lineages are highlighted in yellow. We list the most frequent positively selected amino acid(s) at each site in C_4 grasses. The species in the C_3 Nyctag clade are the North American Xerophytic C_3 species in the PEPC1 gene tree in ***Supplementary Figure 2****, which exhibited identical amino acids at the sites shown. *Gln is only present in B. burbidgeana; the other Boerhavia and Okenia have Arg at position 813.*

branch sister to the C_4 clades. The number of C_4 species in the Nyctaginaceae is 43–46 based on recent assessments that conclude there are 40 *Boerhavia*, one to four *Okenia*, and two *Allionia* species (Spellenberg, 2003; Tropicos, 2020). The typically C_3 $\delta^{13}C$ values in the 31 examined species of *Anulocaulis*, *Commicarpus*, *Cyphomeris*, and *Nyctaginia* indicate that none of the examined species have a type of C_3–C_4 intermediacy termed C_4-like, where a strong C_4 metabolic cycle has been engaged. As a C_4 cycle becomes engaged, the values of $\delta^{13}C$ increase from C_3 toward C_4 values, becoming distinct from typical C_3 values when $\delta^{13}C$ rises above −22‰ (Von Caemmerer, 1992; Alonso-Cantabrana and von Caemmerer, 2016). This occurs because PEPCase does not discriminate against the ^{13}C isotope to the degree that Rubisco does (Farquhar et al., 1989). Carbon isotope screens cannot differentiate between C_3 plants and a type of C_3–C_4 intermediacy utilizing C_2 photosynthesis,

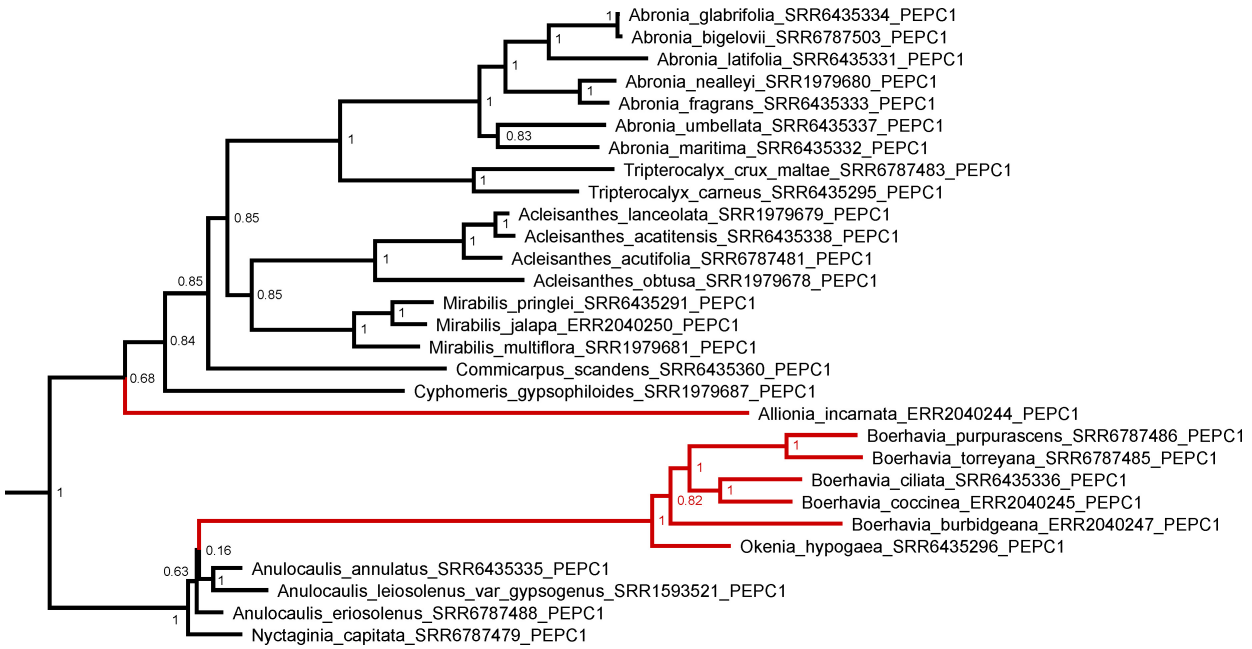

FIGURE 8 | A Baysian gene phylogeny for the PEPCase gene copy co-opted for C_4 function in the Nyctaginaceae. Node values are posterior probabilities. Red branches denote C_4 species. A tree with all three PEPC paralogs is presented in **Supplementary Figure 2**.

a metabolic pathway that concentrates CO_2 into the BS using photorespiratory metabolites to shuttle CO_2 from M to BS cells. In C_2 photosynthesis, the photorespiratory enzyme glycine decarboxylase is expressed only in BS mitochondria, which forces photorespiratory glycine to diffuse from the M cells where it is formed to the BS cell for metabolism (Sage et al., 2012). The released CO_2 from glycine decarboxylation accumulates in the BS cells, where it can be refixed by Rubisco in adjacent chloroplasts with high efficiency. C_2 photosynthesis improves carbon gain at low CO_2 concentrations, but because all CO_2 is fixed by Rubisco, the $\delta^{13}C$ values of C_2 plants reflect those of C_3 plants (Von Caemmerer, 1992). Anatomically, C_2 plants have increased numbers of mitochondria and chloroplasts in BS cells, typically in a centripetal position against the BS wall facing the vascular tissue (Khoshravesh et al., 2016). This characteristic identifies candidate C_2 species, with low CO_2 compensation points of photosynthesis (Γ) confirming the presence of the C_2 physiology. Our examination of the BS structure in *Anulocaulis*, *Commicarpus* and *Nyctaginia*, and leaf gas exchange in *Nyctaginia*, showed no evidence of C_2 photosynthesis as the structural characteristics and Γ were typically C_3. Hence, we conclude that the sister clades to the *Allionia* and *Boerhavia/Okenia* clades are most likely comprised of only C_3 plants. Gas exchange data and anatomical studies demonstrate strong C_4 features in selected species of *Allionia* and *Boerhavia*, which along with the consistently C_4 values of $\delta^{13}C$ in these clades, lead us to conclude they are completely C_4 rather than C_4-like. C_4-like species have a fully functional C_4 cycle but retain limited C_3 photosynthesis in M tissues, and exhibit features suggesting they are newly evolved C_4 plants that have not yet optimized the C_4 pathway (Moore et al., 1989).

In the phylogeny, *Anulocaulis* and *Nyctaginia* species occur in a clade that branches sister to the *Boerhavia/Okenia* clade, while *Cyphomeris* branches sister to *Allionia*, supporting a hypothesis of two independent C_4 origins, one in ancestral *Allionia* and a second in ancestors to the *Okenia/Boerhavia* clade. An alternative possibility is a single C_4 origin followed by reversions to C_3 in ancestors of *Cyphomeris* and the *Anulocaulis/Nyctaginia* clade. This option is less parsimonious, requiring three changes – one acquisition of C_4 and two reversions – rather than two. It is also not favored because reversions are considered unlikely and have never been confirmed (Christin et al., 2010a; Oakley et al., 2014; Bräutigam and Gowik, 2016). In addition, the two C_4 clades exhibit different biochemical and structural subtypes. The high degree of structural, transport, and biochemical specialization associated with each subtype indicates switching subtypes would be difficult, and consistently, no instances of subtype switching have been documented. Gene sequences of C_4 pathway enzymes can also be used to assess reversal possibilities. If a reversion had occurred, the derived ortholog of a C_4-pathway enzyme in any C_3 progeny might retain sequences from their ancestral C_4 function (Christin et al., 2010a; Khoshravesh et al., 2020). In *Anulocaulis*, *Cyphomeris*, and *Nyctaginia*, the C_4-type orthologs of PEPCase, PPDK, and both decarboxylases exhibited no signatures of prior C_4 function. For example, the C_3 species in the Nyctagineae share only one positively selected amino acid substitution with the C_4 Nyctaginaceae and *Zea mays* in the PEPC1 sequence, a serine at position 577. C_3 species outside the clade where an ancestral C_4 transition could have occurred also share this serine,

indicating it is not a relic from a C_4-state. Based on these points, we conclude that *Allionia* and *Boerhavia* represent independent yet closely related C_4 clades that have diverged in terms of biochemical subtype. With other examples of subtype divergence in closely related species of *Portulaca* and possibly other clades (*Salsola* and Sesuvioideae; Voznesenskaya et al., 1999; Bohley et al., 2015), comparative methods of evolutionary biology can be used to address questions of convergence and divergence during C_4 evolution.

In contrast to a prior conclusion that both *Boerhavia* and *Allionia* are NADP-ME (Muhaidat et al., 2007), the biochemical assays demonstrate species in these two clades are different biochemical subtypes of C_4 photosynthesis. Consistently, *Allionia* and *Boerhavia* showed pronounced differences in BS and M ultrastructure that support their designation as NAD-ME and NADP-ME subtypes, respectively. In both *B. coccinea* (high NADP-ME activity, low NAD-ME activity) and *A. incarnata* (high NAD-ME activity, nil NADP-ME activity), chloroplasts and mitochondria occupy the inner half of the BS cells, as is typical in eudicot C_4 species (Edwards and Voznesenskaya, 2011). In *A. incarnata*, there is an abundance of mitochondria scattered among the elongated chloroplasts, while in *B. coccinea,* BS mitochondria are much less frequent and not commonly interspersed within the chloroplast cluster. In NAD-ME species, the juxtaposition of chloroplasts and mitochondria facilitates rapid movement of CO_2 released from mitochondria into adjacent chloroplasts, while the tight packing of chloroplasts and mitochondria reduces the chance of CO_2 escape. This tight packing is a major means by which CO_2 is trapped in species lacking a suberized layer around the BS wall, as is the case in eudicots (von Caemmerer and Furbank, 2003). In *Portulaca* species, the same patterns hold but with one important variation that is not reported in the literature, notably, mitochondria in the BS of NAD-ME *P. oleracea* form distinct attachment structures to adjacent chloroplasts that are not obvious in *Allionia* and other NAD-ME type C_4 species (for example, *Anticharis*, *Cleome, Salsola*, and *Suaeda;* Voznesenskaya et al., 1999, 2018; Khoshravesh et al., 2012). Such structures may facilitate rapid CO_2 diffusion between mitochondria and chloroplasts, enhancing refixation efficiency.

The structural characteristics observed here are consistent with patterns in other C_4 clades. NADP-ME species among C_4 grasses, sedges and eudicots (as shown in *Flaveria*, *Euphorbia*, *Gomphrena, Heliotropium*, *Salsola*, and *Tribulus*) share with *Boerhavia* and *Portulaca pilosa* the pattern of enlarged chloroplasts with weakly developed grana stacks, and few BS mitochondria (Carolin et al., 1978; Kim and Fisher, 1990; Ueno, 1996, 2013; Voznesenskaya et al., 1999; Yoshimura et al., 2004; Muhaidat et al., 2011; Sage T. L. et al., 2011; Lauterbach et al., 2019). NAD-ME eudicots in *Amaranthus*, *Atriplex*, *Anticharis*, *Cleome*, *Gisekia*, *Salsola, Suaeda*, and *Tecticornia* share with *Allionia* and *P. oleracea* the pattern of enlarged chloroplasts with well-developed grana and large numbers of interspersed mitochondria (Carolin et al., 1978; Voznesenskaya et al., 1999, 2007, 2008; Khoshravesh et al., 2012; Bissinger et al., 2014; Oakley et al., 2014). In grasses, Dengler and Nelson (1999) describe NAD-ME species as having 5-to 20-fold more mitochondria than NADP-ME species, and NAD-ME mitochondria are often larger with greater internal membrane surface area in the BS. The differences between C_4 subtypes hold even when different tissue layers are co-opted as the site of CO_2 concentration, whether it is the mestome sheath as in grasses, an inner chlorenchyma layer as occurs in *Salsola* and other succulent chenopods, or single-cell type of C_4 photosynthesis as shown in the NAD-ME chenopods *Bienertia cycloptera* and *Suaeda aralocaspica* (Ueno, 1996, 2013; Voznesenskaya et al., 1999; Pyankov et al., 2000; Edwards and Voznesenskaya, 2011; Khoshravesh et al., 2020). In the middle of M cells of *Bienertia cycloptera*, for example, there are many mitochondria within a ball of Rubisco-containing chloroplasts where CO_2 is concentrated (Voznesenskaya et al., 2002). The central chloroplasts in *B. cycloptera* have well-stacked grana, while peripheral chloroplast that are functionally equivalent to M chloroplasts of typical C_4 species do not.

Lateral transfer of genes encoding C_4 pathway elements has been observed in closely related grass clades within *Alloteropsis* and *Neurachne* (Christin et al., 2012a,b; Dunning et al., 2017; Khoshravesh et al., 2020). Lateral gene transfer is relevant to discussions of evolutionary convergence because the acquisition of previously evolved C_4 genes by a non-C_4 relative could facilitate C_4 evolution in the receptive clade, possibly creating a false impression of convergence. When we examined the phylogenies of C_4 pathway genes in the Nyctagineae, we found no evidence of lateral transfer. The gene trees consistently showed the two C_4 lineages do not share a common gene copy of a C_4-adapted isoforms of *PEPC*, *PPDK,* or either malic enzyme, because the respective orthologs from related C_3 species branch between the C_4 lineages with high support (**Figure 8** and **Supplementary Figures 2–5**). This evidence is consistent with a hypothesis that the two C_4 clades in the Nyctaginaceae arose *de novo* from a completely C_3 state, rather than via assisted origins involving gene introgression from a pre-existing C_4 clade (Christin et al., 2012a,b; Dunning et al., 2017).

With sequencing data becoming widely available, it may be possible to identify C_4-subtypes using transcriptomics (Lauterbach et al., 2019). Such an approach, however, could produce errors if not coupled with enzyme activity assays. Here, the transcriptomes consistently showed high expression of the copies we designate as *NADP-ME1* in *Boerhavia* and *NAD-ME3* in *Allionia*. *Allionia* also exhibited significant expression of *NADP-ME1* transcripts which approach expression of its *NAD-ME3* gene. This data by itself would suggest co-function of NADP-ME and NAD-ME in *Allionia*; however, no NADP-ME activity was detected in *Allionia*, leading us to conclude its *NADP-ME1* transcripts are not translated into functional enzyme. The possibility of error in this conclusion is unlikely because the BS ultrastructure in *Allionia* is consistent with patterns generally seen in NAD-ME subtypes. The high transcript level of both NAD-ME and NADP-ME may be a relic from ancestral *Allionia* species that may have utilized both decarboxylases during an early phase of C_4 photosynthesis. A chance selection event may have subsequently enhanced NAD-ME activity, after which leaf structure and energetics were optimized for the NAD-ME subtype.

One of the notable features of C_4 photosynthesis is that it is concentrated in relatively few orders of higher plants. The Poideae (grasses, sedges) and Caryophyllales, for example, account for over 90% of all C_4 species and about 50 of the estimated 65 independent origins of C_4 photosynthesis (Sage, 2016). The ability of these clades to repeatedly evolve the C_4 pathway is hypothesized to reflect the presence of enabling traits that facilitate C_4 evolution (Sage, 2004; Christin et al., 2013b). Wider BS cells in C_3 species, for example, could be a structural enabling trait (Muhaidat et al., 2011; Christin et al., 2013b; Griffiths et al., 2013), while greater numbers of organelles in BS tissues may reflect an activation of photosynthetic physiology which enables establishment of photorespiratory glycine shuttles (Sage et al., 2013; Schulze et al., 2013). Photosynthetic activation of the BS can facilitate the rise of photorespiratory glycine shuttles because mitochondria and some chloroplasts can reposition to the inner BS region, forcing glycine formed in centrifugal chloroplasts to migrate to the inner BS for processing by glycine decarboxylase (Sage et al., 2013). To evaluate whether enabling traits are present in close C_3 relatives of the C_4 Nyctaginaceae clades, we examined their BS structure and organelle characteristics. In the C_3 species of *Anulocaulus, Cyphomeris,* and *Nyctaginia*, BS cells had similar cross-sectional areas as the C_4 BS cells, and exhibited numerous chloroplasts along the outer wall, indicating photosynthetic activation. We thus conclude the close C_3 relatives of *Allionia* and *Boerhavia* exhibit numerous enabling traits for C_4 evolution, indicating they were present in ancestral C_3 taxa from which the C_4 clades arose.

C_4 Selection Environments

The results indicate C_4 photosynthesis evolved in Nyctaginaceae species from hot and dry climates of the New World, most likely in the arid-to-semi-arid regions of Southwestern North America where the center of diversity occurs for the North American Xerophytic clade of the Nyctaginaceae (Douglas and Manos, 2007). Christin et al. (2011) estimate that the C_4 pathway appeared in the *Boerhavia/Okenia* clade about 4.7 million years ago, and 6.1 million years ago in *Allionia*, based on molecular phylogenies. This was during a period of aridification in Southwestern North American and reduced atmospheric CO_2 (Sage et al., 2018). The environmental habitat of *Allionia* versus *Boerhavia* did not differ much from each other or their C_3 ancestors, indicating the C_3 ancestors were adapted to the same kind of hot, dry environments where the C_4 species are common. This supports a hypothesis that C_4 arose in these sorts of environments, consistent with a model that high levels of photorespiration brought on by heat, drought and/or salinity promoted C_4 evolution in C_3 plants pre-adapted to such harsh environments (Sage et al., 2018).

Surveys of the floristic distribution of NADP-ME versus NAD-ME grasses have identified a trend where NADP-ME species predominate in grass floras of wetter climates while NAD-ME species predominate in drier climates (Vogel et al., 1978; Hattersley, 1992; Ghannoum et al., 2011). The reasons for this trend have not been clarified, although it has been hypothesized that the pattern reflects phylogenetic ancestry where NAD-ME species are more likely to arise in clades from drier areas while NADP-ME species evolve in wetter climate zones (Taub, 2000). Explanations for habitat differences between C_4 subtypes can be evaluated using closely related NADP-ME and NAD-ME clades. In *A. incarnata* and *B. coccinea*, we observed similar photosynthetic capacities and intrinsic WUE (*A/gs*), indicating no obvious differences in photosynthetic parameters that might explain subtype segregation along a moisture gradient. *Allionia* did exhibit a steeper initial slope of the *A/Ci* response than *Boerhavia*, indicating a stronger C_4 metabolic cycle in the NAD-ME plant. Consistently, the NAD-ME clade in *Allionia* had greater PEPCase activity per unit chlorophyll than *Boerhavia* (NADP-ME clade). If a stronger metabolic cycle is an inherent feature of NAD-ME relative to NADP-ME species, there could be a photosynthetic advantage under low intercellular CO_2 concentrations that commonly occur where drought and low humidity reduce stomatal aperture. A comparative analysis using closely related species of differing subtype in the Nyctagineae, *Portulaca* and other lineages could evaluate this possibility.

Convergence Versus Non-convergence in the C_4 Functional Type

As a highly convergent, complex trait, C_4 photosynthesis represents an excellent system to dissect evolutionary convergence, particularly the degree of convergence versus divergence in the mix of traits that give rise to a composite phenotype (Heyduk et al., 2019). Numerous studies have previously examined convergent properties of C_4 components, such as convergence in genes for C_4 enzymes (Christin et al., 2007, 2009, 2010b; Emms et al., 2016); regulatory components (Gowik and Westhoff, 2011; John et al., 2014; Reyna-Llorens and Hibberd, 2017), structural features (Yoshimura et al., 2004; Kadereit et al., 2014; Stata et al., 2014; Danila et al., 2018); and biochemical sub-types (Gutierrez et al., 1974; Hatch et al., 1975; Sage R. F. et al., 2011; Ludwig, 2016b). However, the assembly of multiple datasets into a hierarchical framework has not been attempted in a C_4 context. In **Table 6**, we present a preliminary hierarchy of convergent and divergent traits observed in comparative studies of the many C_4 clades, thereby enabling deeper assessments of when and why convergence is favored, versus when divergent solutions to CO_2 concentration can occur. Convergence within C_4 photosynthesis is apparent in two key steps: the carboxylation of PEP by PEPCase, and the conversion of CO_2 to bicarbonate by CA. The ubiquitous use of PEPCase is probably due to three factors. First, it is readily available for co-option, because it is widely used in C_3 plants for processes such as pH control, metabolite generation for the Krebs cycle and nitrogen assimilation, and shuttling reducing power between cellular compartments (Aubry et al., 2011; Mallmann et al., 2014). Second, there are no obvious alternatives in vascular plants that may be co-opted to provide the carboxylation function of a C_4 cycle. While numerous carboxylases are active in prokaryotes, few occur in higher plants (Erb, 2011). One obvious candidate to recruit into a C_4 cycle is pyruvate carboxylase, which produces OAA from pyruvate and bicarbonate in bacteria, animals and algae, but it is not known to occur in plants (Tsuji et al., 2012). Third, it is worth considering the chain of events giving rise to C_4 photosynthesis. The evolutionary rise of C_2 photosynthesis in C_3–C_4 intermediate

TABLE 6 | A summary list of convergent and divergent traits of C_4 photosynthesis.

Convergent trait	Divergent trait
Biochemical level	**Biochemical level**
a) PEPCase upregulated in M cytosol only. b) Reduced malate sensitivity of PEPCase. c) Carbonic anhydrase expressed in M tissue only. d) Functional rubisco restricted to high CO_2 compartment.	a) Variation in decarboxylating enzymes and associated ultrastructural, transport and regulatory traits. b) PEP-CK species may not require PPDK c) Transport metabolites are malate or aspartate
Structural level	**Structural level**
a) Interior compartment always used for CO_2 concentration. b) Reduced M/BS ratio. c) High vein density. d) More plasmodesmata at M x BS boundary.	a) Cell tissue type for CO_2 concentration (e.g., BS, mestome sheath, inner chloremchyma, central cluster of chloroplasts as in *Bienertia*) b) Variable M and BS arrangement around veins (e.g., partially or incomplete coverage of vein) c) Variable plasmodemata structure (branched, non-branched)
Ultrastructural level	**Ultrastructural level**
a) More chloroplast volume in Kranz sheath tissue b) Less chloroplast investment in M cells c) Reduced mitochondria in M cells	a) Chloroplast size, number, shape and position b) Diffusive trap structure (cell wall thickness, suberin presence, vacuole size, chloroplast cluster arrangement) c) Thylakoid stacking and PSII location d) Variable location of whole-chain versus cyclic electron transport e) Mitochondria number, size and position in BS Note: most of the variable traits above become convergent within a given C_4 subtype.
Genetic level (PEPCase gene only)	**Genetic level (PEPCase gene only)**
a) Use of serine at position 780, with some exceptions b) Use of codons in promoter to target M cell expression and expression strength	a) Variation in amino acids at most sites under positive selection b) Variation in PEPCase promoter structure c) Variation in enzyme paralogs co-opted for the C_4 pathway

BS, bundle sheath; M, mesophyll; PEP-CK, PEPcase, PEP carboxylase; PEP carboxykinase.

species establishes a Kranz-like leaf anatomy and M to BS transport systems that facilitate the subsequent upregulation of PEPCase and a C_4 metabolic cycle (Sage et al., 2012). A leading hypothesis to explain this upregulation is PEPCase provides carbon skeletons to support re-assimilation of photorespiratory ammonia (Mallmann et al., 2014). If so, then the ubiquitous use of PEPCase may arise out of its pre-exiting role supporting N metabolism in C_3 leaves, which make it the ready candidate for optimizing the performance of C_2 photosynthesis, which in turn positions it to be further upregulated as the benefits of the nascent C_4 cycle improve fitness (Heckmann et al., 2013). The convergence upon enhanced CA activity then becomes necessary to enable PEPCase to have enough bicarbonate to maintain rapid activity.

As the most heavily studied enzyme in C_4 photosynthesis, the sequence of the PEPCase gene serves as a model of how convergence versus divergence can operate at the gene sequence level. Convergence is apparent in that *PEPC1* was separately co-opted for C_4 function in the *Allionia* and *Okenia/Boerhavia* lineages, while divergence is apparent in the sequence of *PEPC1* in these clades, as indicated by the distinct branches on the gene trees and variation at specific sites in the amino acid sequences. Is this sequence divergence random, or could it reflect divergent solutions for optimizing PEPcase function in C_4 leaves? In C_4 plants, the affinity of PEPCase for its substrate bicarbonate should be increased to compensate for subsaturating bicarbonate concentrations in M tissues (Gowik and Westhoff, 2011; Di Mario and Cousins, 2019). Also, C_4 orthologs of PEPCase require reduced sensitivity to malate and aspartate, because these allosteric inhibitors of PEPCase must accumulate to high concentration in M cells to drive rapid diffusion into the BS (Jacobs et al., 2008; Gowik and Westhoff, 2011). Numerous studies have documented similar changes in the amino acid sequence of PEPCase in a pattern that is associated with changes in sensitivity to malate, aspartate and PEP, supporting hypotheses of convergent optimization of PEPCase for the C_4 leaf (Engelmann et al., 2003; Gowik et al., 2006; Christin et al., 2007). Convergence is indicated by a widely noted substitution of a serine for alanine at a homologous site in the PEPC sequence, position 780 in maize, which is proposed to alter PEP and possibly bicarbonate sensitivity (Christin et al., 2007; Gowik and Westhoff, 2011; Di Mario and Cousins, 2019). Christin et al. (2007) also observed common sequence substitutions in numerous distinct clades of C_4 grasses, some of which occur in gene regions controlling sensitivity to malate. However, transcriptome surveys of numerous chenopods indicate the sequence convergence observed in grasses is less common in eudicots, suggesting divergent solutions to meeting the kinetic and regulatory requirements of a C_4 PEPCase (Rosnow et al., 2014, 2015). Our results with *Boerhavia* and *Allionia* support a hypothesis of flexibility in how PEPCase is modified for the C_4 leaf. Both clades lack the serine at the position near 780, instead sharing an alanine with their C_3 sisters. They also exhibit differences in half of the sequence positions where Christin et al. (2007) noted convergence in grasses, but do exhibit some of the same substitutions as C_4 grasses and certain C_4 eudicots (at positions 572, 577, 761, and 807; **Table 5**). These patterns raise a number of possibilities. On the one hand, these convergent substitutions may sufficiently alter kinetics to allow the C_4 PEPCase to efficiently function. Alternatively, other sequence shifts may

be able to accomplish the necessary change in sensitivities. A third possibility is the C_4 Nyctaginaceae species may simply have a non-optimized PEPCase for the C_4 function, in which case other mechanisms may compensate for inefficiencies. One compensation mechanism may be environmental, for example, high temperature in the natural habitat may override a need for convergence in PEPCase sequences by stimulating catalytic capacity. Where convergence is beneficial, but not accomplished, the consequences and compensation mechanisms become as interesting as the convergence itself.

We next consider the evolutionary transition at the opposite end of the C_4 pathway, where there is clear divergence in decarboxylase function. The divergence arises because there are enzymatic alternatives to meeting the decarboxyation imperative, with NADP-ME, NAD-ME, and PCK having significant roles in numerous metabolic pathways in C_3 plants (Aubry et al., 2011). Elevated activities of NADP-ME, NAD-ME, and PCK have been detected in vascular tissues and BS cells of C_3 species, where they metabolize organic acids used in long-distance transport from the roots, and may have a role in pH homeostasis, coordinating carbon and nitrogen metabolism, and providing metabolites for numerous biosynthetic pathways (Hibberd and Quick, 2002; Aubry et al., 2011). The mechanism by which one decarboxylase is selected over another is not known, but since all three are active in BS tissues, it is possible that each may function in a nascent C_4 pathway, perhaps to support recovery of photorespired nitrogen in the BS (Mallmann et al., 2014). It may be that chance determines which of the three decarboxylases is upregulated as the C_4 pathway strengthens, with convergence on the distinctive subtype characteristics occurring afterward, as the C_4 pathway is optimized for the selected decarboxylase. Alternatively, pre-existing traits in the C_3 or C_2 ancestors may determine which decarboxylase is selected and also influence the characteristics of each C_4 subtype.

To close, we note that comparative studies of evolutionary convergence and divergence in C_4 plants show that convergence is greatest where constraints are high and biochemical options are limited, while divergence occurs where the constraints are low and multiple alternative solutions can be selected during the evolutionary process. What remains unclear is the influence that chance, ancestry, and environment have over divergent possibilities. By demonstrating multiple sets of closely related clades differing in C_4 subtype, we have identified replicated examples with which to address these issues using comparative methods of evolutionary biology (Harvey and Pagel, 1991). C_4 photosynthesis can thus become a powerful tool to unravel the intricacies of convergence in complex trait evolution.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

RK led the imaging efforts and their quantification and ecological data analysis. MS conducted the phylotranscriptomic and gene sequence analyses. SA performed gas exchange and enzyme assays. TS and RS directed the labs where the work was conducted and supplied funding. RS wrote the manuscript with input from each co-author. All authors contributed to the article and approved the submitted version.

DEDICATION

This paper is dedicated to the memory of Dr. Udo Gowik (1971–2020), a fun-loving friend, helpful to all, a pioneer in C_4 plant biology and a fearsome political debater.

FUNDING

This research was supported by Canadian Natural Science and Engineering Research Council grants RGPIN-04878-2015 to TS and RGPIN-2017-06476 to RS. The contribution of SA was supported by a travel grant from the Institute of Global Innovation Research, Tokyo University of Agriculture and Technology.

ACKNOWLEDGMENTS

We thank Nina Esmail and Perlina Lim for assistance in plant growth and the fixation and embedding of leaves for microscopic imaging. We also thank Gerry Edwards and Elena Vosnesenskaya for the gift of *Portulaca pilosa* seeds.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.578739/full#supplementary-material

REFERENCES

Aberer, A. J., Kobert, K., and Stamatakis, A. (2014). ExaBayes: massively Bayesian tree inference for the whole-genome era. *Mol. Biol. Evol.* 31, 2553–2556. doi: 10.1093/molbev/msu236

Alonso-Cantabrana, H., and von Caemmerer, S. (2016). Carbon isotope discrimination as a diagnostic tool for C_4 photosynthesis in C_3–C_4 intermediate species. *J. Exp. Bot.* 67, 3109–3121. doi: 10.1093/jxb/erv555

Arnon, D. I. (1949). Copper enzymes in isolated chloroplasts: polyphenoloxidase in *Beta vulgaris*. *Plant Physiol.* 24, 1–15. doi: 10.1104/pp.24.1.1

Ashton, A. R., Burnell, J. N., Furbank, R. T., Jenkins, C. L. D., and Hatch, M. D. (1990). "Enzymes of C_4 photosynthesis," in *Methods in Plant Biochemistry*, Vol. 7A, ed. P. J. Lea, (London: Academic Press), 39–72. doi: 10.1016/B978-0-12-461013-2.50010-1

Aubry, S., Brown, N. J., and Hibberd, J. M. (2011). The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway. *J. Exp. Bot.* 62, 3049–3059. doi: 10.1093/jxb/err012

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bissinger, K., Khoshravesh, R., Kotrade, J. P., Oakley, J., Sage, T. L., Sage, R. F., et al. (2014). *Gisekia* (Gisekiaceae): phylogenetic relationships, biogeography, and ecophysiology of a poorly known C_4 lineage in the Caryophyllales. *Am. J. Bot.* 101, 499–509. doi: 10.3732/ajb.1300279

Bivand, R., and Lewin-Koh, N. (2013). *maptools: Tools for Reading and Handling Spatial Objects. R Package Version 0.8–27*. Available online at: https://cran.r-project.org/web/packages/maptools/index.html (accessed August 24, 2020).

Blount, Z. D., Lenski, R. E., and Losos, J. B. (2018). Contingency and determinism in evolution: replaying life's tape. *Science* 362:eaam5979. doi: 10.1126/science.aam5979

Bohley, K., Joos, O., Hartmann, H., Sage, R. F., Liede-Schumann, S., and Kadereit, G. (2015). phylogeny of sesuvioideae (Aizoaceae)-biogeography, leaf anatomy and the evolution of C_4photosynthesis. *Perspect. Plant Ecol.* 17, 116–130.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bräutigam, A., and Gowik, U. (2016). Photorespiration connects C_3 and C_4 photosynthesis. *J. Exp. Bot.* 67, 2953–2962.

Bräutigam, A., and Weber, A. P. M. (2011). "Transport processes: connecting the reactions of C_4 photosynthesis," in *C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms*, eds A. S. Raghavendra, and R. F. Sage, (Dordrecht: Springer Netherlands), 277–300.

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Carolin, R., Jacobs, S., and Vesk, M. (1978). Kranz cells and mesophyll in the Chenopodiales. *Aust. J. Bot.* 26:683. doi: 10.1071/BT9780683

Christin, P.-A., Besnard, G., Samaritani, E., Duvall, M. R., Hodkinson, T. R., Savolainen, V., et al. (2008). Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. *Curr. Biol.* 18, 37–43. doi: 10.1016/j.cub.2007.11.058

Christin, P.-A., Boxall, S. F., Gregory, R., Edwards, E. J., Hartwell, J., and Osborne, C. P. (2013a). Parallel recruitment of multiple genes into C_4 photosynthesis. *Genome Biol. Evol.* 5, 2174–2187. doi: 10.1093/gbe/evt168

Christin, P.-A., Edwards, E. J., Besnard, G., Boxall, S. F., Gregory, R., Kellogg, E. A., et al. (2012a). Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. *Curr. Biol.* 22, 445–449. doi: 10.1016/j.cub.2012.01.054

Christin, P.-A., Freckleton, R. P., and Osborne, C. P. (2010a). Can phylogenetics identify C_4 origins and reversals? *Trends Ecol. Evol.* 25, 403–409. doi: 10.1016/j.tree.2010.04.007

Christin, P.-A., and Osborne, C. P. (2013). The recurrent assembly of C_4 photosynthesis, an evolutionary tale. *Photosynth. Res.* 117, 163–175. doi: 10.1007/s11120-013-9852-z

Christin, P.-A., Osborne, C. P., Chatelet, D. S., Columbus, J. T., Besnard, G., Hodkinson, T. R., et al. (2013b). Anatomical enablers and the evolution of C_4 photosynthesis in grasses. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1381–1386. doi: 10.1073/pnas.1216777110

Christin, P. A., Osborne, C. P., Sage, R. F., Arakaki, M., and Edwards, E. J. (2011). C_4 eudicots are not younger than C_4 monocots. *J. Exp. Bot.* 62, 3171–3181. doi: 10.1093/jxb/err041

Christin, P.-A., Salamin, N., Savolainen, V., Duvall, M. R., and Besnard, G. (2007). C_4 photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr. Biol.* 17, 1241–1247. doi: 10.1016/j.cub.2007.06.036

Christin, P.-A., Samaritani, E., Petitpierre, B., Salamin, N., and Besnard, G. (2009). Evolutionary insights on C_4 photosynthetic subtypes in grasses from genomics and phylogenetics. *Genome Biol. Evol.* 1, 221–230. doi: 10.1093/gbe/evp020

Christin, P.-A., Wallace, M. J., Clayton, H., Edwards, E. J., Furbank, R. T., Hattersley, P. W., et al. (2012b). Multiple photosynthetic transitions, polyploidy, and lateral gene transfer in the grass subtribe Neurachninae. *J. Exp. Bot.* 63, 6297–6308. doi: 10.1093/jxb/ers282

Christin, P. A., Weinreich, D. M., and Besnard, G. (2010b). Causes and evolutionary significance of genetic convergence. *Trends Genet.* 26, 400–405. doi: 10.1016/j.tig.2010.06.005

Conway-Morris, S. (2003). *Life's Solutions: Inevitable Humans in a Lonely Universe*. Cambridge: Cambridge University Press.

Danila, F. R., Quick, W. P., White, R. G., Kelly, S., von Caemmerer, S., and Furbank, R. T. (2018). Multiple mechanisms for enhanced plasmodesmata density in disparate subtypes of C_4 grasses. *J. Exp. Bot.* 69, 1135–1145. doi: 10.1093/jxb/erx456

Dengler, N. G., and Nelson, T. (1999). "Leaf structure and development in C_4 plants," in *C_4 Plant Biology*, eds R. F. Sage, and R. K. Monson, (San Diego: Academic Press), 133–172.

Di Mario, R. J., and Cousins, A. B. (2019). A single serine to alanine substitution decreases bicarbonate affinity of phosphoenol pyruvate carboxylase in C_4 *Flaveria trinervia*. *J. Exp. Bot.* 70, 995–1004. doi: 10.1093/jxb/ery403

Douglas, N. A., and Manos, P. S. (2007). Molecular phylogeny of Nyctaginaceae: taxonomy, biogeography, and characters associated with a radiation of xerophytic genera in North America. *Am. J. Bot.* 94, 856–872. doi: 10.3732/ajb.94.5.856

Douglas, N. A., and Spellenberg, R. (2010). A new tribal classification of Nyctaginaceae. *Taxon* 59, 905–910. doi: 10.1002/TAX.593018

Drincovich, M. F., Lara, M. V., Andreo, C. S., and Maurino, V. G. (2011). "C_4 decarboxylases: different solutions for the same biochemical problem, the provision of CO_2 to Rubisco in the bundle sheath cells," in *C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms*, eds A. S. Raghavendra, and R. F. Sage, (Dordrecht: Springer Netherlands), 277–300.

Dunning, L. T., Lundgren, M. R., Moreno-Villena, J. J., Namaganda, M., Edwards, E. J., Nosil, P., et al. (2017). Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives: complex transitions among photosynthetic types. *Evolution* 71, 1541–1555. doi: 10.1111/evo.13250

Edwards, E. J. (2019). Evolutionary trajectories, accessibility and other metaphors: the case of C_4 and CAM photosynthesis. *New Phytol.* 223, 1742–1755. doi: 10.1111/nph.15851

Edwards, G. E., and Voznesenskaya, E. V. (2011). "C_4 photosynthesis: kranz forms and single-cell C_4 in terrestrial plants," in *C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms*, eds A. S. Raghavendra, and R. F. Sage, (Dordrecht: Springer Netherlands), 29–61.

Edwards, G. E., and Walker, D. A. (1983). *C_3, C_4: Mechanism, and Cellular and Environmental Regulation, of Photosynthesis*. Oxford: Blackwell Scientific Publications.

Emms, D. M., Covshoff, S., Hibberd, J. M., and Kelly, S. (2016). Independent and parallel evolution of new genes by gene duplication in two origins of C_4 photosynthesis provides new insight into the mechanism of phloem loading in C_4 species. *Mol. Biol. Evol.* 33, 1796–1806. doi: 10.1093/molbev/msw057

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157. doi: 10.1186/s13059-015-0721-2

Engelmann, S., Bläsing, O. E., Gowik, U., Svensson, P., and Westhoff, P. (2003). Molecular evolution of C_4 phosphoenolpyruvate carboxylase in the genus *Flaveria*? A gradual increase from C_3 to C_4 characteristics. *Planta* 217, 717–725. doi: 10.1007/s00425-003-1045-0

Erb, T. J. (2011). Carboxylases in natural and synthetic microbial pathways. *Appl. Environ. Microbiol.* 77, 8466. doi: 10.1128/AEM.05702-11

Farquhar, G. D., Ehleringer, J. R., and Hubick, K. T. (1989). Carbon isotope discrimination and photosynthesis. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* 40, 503–537. doi: 10.1146/annurev.pp.40.060189.002443

Fick, S. E., and Hijmans, R. J. (2017). Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. doi: 10.1002/joc.5086

Furbank, R. T. (2011). Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? *J. Exp. Bot.* 62, 3103–3108. doi: 10.1093/jxb/err080

Ghannoum, O., Evans, J. R., and Caemmerer, S. (2011). "Nitrogen and water use efficiency of C_4 plants"," in *C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms*, eds A. S. Raghavendra, and R. F. Sage, (Dordrecht: The Netherlands: Springer), 129–146.

Gowik, U., Bräutigam, A., Weber, K. L., Weber, A. P. M., and Westhoff, P. (2011). Evolution of C_4 photosynthesis in the genus *Flaveria*: how many and which genes does it take to make C_4? *Plant Cell* 23, 2087–2105. doi: 10.1105/tpc.111.086264

Gowik, U., Engelmann, S., Bläsing, O. E., Raghavendra, A. S., and Westoff, P. (2006). Evolution of C_4 phosphoenolpyruvate carboxylase in the genus

Alternanthera: gene families and the enzymatic characterisitics of the C_4 isozyme and its orthologues in C_3 and C_4 *Alternanthera*. *Planta* 223, 359–368. doi: 10.1007/s00425-005-0085-z

Gowik, U., and Westhoff, P. (2011). "C_4-phosphoenolpyruvate carboxylase," in *C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms*, eds A. S. Raghavendra, and R. F. Sage, (Dordrecht: Springer Netherlands), 257–275.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Griffiths, H., Weller, G., Toy, L. F. M., and Dennis, R. J. (2013). You're so vein: bundle sheath physiology, phylogeny and evolution in C_3 and C_4 plants: origins and function of bundle sheath in C_3 plants. *Plant Cell Environ.* 36, 249–261. doi: 10.1111/j.1365-3040.2012.02585

Gutierrez, M., Gracen, V. E., and Edwards, G. E. (1974). Biochemical and cytological relationships in C_4 plants. *Planta* 119, 279–300. doi: 10.1007/BF00388331

Harvey, P. H., and Pagel, M. D. (1991). *The Comparative Method in Evolutionary Biology*. Oxford: Oxford Press.

Hatch, M., Kagawa, T., and Craig, S. (1975). Subdivision of C_4-pathway species based on differing C_4 acid decarboxylating systems and ultrastructural features. *Funct. Plant Biol.* 2:111. doi: 10.1071/PP9750111

Hatch, M. D. (1987). C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochim. Biophys. Acta Rev. Bioenerget.* 895, 81–106. doi: 10.1016/S0304-4173(87)80009-5

Hattersley, P. W. (1992). "C_4 photosynthetic pathway variation in grasses (Poaceae): its significance for arid and semi-arid lands," in *Desertified Grasslands: their Biology and Management*, ed. G. P. Chapman, (London: Academic Press), 181–212.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P. M., et al. (2013). Predicting C_4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell* 153, 1579–1588. doi: 10.1016/j.cell.2013.04.058

Heyduk, K., Moreno-Villena, J. J., Gilman, I. S., Christin, P.-A., and Edwards, E. J. (2019). The genetics of convergent evolution: insights from plant photosynthesis. *Nat. Rev. Genet.* 20, 485–493. doi: 10.1038/s41576-019-0107-5

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

Hijmans, R. J., and van Etten, J. (2012). *raster: Geographic Analysis and Modeling with Raster Data. R package Version 2.0-12*. Available online at: http://CRAN.R-project.org/package=raster (accessed July 17, 2020).

Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754

Jacobs, B., Engelmann, S., Westhoff, P., and Gowik, U. (2008). Evolution of C_4 phosphoenolpyruvate carboxylase in *Flaveria*: determinants for high tolerance towards the inhibitor L-malate. *Plant Cell Environ.* 31, 793–803. doi: 10.1111/j.1365-3040.2008.01796

John, C. R., Smith-Unna, R. D., Woodfield, H., Covshoff, S., and Hibberd, J. M. (2014). Evolutionary convergence of cell-specific gene expression in independent lineages of C_4 grasses. *Plant Physiol.* 165, 62–75. doi: 10.1104/pp.114.238667

Kadereit, G., Ackerly, D., and Pirie, M. D. (2012). A broader model for C_4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae s.s.). *Proc. R. Soc. B.* 279, 3304–3311.

Kadereit, G., Lauterbach, M., Pirie, M. D., Arafeh, R., and Freitag, H. (2014). When do different C_4 leaf anatomies indicate independent C_4 origins? Parallel evolution of C_4 leaf types in Camphorosmeae (Chenopodiaceae). *J. Exp. Bot.* 65, 3499–3511. doi: 10.1093/jxb/eru169

Kanai, R., and Edwards, G. (1999). "The biochemistry of C_4 photosynthesis," in *C_4 Plant Biology*, eds R. F. Sage, and R. K. Monson, (San Diego, CA: Academic Press), 49–87.

Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Khoshravesh, R., Akhani, H., Sage, T. L., Nordenstam, B., and Sage, R. F. (2012). Phylogeny and photosynthetic pathway distribution in *Anticharis* Endl. (Scrophulariaceae). *J. Exp. Bot.* 63, 5645–5658. doi: 10.1093/jxb/ers218

Khoshravesh, R., Lundsgaard-Nielsen, V., Sultmanis, S., and Sage, T. L. (2017). "Light microscopy, transmission electron microscopy, and immunohistochemistry protocols for studying photorespiration," in *Photorespiration: Methods and Protocols*, eds A. R. Fernie, H. Bauwe, and A. P. M. Weber, (New York, NY: Springer), 243–270.

Khoshravesh, R., Stata, M., Busch, F. A., Saladié, M., Castelli, J. M., Dakin, N., et al. (2020). The evolutionary origin of C_4 photosynthesis in the grass subtribe Neurachninae. *Plant Physiol.* 182, 566–583. doi: 10.1104/pp.19.00925

Khoshravesh, R., Stinson, C. R., Stata, M., Busch, F. A., Sage, R. F., Ludwig, M., et al. (2016). C_3–C_4 intermediacy in grasses: organelle enrichment and distribution, glycine decarboxylase expression, and the rise of C_2 photosynthesis. *J. Exp. Bot.* 67, 3065–3078. doi: 10.1093/jxb/erw150

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Kim, I., and Fisher, D. G. (1990). Structural aspects of the leaves of seven species of *Portulaca* growing in Hawaii. *Can. J. Bot.* 68, 1803–1811. doi: 10.1139/b90-233

Lauterbach, M., Zimmer, R., Alexa, A. C., Adachi, S., Sage, R., Sage, T., et al. (2019). Variation in leaf anatomical traits relates to the evolution of C_4 photosynthesis in Tribuloideae (Zygophyllaceae). *Perspect. Plant Ecol.* 39:125463. doi: 10.1016/j.ppees.2019.125463

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: a package for multivariate analysis. *J. Stat. Softw.* 25, 1–18. doi: 10.18637/jss.v025.i01

Leegood, R. C., and Walker, R. O. (1999). "Regulation of the C_4 pathway," in *C_4 Plant Biology*, eds R. F. Sage, and R. K. Monson, (San Diego, CA: Academic Press), 89–131.

Losos, J. B. (2011). Convergence, adaptation, and constraint. *Evolution* 65, 1827–1840. doi: 10.1111/j.1558-5646.2011.01289

Ludwig, M. (2016a). Evolution of carbonic anhydrases in C_4 plants. *Curr. Opin. Plant Sci.* 31, 16–22. doi: 10.1016/j.pbi.2016.03.003

Ludwig, M. (2016b). The roles of organic acids in C_4 photosynthesis. *Front. Plant Sci.* 7:647. doi: 10.3389/fpls.2016.00647

Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M. J., Weber, A. P., Westhoff, P., et al. (2014). The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. *eLife* 3:e02478. doi: 10.7554/eLife.02478

Monson, R. K., Teeri, J. A., Ku, M. S. B., Gurevitch, J., Mets, L. J., and Dudley, S. (1988). Carbon-isotope discrimination by leaves of *Flaveria* species exhibiting different amounts of C_3–and C_4-cycle co-function. *Planta* 174, 145–151. doi: 10.1007/BF00394765

Moore, B. D., Ku, M. S. B., and Edwards, G. E. (1989). Expression of C_4-like photosynthesis in several species of *Flaveria*. *Plant Cell Environ.* 12, 541–549. doi: 10.1111/j.1365-3040.1989.tb02127

Muhaidat, R., Sage, R. F., and Dengler, N. G. (2007). Diversity of Kranz anatomy and biochemistry in C_4 eudicots. *Am. J. Bot.* 94, 362–381. doi: 10.3732/ajb.94.3.362

Muhaidat, R., Sage, T. L., Frohlich, M. W., Dengler, N. G., and Sage, R. F. (2011). Characterization of C_3–C_4 intermediate species in the genus *Heliotropium* L. (Boraginaceae): anatomy, ultrastructure and enzyme activity. *Plant Cell Environ.* 34, 1723–1736. doi: 10.1111/j.1365-3040.2011.02367

Oakley, J. C., Sultmanis, S., Stinson, C. R., Sage, T. L., and Sage, R. F. (2014). Comparative studies of C_3 and C_4 *Atriplex* hybrids in the genomics era: physiological assessments. *J. Exp. Bot.* 65, 3637–3647. doi: 10.1093/jxb/eru106

Ocampo, G., Koteyeva, N. K., Voznesenskaya, E. V., Edwards, G. E., Sage, T. L., Sage, R. F., et al. (2013). Evolution of leaf anatomy and photosynthetic pathways in Portulacaceae. *Am. J. Bot.* 100, 2388–2402. doi: 10.3732/ajb.1300094

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Pyankov, V. I., Voznesenskaya, E. V., Kuz'min, A. N., Ku, M. S. B., Ganko, E., Franceschi, V. R., et al. (2000). Occurrence of C_3 and C_4 photosynthesis in cotyledons and leaves of *Salsola* species (Chenopodiaceae). *Photosynth. Res.* 63, 69–84. doi: 10.1023/A:1006377708156

Rao, X., and Dixon, R. A. (2016). The differences between NAD-ME and NADP-ME subtypes of C_4 photosynthesis: more than decarboxylating enzymes. *Front. Plant Sci.* 7:1525. doi: 10.3389/fpls.2016.01525

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things)." *Methods Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169

Reyna-Llorens, I., and Hibberd, J. M. (2017). Recruitment of pre-existing networks during the evolution of C_4 photosynthesis. *Phil. Trans. R. Soc. B* 372:20160386. doi: 10.1098/rstb.2016.0386

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277.

Rosnow, J. J., Edwards, G. E., and Roalson, E. H. (2014). Positive selection of Kranz and non-Kranz C_4 phosphoenolpyruvate carboxylase amino acids in Suaedoideae (Chenopodiaceae). *J. Exp. Bot.* 65, 3595–3607. doi: 10.1093/jxb/eru053

Rosnow, J. J., Evans, M. A., Kapralov, M. V., Cousins, A. B., Edwards, G. E., and Roalson, E. H. (2015). Kranz and single-cell forms of C_4 plants in the subfamily Suaedoideae show kinetic C_4 convergence for PEPC and Rubisco with divergent amino acid substitutions. *J. Exp. Bot.* 66, 7347–7358. doi: 10.1093/jxb/erv431

Sage, R. F. (2004). The evolution of C_4 photosynthesis. *New Phytol.* 161, 341–370. doi: 10.1111/j.1469-8137.2004.00974

Sage, R. F. (2016). A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *J. Exp. Bot.* 67, 4039–4056. doi: 10.1093/jxb/erw156

Sage, R. F., Christin, P.-A., and Edwards, E. J. (2011). The C_4 plant lineages of planet Earth. *J. Exp. Bot.* 62, 3155–3169. doi: 10.1093/jxb/err048

Sage, R. F., Khoshravesh, R., and Sage, T. L. (2014). From proto-Kranz to C_4 Kranz: building the bridge to C_4 photosynthesis. *J. Exp. Bot.* 65, 3341–3356. doi: 10.1093/jxb/eru180

Sage, R. F., Monson, R. K., Ehleringer, J. R., Adachi, S., and Pearcy, R. W. (2018). Some like it hot: the physiological ecology of C_4 plant evolution. *Oecologia* 187, 941–966. doi: 10.1007/s00442-018-4191-6

Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and the evolution of C_4 photosynthesis. *Annu. Rev. Plant Biol.* 63, 19–47. doi: 10.1146/annurev-arplant-042811-105511

Sage, T. L., Busch, F. A., Johnson, D. C., Friesen, P. C., Stinson, C. R., Stata, M., et al. (2013). Initial events during the evolution of C_4 photosynthesis in C_3 species of *Flaveria*. *Plant Physiol.* 163, 1266–1276. doi: 10.1104/pp.113.221119

Sage, T. L., Sage, R. F., Vogan, P. J., Rahman, B., Johnson, D. C., Oakley, J. C., et al. (2011). The occurrence of C_2 photosynthesis in *Euphorbia* subgenus *Chamaesyce* (Euphorbiaceae). *J. Exp. Bot.* 62, 3183–3195. doi: 10.1093/jxb/err059

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089

Schulze, S., Mallmann, J., Burscheidt, J., Koczor, M., Streubel, M., Bauwe, H., et al. (2013). Evolution of C_4 photosynthesis in the genus *Flaveria*: establishment of a photorespiratory CO_2 pump. *Plant Cell* 25, 2522–2535. doi: 10.1105/tpc.113.114520

Spellenberg, R. W. (2003). Nyctaginaceae. *Flora North Am.* 4, 14–74.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.

Stata, M., Sage, T. L., Rennie, T. D., Khoshravesh, R., Sultmanis, S., Khaikin, Y., et al. (2014). Mesophyll cells of C_4 plants have fewer chloroplasts than those of closely related C_3 plants: C_3 versus C_4 mesophyll chloroplasts. *Plant Cell Environ.* 37, 2587–2600. doi: 10.1111/pce.12331

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–W612. doi: 10.1093/nar/gkl315

Taub, D. R. (2000). Climate and the U.S. distribution of C_4 grass subfamilies and decarboxylation variants of C_4 photosynthesis. *Am. J. Bot.* 87, 1211–1215. doi: 10.2307/2656659

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., et al. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* 37, 914–939. doi: 10.1111/j.1365-313x.2004.02016.x

Tropicos (2020). *Tropicos.org. Missouri Botanical Garden* http://www.tropicos.org (accessed October 20, 2020).

Tsuji, Y., Suzuki, I., and Shiraiwa, Y. (2012). Enzymological evidence for the function of a plastid-located pyruvate carboxylase in the haptophyte alga *Emiliania huxleyi*: a novel pathway for the production of C_4 compounds. *Plant Cell Physiol.* 53, 1043–1052. doi: 10.1093/pcp/pcs045

Ueno, O. (1996). Structural characterization of photosynthetic cells in an amphibious sedge, *Eleocharis vivipara*, in relation to C_3 and C_4 metabolism. *Planta* 199, 382–393. doi: 10.1007/BF00195730

Ueno, O. (2013). Ultrastructure and carbon isotope ratios of leaves in C_4 species of *Rhynchospora* (Cyperaceae) that differ in the location of Kranz cells. *Int. J. Plant Sci.* 174, 702–709. doi: 10.1086/669912

Vogel, J. C., Fuls, A., and Ellis, R. P. (1978). The geographical distribution of Kranz grasses in South Africa. *S. Afr. J. Sci.* 74, 209–215.

Von Caemmerer, S. (1992). Carbon isotope discrimination in C_3−C_4 intermediates. *Plant Cell Environ.* 15, 1063–1072. doi: 10.1111/j.1365-3040.1992.tb01656

von Caemmerer, S., and Furbank, R. T. (2003). The C_4 pathway: an efficient CO_2 pump. *Photosynth. Res.* 77, 191–207. doi: 10.1023/A:1025830019591

Voznesenskaya, E. V., Akhani, H., Koteyeva, N. K., Chuong, S. D. X., Roalson, E. H., Kiirats, O., et al. (2008). Structural, biochemical, and physiological characterization of photosynthesis in two C_4 subspecies of *Tecticornia indica* and the C_3 species *Tecticornia pergranulata* (Chenopodiaceae). *J. Exp. Bot.* 59, 1715–1734. doi: 10.1093/jxb/ern028

Voznesenskaya, E. V., Chuong, S. D. X., Koteyeva, N. K., Franceschi, V. R., Freitag, H., and Edwards, G. E. (2007). Structural, biochemical, and physiological characterization of C_4 photosynthesis in species having two vastly different types of Kranz anatomy in genus *Suaeda* (Chenopodiaceae). *Plant Biol.* 9, 745–757. doi: 10.1055/s-2007-965579

Voznesenskaya, E. V., Franceschi, V. R., Chuong, S. D. X., and Edwards, G. E. (2006). Functional characterization of phosphoenolpyruvate carboxykinase-type C_4 leaf anatomy: immuno-, cytochemical and ultrastructural analyses. *Ann. Bot.* 98, 77–91. doi: 10.1093/aob/mcl096

Voznesenskaya, E. V., Franceschi, V. R., Kiirats, O., Artyusheva, E. G., Freitag, H., and Edwards, G. E. (2002). Proof of C_4 photosynthesis without Kranz anatomy in *Bienertia cycloptera* (Chenopodiaceae). *Plant J.* 31, 649–662. doi: 10.1046/j.1365-313X.2002.01385

Voznesenskaya, E. V., Franceschi, V. R., Pyankov, V. I., and Edwards, G. E. (1999). Anatomy, chloroplast structure and compartmentation of enzymes relative to photosynthetic mechanisms in leaves and cotyledons of species in the tribe Salsoleae (Chenopodiaceae). *J. Exp. Bot.* 50, 1779–1795. doi: 10.1093/jxb/50.341.1779

Voznesenskaya, E. V., Koteyeva, N. K., Cousins, A., and Edwards, G. E. (2018). Diversity in structure and forms of carbon assimilation in photosynthetic organs in *Cleome* (Cleomeaceae). *Funct. Plant Biol.* 45:983. doi: 10.1071/FP17323

Voznesenskaya, E. V., Koteyeva, N. K., Edwards, G. E., and Ocampo, G. (2010). Revealing diversity in structural and biochemical forms of C_4 photosynthesis and a C_3–C_4 intermediate in genus *Portulaca* L. (Portulacaceae). *J. Exp. Bot.* 61, 3647–3662. doi: 10.1093/jxb/erq178

Yoshimura, Y., Kubota, F., and Ueno, O. (2004). Structural and biochemical bases of photorespiration in C_4 plants: quantification of organelles and glycine decarboxylase. *Planta* 220, 307–317. doi: 10.1007/s00425-004-1335-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 05 November 2020
doi: 10.3389/fpls.2020.546518

A Review of C_4 Plants in Southwest Asia: An Ecological, Geographical and Taxonomical Analysis of a Region With High Diversity of C_4 Eudicots

Alexander Rudov[1], *Marjan Mashkour*[2], *Morteza Djamali*[3] *and Hossein Akhani*[1]*

[1] Halophytes and C_4 Plants Research Laboratory, Department of Plant Sciences, School of Biology, College of Sciences, University of Tehran, Tehran, Iran, [2] Archéozoologie, Archéobotanique: Sociétés, Pratiques et Environnements (AASPE/UMR7209)—CNRS (Centre national de Recherche Scientifique) et MNHN (Muséum national d'Histoire naturelle), Paris, France, [3] Institut Méditerranéen de Biodiversité et d'Ecologie (IMBE/UMR7263), Aix Marseille Univ, Avignon Univ, CNRS, IRD, IMBE, Aix-en-Provence, France

OPEN ACCESS

Edited by:
Sarah Covshoff,
Independent Researcher, Las Vegas, NV, United States

Reviewed by:
Colin Osborne,
The University of Sheffield, United Kingdom
Isabel Larridon,
Royal Botanic Gardens, Kew, United Kingdom

***Correspondence:**
Hossein Akhani
hakhani@ut.ac.ir

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: 17 April 2020
Accepted: 19 August 2020
Published: 05 November 2020

Citation:
Rudov A, Mashkour M, Djamali M and Akhani H (2020) A Review of C_4 Plants in Southwest Asia: An Ecological, Geographical and Taxonomical Analysis of a Region With High Diversity of C_4 Eudicots. Front. Plant Sci. 11:546518. doi: 10.3389/fpls.2020.546518

Southwest Asia is climatically and topographically a highly diverse region in the xeric belt of the Old World. Its diversity of arid habitats and climatic conditions acted as an important area for the evolution and diversification of up to 20 (of 38 known) independent Eudicot C_4 origins. Some of these lineages present unique evolutionary strategies like single-cell functioning C_4 and C_3–C_4 switching mechanisms. The high diversity of C_4 taxa in Southwest (SW) Asia is also related to the presence of seven phytogeographic zones including the Irano-Turanian region as a center of diversification of many Caryophyllales lineages and the Somali-Masai region (Southern Oman and Yemen) as a center of diversification for C_4 Monocots. Nevertheless, the C_4 flora of SW Asia has not received detailed attention. This paper presents a comprehensive review of all known C_4 species in the area based on a literature survey, own floristic observations, as well as taxonomic, phylogenetic and herbarium data, and δ^{13}C-isotope ratio analysis. The resulting checklist includes a total number of 923 (861 native, of which 141 endemic, and 62 introduced) C_4 species, composed of 350 Eudicots and 509 Monocots, most of which are therophytic and hemicryptophytic xerophytes with pluriregional and Irano-Turanian distribution. Two hundred thirty-nine new δ^{13}C-isotope ratios of C_4 and C_3 plants, as well as some taxonomic changes are presented. An analysis of the distribution of the three main C_4 plant families (Chenopodiaceae, Poaceae, and Cyperaceae) in the region in relation to climatic variables indicates that the increase of C_4 species follows more or less a latitudinal gradient similar to global patterns, while separate taxonomic groups seem to depend on specific factors as continentality (Chenopodiaceae), average annual temperature (Cyperaceae), and the presence of summer precipitation (Poaceae). An increase of C_4 Eudicots in W-E direction even in similar longitudinal belts is explained by a combination of edaphic and climatic conditions. The provided data should encourage a deeper interest in the evolution of C_4 lineages in SW Asia and their adaptation to ecological and climatical

conditions and awaken interest in the importance of local C_4 crops, the conservation of threatened C_4 taxa, and awareness of human impacts on the rapid environmental changes in the region.

Keywords: C_3–C_4 switching plants, C_4 crops, Chenopodiaceae, conservation, Irano-Turanian region, Poaceae, single-cell C_4, $\delta^{13}C$-isotope ratio

INTRODUCTION

Since its discovery, during the seventh decade of the twentieth century, the C_4 photosynthetic pathway has received attention of extensive studies (Hatch, 1999). In contrast to C_3 photosynthesis, which evolved under high atmospheric CO_2 levels and mesic conditions, C_4 photosynthesis developed under low CO_2 levels and arid conditions. The climatic changes during the Oligocene (30–25 M.y.a.) and the following Miocene, marked by dropping CO_2 levels and increasing seasonality with hot and dry periods and the resulting expansion of arid habitats, favored the convergent evolution and diversification of various C_4 lineages (Sage, 2001; Osborne and Beerling, 2006; Christin et al., 2008; Sage, 2016). C_4 photosynthesis involves a CO_2 concentrating mechanism in hot and arid conditions through the activity of the phosphoenolpyruvate carboxylase (PEPC), an enzyme with a high affinity for HCO_3^-. The mechanism avoids photorespiration by concentrating CO_2 levels around Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) using a special dual compartmentation named Kranz-anatomy (Sage et al., 2012; Bräutigam and Gowik, 2016). This structure allows fixation of CO_2 by PEPC in the mesophyll and its decarboxylation and concentration around Rubisco in the bundle sheath cells. Subsequently, it has been shown, however, that Kranz-anatomy is not always required for C_4 photosynthesis in terrestrial plants (Voznesenskaya et al., 2001).

For the classification as a C_4 plant, the $\delta^{13}C$ ratio is a decisive indicator in all known fully functional C_4 plants. It is related to differences in the fractionation of stable carbon isotopes ^{12}C and ^{13}C between C_3 and C_4 plants (O'Leary, 1988; Von Caemmerer et al., 2014). Based on the type of decarboxylating enzymes, C_4 plants have been distinguished according to their metabolic type in NAD-dependent malic enzyme (NAD-ME)-type, NADP-dependent malic enzyme (NADP-ME)–type, and phosphoenolpyruvate carboxykinase (PEP-CK)-type (Pyankov et al., 2010). Recent studies, however, question this classification, indicating that C_4 plants can be classified in relation to the malate-decarboxylating enzymes as NAD-ME or NADP-ME, while PEP-CK may be considered as an additional decarboxylating pathway (Bräutigam et al., 2014; Wang et al., 2014; Rao and Dixon, 2016).

C_4 photosynthesis has been a metabolic revolution within the plant kingdom and a highly favorable metabolic pathway in plants growing under hot and arid conditions. In fact, while C_4 plants comprise only 3% of vascular plants, they account for 25% of terrestrial photosynthesis (Sage, 2004; Osborne, 2010). Furthermore, this photosynthetic pathway evolved at least 64 times convergently in different plant families (Sage, 2016). Three most numerous C_4 taxonomic groups can be distinguished: C_4 Poaceae (with around 19 independent C_4 clades including ca. 321 genera and over 5,000 species), C_4 Cyperaceae (including 6 independent C_4 clades in around 7 genera and over 1,300 species) and C_4 Caryophyllales [including 24 independent C_4 clades within 8 families (namely Amaranthaceae s. str., Aizoaceae, Caryophyllaceae, Chenopodiaceae, Gisekiaceae, Molluginaceae, Polygonaceae, and Portulacaceae] ca. 50 genera and over 1000 species) (Sage, 2016). The most interesting case in Caryophyllales is the particular diversity of independent C_4 clades within the Chenopodiaceae family, that may include up to 13 independent C_4 clades and 15 different C_4 leaf anatomical types (Pyankov et al., 2001; Kadereit et al., 2003; Kadereit et al., 2012; Sage, 2016). [Following Hernandez-Ledesma et al., 2015; Walker et al., 2018, Chenopodiaceae is treated in this article as a separate family and not as a part of Amaranthaceae, although APG III and IV suggest it to be included in Amaranthaceae (APG III (The Angiosperm Phylogeny Group), 2009; APG IV (The Angiosperm Phylogeny Group), 2016)].

Previous studies have shown that different taxonomic groups of C_4 plants [e.g., Monocots (Poaceae, Cyperaceae) and Eudicots (Chenopodiaceae)] follow different distributions in relation to climatic variables (Stowe and Teeri, 1978; Pyankov et al., 2010). In the case of the European continent, the distribution of C_4 Monocots seems to be related to high temperatures, while the distribution of C_4 Chenopodiaceae and several other C_4 Eudicot lineages shows a relation to aridity (Pyankov et al., 2010). Similar tendencies can be observed also in other regions. E.g., the adaptation to aridity extends the dominion of C_4 Chenopods to the highly continental Gobi desert and over 4,000 m altitude of the Pamir (Pyankov et al., 2000a). In tropical and subtropical Asia, Africa, Australia, and South America the majority of C_4 Poaceae confirm the trend observed in Europe, being mainly distributed in hot climates with the obligatory presence of summer rainfall (Hattersley, 1983; Cabido et al., 2008; Ehleringer et al., 1987; Schulze et al., 1996; Wooler et al., 2001). C_4 Cyperceae, finally, have been reported to be abundant in warm tropical temporary wetlands (Stock et al., 2004). The biodiversity of C_4 plants has been so far reported for Europe (Pyankov et al., 2010) and China (Wang and Ma, 2016), while further studies on the distribution of C_4 plants in relation to climatic and ecological parameters are available for various C_4 lineages or specific geographical areas of the Old and New World (Raghavendra and Das, 1976; Teeri and Stowe, 1976; Tieszen et al., 1979; Waller and Lewis, 1979; Winter, 1981; Hattersley, 1983; Ehleringer et al., 1987; Batanouny et al., 1988; Medina et al., 1989; Schulze et al., 1996; Akhani et al., 1997; Rundel et al., 1999; Pyankov et al., 2000a; Wooler et al., 2001; Stock et al., 2004; Cabido et al., 2008; Mantlana et al., 2008).

The C_4 flora of Southwest Asia is of great interest because of the remarkable diversity of C_4 Eudicots and discovery of single-cell functioning C_4 (Rechinger, 1963-2015; Nikitin and Geldikhanov, 1988; Akhani et al., 1997; Edwards et al., 2004; Akhani and Ghasemkhani, 2007). Southwest (SW) Asia in an extended sense including the Middle East, European parts of Turkey, Transcaucasia, Turkmenistan, Afghanistan, and Pakistan presents a topographically very diverse region.

Southwest and Central Asia have been proposed to be the origin of at least 20 of the 38 accepted C_4 Eudicot lineages (Sage, 2011; Kadereit and Freitag, 2011; Sage, 2016). In fact, both regions can be considered an exceptional areas for the evolution and diversity of C_4 Eudicots in a predominantly Monocot dominated "C_4 world." This and the high diversity of ecological and morphological features and adaptations of SW Asian C_4 plants, as well as the quick degradation of the arid regions in the Middle East by improper agriculture, overgrazing, water mismanagement, desertification, and climate change (Motagh et al., 2008; Akhani, 2015b) and finally the need of detailed information on the C_4 taxa of this region, inspired compilation of this work.

The aims of this work are: 1) to present an "as complete as possible" checklist of C_4 species of SW Asia, with information on their respective ecological, floristic, and anatomical characteristics, where available; 2) to publish new $\delta^{13}C$ stable isotope ratios for taxa with previously unpublished isotope data; 3) to analyze the distribution of C_4 plants in SW Asia in relation to climatic variables and to evaluate different adaptations of the three main taxonomic groups of C_4 plants (Chenopodiaceae, Cyperaceae, Poaceae) in relation to climate; 4) to discuss shortly the economical and ecological importance of major regionally cultivated or wild growing C_4 crops in a region highly affected by climate change and desertification.

The biodiversity of C_4 plants of Southwest Asia is of particular interest to comprehend the evolutionary history of C_4 plants and the interaction of habitats, biogeographic regions, climate and soil in relation to C_3–C_4 domination. This allows to understand and predict future scenarios of the arid belts of the world, affected by global warming.

MATERIAL AND METHODS

Data Collection and Nomenclature

In this article, we treated SW Asia in an extended way (**Figure 1**). The geographical area considered in this article (extended Southwest Asia) includes the territories of the following countries and geographic areas: Afghanistan, Armenia, Azerbaijan, Bahrain, Iran, Iraq, Israel/Palestine, Jordan, Kuwait, Lebanon, Oman, Pakistan, Qatar, Saudi Arabia, Sinai Peninsula, Syria, Turkey, Turkmenistan, United Arab Emirates, and Yemen (**Figure 1**). The C_4 species distribution data, habitat, and altitude preferences used for compilation of the checklist of C_4 plants of SW Asia were obtained from standard floras, regional contributions, revisions, monographs, reports, and data bases and electronics sources (**Supplementary Appendix Table 1**). Data from Herbarium collections were obtained from the Herbarium of H. Akhani (Halophytes and C_4 Plants Research Laboratory, School of Biology, University of Tehran), the Royal Botanic Garden Edinburgh Herbarium (E), and the Herbarium of Russian Academy of Sciences—V. L. Komarov Botanical Institute (LE). Finally, we received some unpublished data such as Cyperaceae of the Arabian Peninsula, kindly provided by Dr. David A. Simpson through personal communications (Royal Botanical Gardens Kew).

The published literature has been screened for all C_4 species in the area, their respective habitat, distribution, life form, choro-, morpho-, and ecotypes. Taxonomic treatment and nomenclature of the C_4 species were mainly based on global databases such as IPNI (2019) and POWO (2019). The naming of families followed the Angiosperm Phylogeny Group classification (APG IV, 2016) with the exception of the family Chenopodiaceae, which is treated as a separate family and not as a part of Amaranthaceae following Hernandez-Ledesma et al. (2015). The polymorphic genus *Calligonum* was treated taxonomically in accordance with the taxonomic simplifications proposed by Soskov (2011) and the genus *Tribulus* according to the simplifications proposed by Thomas and colleagues (Al-Hemaid and Thomas, 1996; Varghese et al., 2006) (see *Discussion* for further notes). Plants were classified as C_4 species based on stable carbon isotope ratios ($\delta^{13}C$-values) as far as previous or own data support, leaf anatomy (presence of Kranz-anatomy), and biochemical subtypes (**Supplementary Appendix Table 1**). Species with no specific data available, but taxonomically belonging to pure C_4 clades, were as well included in the list but marked with AR (analysis required). The C_4 biochemical subtypes are based on published literature (**Supplementary Appendix Table 1**). In many cases we have extrapolated the "deduced" subtypes based on respective lineage unless there are evidences of multiple subtypes. Furthermore, species with lacking data, belonging to genera with both C_4 and C_3 clades and unknown attribution were preliminarily excluded from the list and listed separately (**Supplementary Table 2**). Life form, eco- and morphotype categorization for each species were based on the above-mentioned sources and/or own observations. The consideration of a species as native or introduced was based on distribution data and indications from the standard sources. Chorotypes were proposed based on a species distribution data in relation to the boundaries of the phytochoria. We used the phytogeographical system suggested for SW Asia and Africa by White and Léonard (1991) and considered other references such as Zohary (1973); Takhtajian (1992); Miller and Cope (1996); Djamali et al. (2012); Welk (2015) (**Figure 1**).

$\delta^{13}C$ Analysis

A total of 234 plant samples (**Supplementary Appendix Table 1**) with unpublished $\delta^{13}C$-values ($^{13}C/^{12}C$ ratios) have been sampled from herbarium samples. The $\delta^{13}C$ were analyzed according to the standard procedure relative to PDB (Pee Dee Belemnite) limestone as the carbon isotope standard and calculated according to this formula: $\delta = 1{,}000 \times (R_{sample}/R_{standard} - 1)$ (Osmond et al., 1975; Akhani et al., 2009). The samples have been fine ground using a Retsch ball grinder and transferred in microtubes for isotopic measurement. Each sample was weighted to a mass between 1.50 to 1.90 mg at the

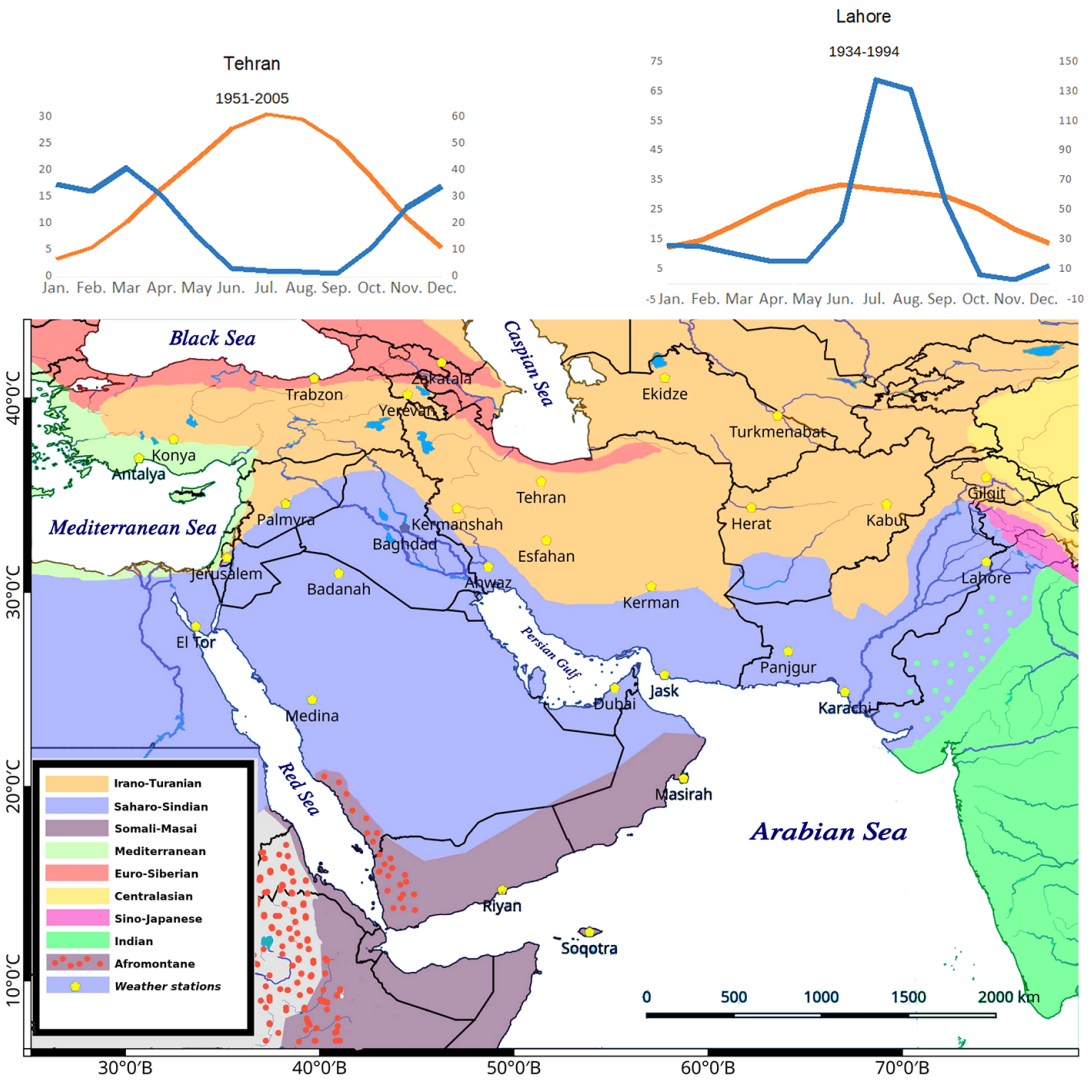

FIGURE 1 | The study area includes Southwest (SW) Asia in an extended sense (**Figure 1**). Phytogeographic boundaries after White and Léonard (1991). We considered 29 weather stations for the evaluation of distribution of main C_4 taxonomic groups in relation to climate. The two extreme climatic conditions showing winter rainfall regime (Tehran) and summer rainfall regime (Lahore) are shown. All other climatic diagrams are depicted in **Supplementary Figure 1**.

SSMIM Mass Spec Lab of the National Museum of Natural History of Paris and burnt in an automated combustion system (EA Flash 2000 Thermo device), interfaced with a DeltaV Advantage Thermo isotope ratio mass spectrometer (continuous flow). The analytical uncertainty within each run estimated from repeated analyses of our laboratory standard (alanine, normalized to IAEA caffeine-600) was lower than 0.08‰ (k = 1) for $\delta^{13}C$ values.

Climate Data and Statistical Analysis

For the correlation of C_4 taxonomic group distributions within the study area and climatic variables, bioclimatic data were extracted from the Worldwide Bioclimatic Classification System (Rivas-Martinez and Rivas-Saenz, 1996-2019; Djamali et al., 2011). For a few stations we obtained climatic data of the Iranian Meteorological Organization (IRIMO) and Scholte and De Geest (2010) and Raza et al. (2015) (**Supplementary Figure 1**). Variables, representative for the SW Asia, have been extracted and/or calculated:

a. Mean annual daily temperature (T)
b. Mean annual precipitation (P)
c. Continentality index [I_c=T_{max} (mean temperature of warmest month) – T_{min} (mean temperature of coldest month)]
d. De Martonne Annual Aridity Index [P/(T+10)]
e. Duration of dry season (number of months with P<2T)
f. Mean summer precipitation (P_s—mean precipitation of warmest 3 months)

g. Ombrothermic index of summer ($Ios_3=P_{p3}/T_{p3}*10$), where P_{p3} is the precipitation of the whole summer and T_{p3} the sum of the mean temperatures for each month of the summer).

The distribution and diversity of C_4 plants in relation to climatic variables (annual mean daily temperature, annual mean precipitation, De Martonne aridity index, continentality index, mean summer precipitation, duration of dry season, and ombrothermic index of summer) has been calculated by linear correlation analysis. The correlation is considering the number of total C_4 plant species, as well as the numerically and ecologically most important taxonomic groups of C_4 plants, e.g., number of C_4 Poaceae, C_4 Cyperaceae and C_4 Chenopodiaceae, and the C_4 Monocot/Eudicot ratio. The data have been imported into "OriginPro," which has been used to calculate the linear correlation and Pearson correlation coefficient and for graphical design (Origin Pro, 2017).

RESULTS

A complete list of all SW Asian C_4 species with life form, chorotype, ecotypes, $\delta^{13}C$-values, metabolic and Kranz anatomical subtypes is presented in **Supplementary Appendix Table 1**.

General Statistics

A total number of 923 (861 native and 62 introduced) C_4 species belonging to 166 genera, 48 independent C_4 lineages, 19 families, and 9 orders have been known from extended SW Asia (**Supplementary Appendix Table 1**, **Figure 1**). For the taxonomic diversity of SW Asian C_4 plants view **Figure 2**.

In the case of *Polycarpaea* (a polyphyletic genus with both C_3 and C_4 species), we have only included one species in our list, considering the fact that we could not verify the photosynthetic pathway of seven additional species, reported from the area (mostly from Socotra). A list of these species is given in the **Supplementary Table 2**.

Distribution of C_4 Species by Country

The number of native and introduced C_4 species in individual SW Asian countries/regions and respective Monocot/Eudicot proportions are shown in **Figures 3** and **4** respectively. The highest diversity of C_4 plants has been documented for Pakistan (account for 43% of all known native SW Asian species), Yemen (38%), Iran (36%), Saudi Arabia (36%), Afghanistan (30%), and Oman (28%), respectively. The differences in the proportion of C_4 Monocot/Eudicots allowed us to categorize countries into three groups: 1) countries with remarkably high percentage of C_4 Eudicots, such as Turkmenistan and Iran; 2) countries with more

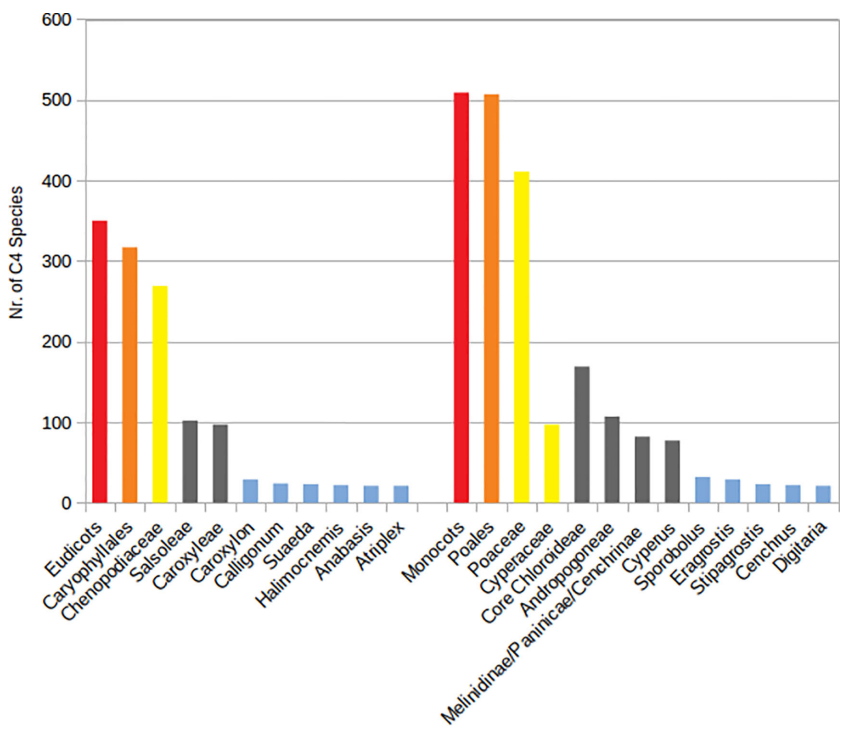

FIGURE 2 | Main taxonomic groups of Southwest (SW) Asian C_4 plants. Red—clades, orange—species richest orders, yellow—species richest families, gray—species richest independent C_4 lineages, blue—genera with more than 20 species.

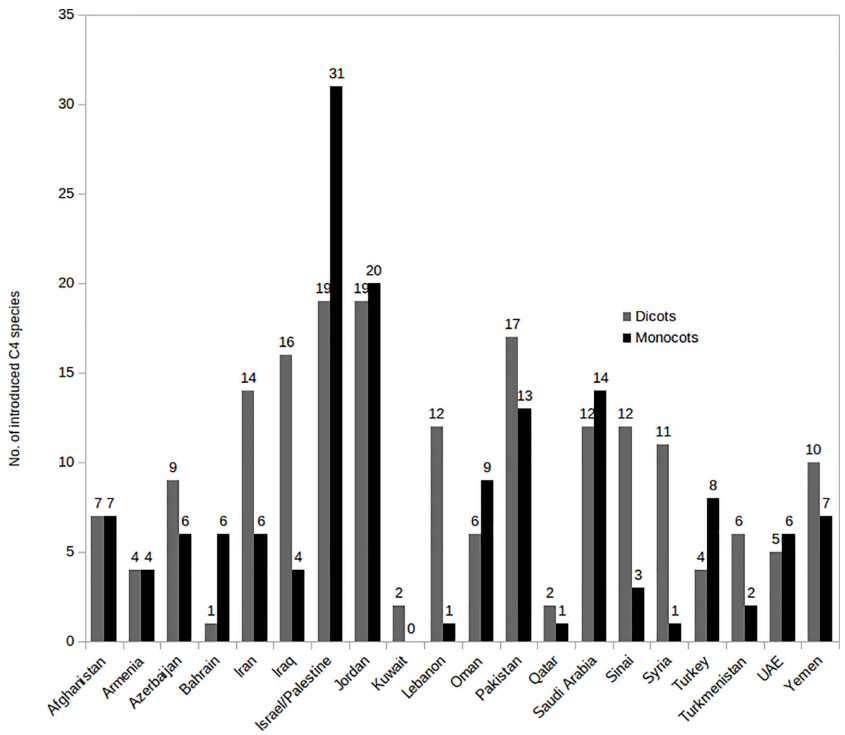

FIGURE 3 | Distribution of introduced C_4 Monocots and Eudicots per country in various Southwest (SW) Asian countries.

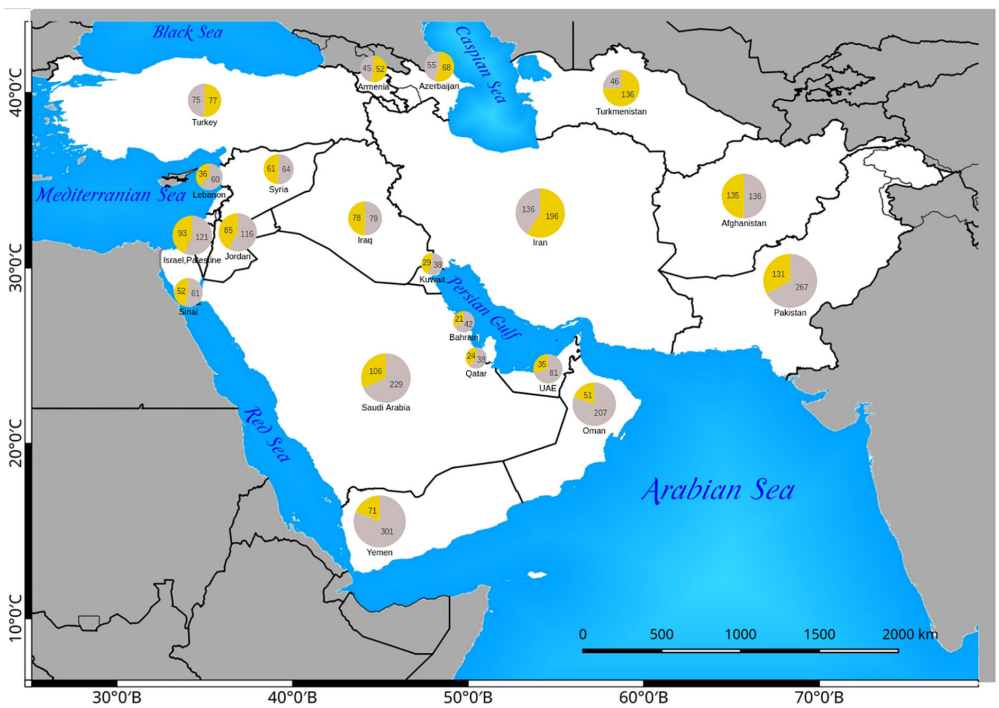

FIGURE 4 | Distribution of native Eudicot (yellow) and Monocot (gray) species in Southwest (SW) Asian countries. The absolute number of C_4 Monocots and Eudicots are indicated in the respective pie charts.

or less equal proportion of Monocot/Eudicots such as Armenia, Azerbaijan, Afghanistan, Turkey, Syria, Iraq, and Sinai Peninsula; 3) counties with higher percentage of C_4 Monocots, that is all other countries located in southern parts of the region.

The countries with the highest number of introduced C_4 plants are Israel and Palestine with 50 introduced species, Jordan with 39 introduced species, and Pakistan with 30 introduced species respectively. The percentage of C_4 plants in relation to total number of recorded plant species per country is highest in Kuwait (23%), Bahrain (22.5%), Qatar (22%), and Oman (20%), respectively (**Figure 5**).

C_4 Endemics of Southwest Asia

One hundred forty-one C_4 species (36 Monocots and 105 Eudicots, 93 of which are Chenopodiaceae) are endemics of SW Asia; 74 of those (22 Monocots and 52 Eudicots) are strict "country endemics." The highest number of "country endemics" are documented in Iran (27 species), Yemen (14 species, 9 of which are endemic to the island of Socotra), Afghanistan (8 species), and Oman (7 species), respectively. The highest number of endemism occurs in Chenopodiaceae (93 species) and Poaceae (31 species). The three endemic richest genera (with 10 or more endemic species) are *Halothamnus* (12 endemic sp. in SW Asia), *Halimocnemis*, and *Climacoptera* (respectively 10 endemic sp. in SW Asia).

The only generic C_4 endemic of the area is the monotypic genus *Halarchon* (*Halarchon vesiculosum*) restricted to Afghanistan. Except a few old records outside of SW Asia, the range of three known species of *Bienertia* is limited to this area.

Climate Correlation

The species-richness of C_4 Chenopodiaceae increases with increasing continentality and decreases with increasing mean summer precipitation (**Figures 6A, E**). The C_4 richness of Cyperaceae increases with increasing mean annual temperature and is negatively affected by increasing continentality (**Figures 6B, G**). The number of C_4 Poaceae increases with increasing mean annual daily temperature (**Figure 6H**). Totally, the diversity and abundance of C_4 plants increases with increasing annual daily temperature and duration of the dry season and decreases with increasing continentality (**Figures 6C, F, I**). Finally, the prevalence of Monocots over Eudicots is related to increasing average summer precipitation (**Figure 6D**).

Life Forms and Ecotypes

Life forms and ecotypes of the C4 plants are shown in **Figures 7A, B**.

Phytogeographic Distribution of C_4 Species of Southwest Asia

The phytogeographic distribution of C_4 species of SW Asia are shown in **Figure 7C**.

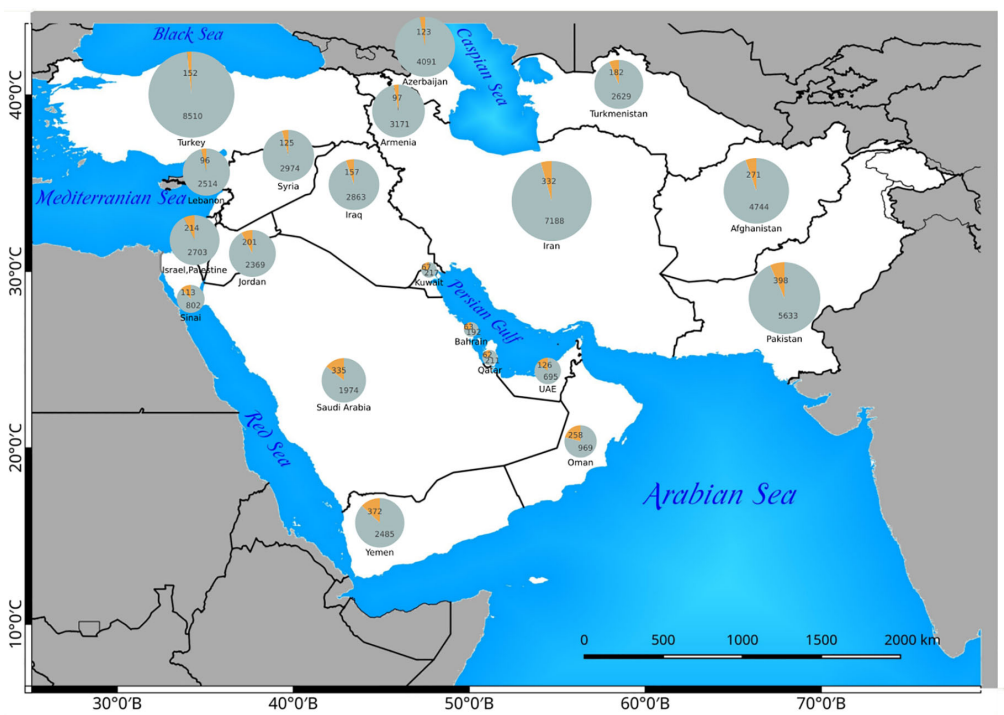

FIGURE 5 | The distribution of C_4 plants (orange) in relation to the total flora (gray) of any Southwest (SW) Asian country. The absolute number of C_4 and C_3 species are indicated in the related pie charts.

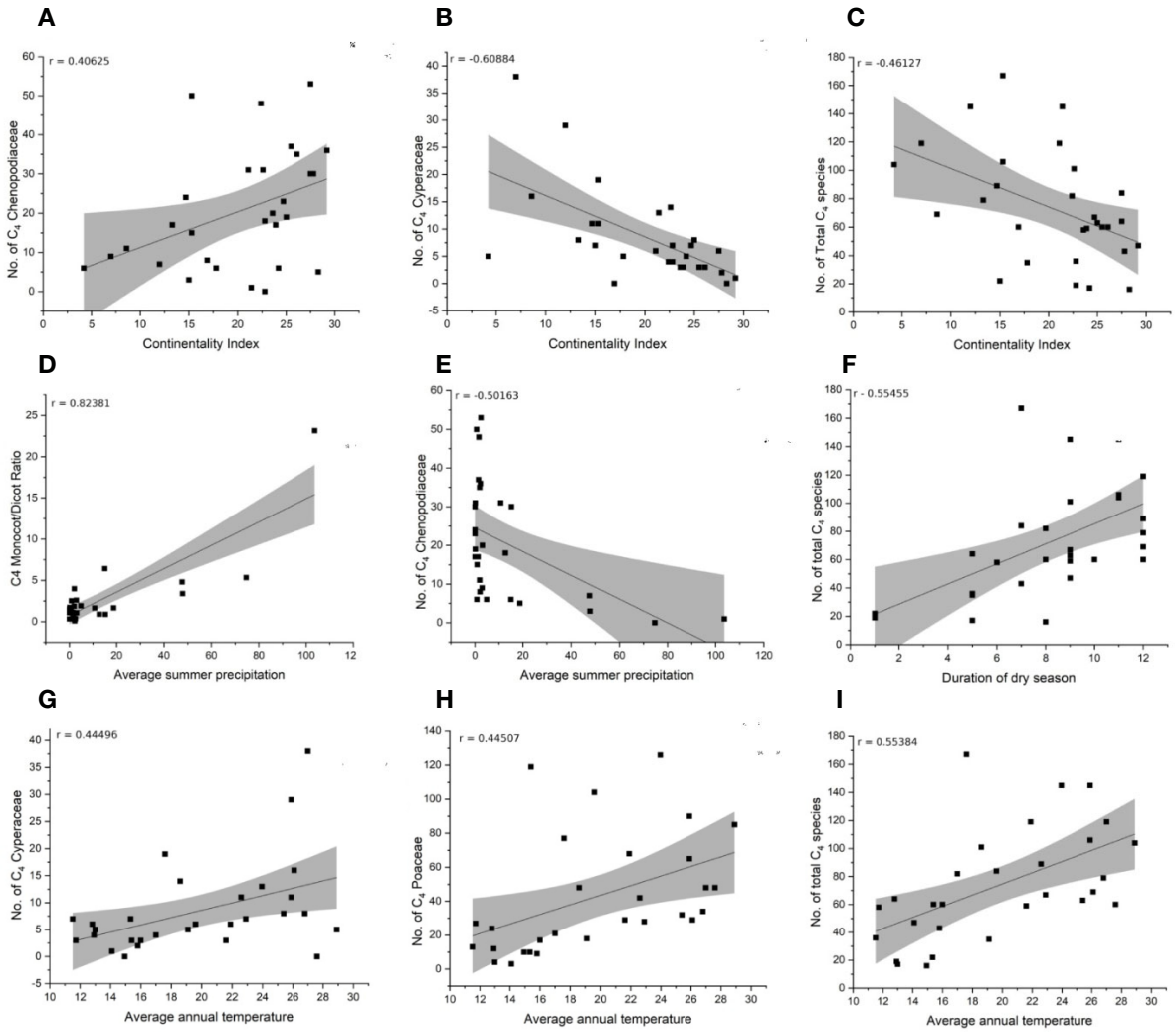

FIGURE 6 | Correlation of climate variables and distribution of C_4 main taxonomic groups (Chenopodiaceae, Cyperaceae, Poaceae, Monocot/Eudicot ratio). **(A–C)** Continentality index. **(D, E)** Average summer precipitation. **(F)** Duration of dry season. **(G–I)** Average annual temperature.

Occurrence of C_4 Biochemical Subtypes

Based on major taxonomic groups (C_4 Poaceae, C_4 Cyperaceae, and C_4 Eudicots), the division in metabolic subtypes is as follows. Within the C_4 Eudicots NAD-ME metabolic types occur in the *Cleome gynandra*, *Cleome angustifolia*, *Amaranthus* spp., C_4 *Atriplex*, Caroxyleae tribe, *Bienertia*, *Suaeda* sect. *Salsina*, *Suaeda* sect. *Schoeberia*, *Calligonum*, *Gisekia*, *Hypertelis*, *Blepharis*, *Anticharis*, and *Tetraena simplex* C_4 lineages (ca. 213 species). The NADP-ME subtype occurs in *Flaveria* clade A, *Euploca*, C_4 *Polycarpaea*, C_4 *Aerva*, *Gomphrena*, C_4 *Althernanthera*, C_4 *Camphorosmeae*, C_4 *Salsoleae*, *Boerhavia*, *Euphorbia* subgen. *Chamaesyce*, and *Tribulus/Kallstroemia* C_4 lineages (ca. 167 species). C_4 Sesuvioideae and *Portulaca* include both NADP-ME and NAD-ME metabolic types.

Within Poaceae a NAD-ME subtype (including mixed NAD-ME/PEP-CK subtypes) is present in *Centropodia*, Melinidinae, Panicinae, and core Chloroideae (ca. 213 species), while a NADP-ME subtype (including mixed PEP-CK/NADP-ME subtypes—for further information consult **Supplementary Appendix Table 1**) is present within *Alloteropsis*, tribe Andropogoneae, *Aristida*, tribe Arundinelleae, Tristachyideae, *Digitaria*, tribe Cenchrinae, *Paspalum*, and *Stipagrostis* (228 species). Within the Cyperaceae (C_4 *Cyperus*, C_4 *Fimbristylis*, and *Bulbostylis* C_4 lineages, 100 species) and within the Hydrocharitaceae (*Hydrilla* and *Elodea* C_4 lineages) all species known to perform NADP-ME metabolic type.

Isotope Data

Two hundred thirty-four δ^{13}C-isotope ratios, mainly for species with previously unpublished data, are presented in the **Supplementary Appendix Table 1** (data marked with * and +). Additionally, the δ^{13}C-isotope ratios of the C_3 plants *Polycarpaea caespitosa* Balf., *Polycarpaea spicata* Wight ex Arn., *Polycarpaea repens* (Forsskal)

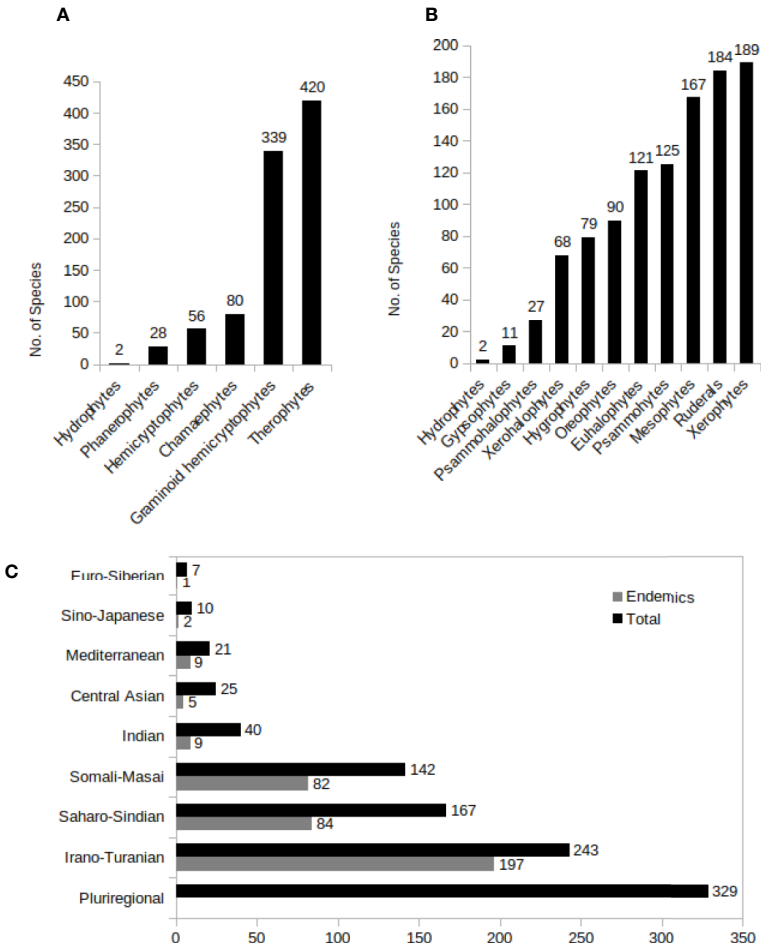

FIGURE 7 | (A–C) C_4 Species distribution in different growth types **(A)**, predominant ecotypes **(B)**, and different phytogeographic regions in Southwest (SW) C_4 plants. A plant was considered as pluriregional, when occur in more than three phytochorion, and as endemic, when present in just one phytochorion.

Aschers. and Schweinf., *Cyperus pulcherrimus* Willd. ex Kunth, and *Fimbristylis turkestanica* (Regel) B. Fedtsch., distributed in Southwest Asia, are published in **Supplementary Table 3**.

DISCUSSION

Southwest Asia Center of Origin of Major C_4 Flora

The checklist appended in this paper has been compiled with great caution to include as much data as available, thanks to intensive Flora compilation in SW Asian countries published during more than half a century (Rechinger, 1963-2015; Davis, 1966-2001; Guest and Ghazanfar, 1966-2013; Nasir and Ali, 1970-2003; Nikitin and Geldikhanov, 1988; Miller and Cope, 1996) and intensive botanical explorations in the region. However, as a first contribution, it requires more data to fill some gaps on local flora of the Levant (Syria, Jordan and Lebanon) and the absence of up-to-date information on specific groups such as Cyperaceae in the Arabian Peninsula.

In spite of our efforts to check photosynthetic types of all putative C_4 groups, we could not get enough samples for a few *Polycarpaea* species. The preliminary phylogenetic studies show that *Polycarpaea* is polyphyletic including both C_3 and C_4 species. It has been suggested to segregate C_4 species in a separate genus *Polia* (Kool, 2012). Within our area, however, only *Polycarpaea corymbosa* (L.) Lam. is a reliable C_4 species. Some other SW Asian species such as *P. repens* (Forsskal) Aschers. and Schweinf., *P. spicata* Wight ex Arn., *P. hassalensis* Chamberlain, and *P. haufensis* A.G. Miller) have been reported to be C_3 according to Kool (2012) and a few ones (**Supplementary Table 3**) require further investigation.

So far 174 species of *Calligonum* have been described worldwide (Soskov, 2011), however phylogenetic studies and attempts to barcode these species revealed little information to support this diversity (Tavakkoli et al., 2010; Li et al., 2014; Doostmohammadi et al., 2020). We followed the recent

monography of the genus which accepts a wide species concept including only 28 species and 8 interspecific hybrids (Soskov, 2011) and recent minor changes by Shi et al. (2016).

The polymorphic, mainly Saharo-Sindian genus *Tribulus*, has been partly reviewed for India and Saudi Arabia by synonymizing many species (Al-Hemaid and Thomas, 1996; Varghese et al., 2006). A systematic review of the genus in our area is highly welcome.

Despite recent progress in the taxonomy and phylogeny of Chenopodiaceae (Kadereit et al., 2003; Akhani et al., 2007), still there are ambiguities in monophyly of some genera such as *Hammada*. The phylogenetic tree based on combined nuclear and chloroplast markers showed polyphyly of three species *Hammada articulata*, *H. salicornica*, and *H. griffithii* (Akhani et al., 2007; Schüssler et al., 2017). In a long debate on the nomenclatural status of the *Kali*-clade within Salsoleae which has been separated from *Salsola* s.l. based on strong molecular and morphological data, recent unexpected decision on replacing the type of the genus *Salsola* by *Salsola kali* L. instead of *S. soda* by International Code of Nomenclature (Akhani et al., 2014; Mosyakin et al., 2017; Turland et al., 2018) resulted in instability and chaos of all names used since 2007. Therefore, in order to keep phylogenetic classification of Salsoloideae we are pushed to change the name of many species traditionally classified in *Salsola* into *Soda* (see **Nomenclatural Appendix**). Furthermore, we provide new combinations for some species which have been overlooked in the phylogenetically based system of Salsoleae by Akhani et al. (2007).

The phylogenetic relationships within Poaceae have seen several recent revisions (Hodkinson et al., 2002; Kellogg, 2015; Soreng et al., 2015; Soreng et al., 2017). These revisions regarded also changes within the PACMAD clade (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, and Danthonioideae subfamiles), which includes all C_4 Poaceae lineages. While some phylogenetic relationships may need further clarification (e.g., genera *Saccharum*, *Miscanthus*, *Miscanthidium*, etc.), several clades have been synonymized [e.g., *Urochloa* (=*Brachiaria, Snowdenia*)] or separated (e.g., *Narenga* separated from *Saccharum*). In the nomenclatural appendix we provide new combinations for the species that have been affected by the last revision by Soreng et al. (2017).

Based on our data (**Table 1**), the C_4 flora of SW Asia includes 923 (ca. 11% of world known C_4 species) and represent 48 of 65 known C_4 lineages of the world (Sage, 2016; Akhani et al., 2007; Kadereit and Freitag, 2011). The area, as one of the major center of diversity of C_4 Eudicots, harbors the origin of ca. 19 C_4 Eudicot lineages and has representatives of all families known to have C_4 species either as native or introduced (Sage, 2016) (**Table 1**, **Supplementary Appendix Table 1**).

SW and Central Asia which represent largely the Irano-Turanian flora are the center of origin of at least 12 C_4 Chenopodiaceae lineages (Akhani et al., 2007; Sage, 2011; Sage, 2016). We consider Salsoleae s. str. as one C_4 origin with understanding that present topologies suggest two additional origins that are not well resolved (Akhani et al., 2007; Kadereit and Freitag, 2011).

The region shows also the highest diversity of species, anatomical and ecological types, and life forms in C_4 Eudicots (see **Supplementary Appendix Table 1**). Of those lineages, 11 (*Suaeda* sect. *Salsina*, *Suaeda* sect. *Schoberia*, *Bienertia, Camphorosma, Bassia*, C_4 Salsoleae s. str., Caroxyleae, *Salsola* (=*Kali*), *Nanophyton*, C_4 *Atriplex*, and C_4 *Tecticornia*) are distributed throughout SW Asia. The origin of the strictly psammophytic genus *Calligonum* (Polygonaceae) has been proposed to be in northern Iran (Irano-Turanian region), on the former shores of the Tethys sea, from where it spread to Central, South, and Southwest Asia and northern Africa (Soskov, 2011). On the other hand, C_4 Aizoaceae, C_4 Zygophyllaceae (*Tetraena simplex* and *Tribulus*/*Kallstroemia* lineages), the *Cleome angustifolia*, and C_4 *Aerva* lineages and probably C_4 *Gisekia* and C_4 *Polycarpaea* have originated on the Arabian Peninsula and adjacent Africa (Sage, 2011).

C_4 Dominated Vegetation in Southwest Asia

C_4 dominated plant communities in SW Asia occur in a wide range of habitats:

1) *Psammophytic C_4 vegetation* occurs in inland sandy deserts and coastal dunes in the following types of vegetation: a) The large sand deserts of the Irano-Turanian and Saharo-Sindian vegetation in Iran, Turkmenistan, Afghanistan, Iraq, and the Arabian Peninsula. The Irano-Turanian sandy deserts are mainly dominated by pure or mixed communities of highly specialized psammophytic C_4 Eudicots [(*Calligonum* spp. (**Figures 8J**, **9E**)], *Haloxylon* spp., *Xylosalsola* spp.)] and Monocots (*Stipagrostis* spp.) (Nechayeva, 1992; Ghasemkhani et al., 2008; Fayvush and Aleksanyan, 2016). In the deserts of Turkmenistan and Central Iran *Haloxylon* communities may be rather densely populated, forming "Saxaul forests" (**Figure 8K**). From central Iran toward the Levant and the Rub Al-Khali the dominating C_4 Eudicots change [(*Calligonum* spp., *Hammada salicornica* (**Figure 8I**), *Cornulaca* spp., *Anabasis articulata*)] while the diversity of psammophytic C_4 Monocot communities increases [*Cyperus* spp. (**Figure 9A**), *Stipagrostis* spp. (**Figure 9E**), *Aristida* spp., *Centropodia* spp. (**Figure 9D**), *Panicum* spp.] (Ghazanfar and Fisher, 1998), similarly the sand dune vegetation in Lut desert in SE Iran and the deserts of South-East Pakistan are dominated by communities of *Calligonum polygonoides, Calligonum mongolicum*, *Soda stocksii* (*Salsola stocksii*), *Tribulus* spp. (**Figure 10H**), *Aerva javanica, Lasiurus scindicus* and *Cymbopogon jwarancusa*, *Stipagrostis multinerva, Desmostachya bipinnata* (Dasti and Agnew, 1994). b) Hamadas, gravel, and coarse sand deserts show rather sparse vegetation, sometimes dominated by sparse communities of chenopods like *Hammada salicornica* (**Figure 8I**), *Cornulaca monacantha*, and *Anabasis* spp. in the tropical deserts of the South SW Asia (Ghazanfar and Fisher, 1998; Akhani, 2015a). However, the temperate deserts of Iran are covered by dense or sparse grasslands of *Stipagrostis plumosa*, often mixed with *Artemisia* subshrubs. c) Coastal dune vegetation varies within the region. E.g., the sandy dunes of the Indus delta and Pakistani Balochistan present mixed

TABLE 1 | C4 Lineages and number of species in each genus in 20 Southwest (SW) Asian countries.

Lineage	Genus	Distribution in SW Asia																		
		Afghanistan	Armenia	Azerbaijan	Bahrain	Iran	Iraq	Israel/Palestine	Jordan	Kuwait	Lebanon	Oman	Pakistan	Qatar	Saudi Arabia	Sinai	Syria	Turkey	Turkmenistan	UAE	Yemen
Monocots																					
ALISMATALES																					
Hydrocharitaceae																					
1) Egeria	*Elodea*																	1			
2) Hydrilla	*Hydrilla*	1				1							1								
POALES																					
Cyperaceae																					
3) Bulbostylis	*Bulbostylis*												2					2			4
4) C4 Cyperus	*Cyperus*	22	6	12	6	24	16	17	14	4	10	22	46	16	23	9	11	15	5	9	39
5) C4 Fimbrystylis	*Fimbristylis*	4	1	3		3	3	3	3		2	3	13		4	2	3	3	1	2	6
Poaceae																					
6) C4 Alloteropsis	*Alloteropsis*												1								
7) Andropogoneae	*Andropogon*							1	1		1	1			1	1	1	1			6
	Apluda	1					1					1	1								1
	Arthraxon	3	1			1						6	3		1			1			3
	Bothriochloa	3	2	2		2	1	1			1	3	3		2		1	1	1		2
	Capillipedium											1	2								
	Chrysopogon	2	1	1		2	1	1	1		1	3	5	2	2		1	1			3
	Cleistachne											1									
	Coix	1																			
	Cymbopogon	3			3	2	2	2	2		1	4	6	1	3		2			2	5
	Dichanthium	1			2	2	1	2	2	2		4	2	2	3	2		1		2	3
	Diectomis																				1
	Dimeria											1									
	Elionurus					1	1					1	1		1						2
	Euclasta											1									
	Eulaliopsis	1											1								
	Hackelochloa											1									1
	Hemarthria	1					1	1	1		1		2		1		1	1			
	Heteropogon	1				1	1				1	2	1		1						1
	Hyparrhenia	1			1	1	1	1	1		1	1	1		1	1	1	1			7
	Imperata	1	1	1		1	1	1	1	1	1	1	1		1	1	1	1	1		1
	Ischaemum												2		1						1
	Iseilema												1								
	Lasiurus	1				1	1	1	1	1		1	1	1	1					1	1
	Microstegium			1		1							2								
	Miscanthus	1																			
	Mnesithea	1											1								
	Narenga												1								
	Ophiuros												1								
	Phacelurus	1						1	1		1		1				1	1			
	Pogonatherum	1											2		1						
	Polytoca												1								
	Pseudodichanthium											1									
	Pseudopogonatherum												1								
	Rottboellia											1									1
	Saccharum	4				3		1	1		1	2	5		2	1	1		1	1	2
	Schizachyrium											1	2								
	Sehima											1	2		1						2
	Sorghum	1	1	1		1	1	2	2	1	2	2	3	1	2	2	2	1	1	1	1
	Spodiopogon						1	1	1		1		1				1	1			
	Thelepogon												1								
	Themeda	1					1				1	1	1		2		1	2			1
	Tripidium	2	1	1		2	2	2	2		3	1	2		1		3	2	1	1	1
8) Aristida	*Aristida*	2	1	1	1	2	1	2	2		1	7	7	3	9	2	1	1	1	4	12
9) Centropodia	*Centropodia*	1			1	1	1	1	1	1		1		1	2	1				2	2
10) Core Chloridoideae	*Acrachne*	1										1	1		1						1
	Aeluropus	3	2	2	2	3	2	2	2	2	1	1	3	1	2	2	2	2	3	2	1
	Chloris	2		1	3	2	2	4	4	1		5	5	3	4				1	4	5
	Cleistogenes	1	1	1		1							1					1	2		
	Coelachyrum					1						3	1		3					2	3
	Ctenium														1						
	Cynodon	1	1	1	1	2	1	1	1	1	1	1	2	1	2	1	1	1	1	1	1
	Dactyloctenium	2			1	2	1	1	1	2	1	4	3	1	3	1	1	1		2	5
	Desmostachya	1				1	1	1	1			1	1		1	1				1	1
	Dignathia											1									1
	Dinebra	1				1	1	1	1	1		1	1		2						1
	Diplachne					1		1	1	1			1	1	1					1	1
	Disakisperma											1			2						2
	Eleusine	2	2	1	1	1		1	1			3	2		4	1		1	1		4
	Enneapogon	1	1	1		2	1	3	3			4	3		5				1	2	6
	Enteropogon											1	1		2						2
	Eragrostis	1	5	5		8	7	11	10	2	3	11	15	3	17	3	4	6	4	6	19
	Eustachys											1									1

(Continued)

TABLE 1 | Continued

Lineage	Genus	Distribution in SW Asia																		
		Afghanistan	Armenia	Azerbaijan	Bahrain	Iran	Iraq	Israel/Palestine	Jordan	Kuwait	Lebanon	Oman	Pakistan	Qatar	Saudi Arabia	Sinai	Syria	Turkey	Turkmenistan	UAE	Yemen
	Fingerhuthia	1										1	1		1						1
	Halopyrum					1	1				1		1	1		1				1	1
	Harpachne														1						1
	Leptocarydion																				1
	Leptochloa	1						2	2												
	Leptothrium											1	1		1						1
	Lepturus																				4
	Melanocenchris					1						1	2		1						2
	Microchloa											1			1						1
	Muhlenbergia	3											3								
	Neyraudia	1											1								
	Orinus												1								
	Oropetium											1			3						2
	Schmidtia																				1
	Schoenefeldia														1						1
	Sporobolus	4	4	3	2	5	4	9	9	1	5	16	8	2	14	2	5	5	3	2	23
	Tetrachaete											1									1
	Tetrapogon	1			1	1		1	1		1	2	2		3	1	1		1	1	4
	Tragus	1	1	1		2					1	2	3		2		1	1	1	2	2
	Trichoneura											1			1						1
	Trigonochloa																				2
	Tripogon	1										6	2		4						6
	Triraphis											1			1						1
	Urochondra											1	1		1						1
	Zaqiqah														1						1
11) Arundinellae	*Arundinella*											2	1								
	Garnotia											1									
12) Tristachyideae	*Danthoniopsis*					1							1		1						1
	Loudetia											1									
13) Digitaria	*Anthephora*					1		1	1						2	1					2
	Digitaria	5	2	4	1	6	1	2	2	1	1	6	13		6	1	1	2	2	2	8
14) Echinochloa	*Echinochloa*	4	1	2	1	3	2	2	2	1	2	2	5	1	4	2	2	3	2	2	2
15) Panicum/ Cenchrus/Urochloa/ Setaria	*Cenchrus*	5	1	1	5	9	3	9	9	4	4	8	10	4	19	6	4	1	1	5	18
	Eriochloa		1	1		1	1					1	2		1				1		2
	Melinis									1		1	1		1						2
	Moorochloa	1	1	1			1	1	1			1	1		1			1	1	1	1
	Panicum	3	2	1	1	6	3	6	6	2	1	5	10	1	6	2	1	4	2	3	8
	Setaria	3	2	2	2	3	3	5	5	3	4	6	7	2	7	2	4	3	2	2	10
	Tricholaena	1			1	1		1	1		1	1	1		1		1			1	2
	Urochloa	3			1	2		4	4		1	7	7		10		1			3	15
16) Paspalum	*Paspalum*	1	1	1	1	1	1	1	1		1	1	3		3	2	1	3	1	2	3
17) Stipagrostis	*Stipagrostis*	9	1	1	2	13	5	9	9	4	2	11	7	4	12	10	2	1	4	8	13
Eudicots																					
ZYGOPHYLLALES																					
Zygophyllaceae																					
18) Tribulus/ Kallstroemia	*Tribulus*	3	1	1	1	3	3	3	3	1	1	3	3	3	3	2	1	1	2	2	2
19) Tetraena	*Tetraena*				1	1		1	1			1	1		1					1	1
MALPIGHIALES																					
Euphorbiaceae																					
20) Euphorbia subgen. *Chamaesyce*	*Euphorbia*	6	5	7	1	9	7	9	9	3	6	6	12	1	11	3	5	8	2	3	10
BRASSICALES																					
Cleomaceae																					
21) Cleome angustifolia	*Cleome*																				1
22) Cleome gynandra								1	1			1	1		1					1	1
CARYOPHYLLALES																					
Polygonaceae																					
23) Calligonum	*Calligonum*	5	1	3	1	17	3	2	1	1	1	3	2	1	3	1	1	2	15	2	2
Caryophyllaceae																					
24) C4 Polycarpea	*Polycarpaea*												1		1						1

(Continued)

TABLE 1 | Continued

Lineage	Genus	Distribution in SW Asia																		
		Afghanistan	Armenia	Azerbaijan	Bahrain	Iran	Iraq	Israel/Palestine	Jordan	Kuwait	Lebanon	Oman	Pakistan	Qatar	Saudi Arabia	Sinai	Syria	Turkey	Turkmenistan	UAE	Yemen
Amaranthaceae																					
25) C4 Aerva	*Aerva*	1			1	1	1	1	1	1		1	1	1	1	1				1	1
26) C4 Alternanthera	*Alternanthera*							1	1				2		1						1
27) Amaranthus	*Amaranthus*	7	4	8	2	11	10	16	15	2	11	5	12	3	9	12	11	15	8	4	8
28) Gomphreneae	*Gomphrena*						1						2		1						2
Chenopodiaceae																					
29) C4 Atriplex	*Atriplex*	10	3	3	1	11	2	6	5	3	4	2	9	1	6	6	5	7	6	2	4
30) Camphorosmeae	*Bassia*	9	5	5	1	8	1	5	4	3	1	1	8	1	7	3	3	3	6	1	1
31) Bassia	*Camphorosma*	1	1	1		1							1					1	1		
32) Tecticornia	*Tecticornia*												1								
33) Caroxyleae	*Caroxylon*	13	6	7	2	20	9	6	6	3		4	9	3	12	3	3	10	9	3	4
	Cimacoptera	8	1	1		9	3	1					2				2	2	9		
	Halarchon	1																			
	Halimocnemis	5	4	5		12	2	1	1				2					3	12		
	Halocharis	6				3	2						4		1		1		3		
	Kaviria	4	2	3		7	3					1	2						3	1	1
	Petrosimonia	1	2	3		3												3	2		
	Piptoptera	1				1													1		
	Pyankovia	1				1													1		
34) Nanophyton	*Nanophyton*																		1		
35) C4 Salsoleae	*Anabasis*	5	2	4	2	12	3	5	4	1	1	1	3	1	4	2	1	1	9	1	1
	Arthrophytum	1																	1		
	Cornulaca	2			1	2	3	1	1	2	1	2	2	2	4	1	2	1	1	1	2
	Girgensohnia	4	1	1		4	1	1					1					1	2		
	Halogeton	2					1	1	1	1	1		3		1	1	1		1		
	Halothamnus	7	1	1		9	3	2	2	1	1	1	3		3		3	1	5	1	1
	Haloxylon	2				2	1	1	1			1	1		1	1			2	1	
	Hammada	4			1	1	3	5	5	1	2	1	4	1	1	4	2		2	1	
	Horaninovia	2				4													3		
	Iljinia																		1		
	Lagenantha																				1
	Noaea	2	2	1		1	1	1	1				1		1			2	1		
	Soda	2	2	3	1	8	2	5	4		3	5	1	3	3	3	4	3	1	1	
	Sevada											1			1						1
	Traganum						1	1	1	1	1			1	1	1	1				
	Turania	1				1													5		
	Xylosalsola	3				2							2						5		
36) Salsola	*Salsola*	6	2	5		6	1	1					4		1			3	3		1
37) Bienertia	*Bienertia*	1	1	1		3	1			1			1	1	1			1	1	1	
38) Suaeda sect. Salsina	*Suaeda*	6	3	3	2	8	5	6	6	2	3	4	5	2	4	4	4	2	4	2	4
39) Suaeda sect. Schoberia		1	1	1		5	1	1	1		1		2			1	1	6	2		
Gisekiaceae																					
40) C4 Gisekia	*Gisekia*	1				1		1	1			1	1		1					1	1
Aizoaceae																					
41) Sesuvium/ Trianthema/Zaleya	*Sesuvium*												1		1						1
	Trianthema					1		1	1			1	2		4						4
	Zaleya	1				1		1	1			1	1		1					1	1
Nyctaginaceae																					
42) Boerhavia	*Boerhavia*					1		1	1			1	4		2	1				1	2
Molluginaceae																					
43) Hypertelis cerviana/H. fragilis	*Hypertelis*					1							1		1			1			1
Portulacaceae																					
44) Portulacaceae	*Portulaca*	2	1	2	2	1	2	1	1	1	1	2	5	1	5	1	1	1	1	1	3
BORAGINALES																					
Boraginaceae																					
45) Euploca	*Euploca*					1						1	2		2					1	4

(Continued)

TABLE 1 | Continued

Lineage	Genus	Distribution in SW Asia																	
		Afghanistan	Armenia	Azerbaijan	Bahrain	Iran	Iraq	Israel/Palestine	Jordan	Kuwait	Lebanon	Oman	Pakistan	Qatar	Saudi Arabia	Sinai	Syria	Turkey	Turkmenistan	UAE	Yemen
LAMIALES																					
Scrophulariaceae																					
46) Anticharis Acanthaceae	*Anticharis*					1		1	1				2		1	1				1	2
47) Blepharis ASTERALES	*Blepharis*					2		2	2			2	3		2	1					3
Asteraceae																					
48) Flaveria clade A	*Flaveria*						2					1			1	1				1	1

psammophytic communities with frequent presence of *Aerva javanica* (**Figure 10D**), *Cyperus arenarius*, *Soda stocksii* (*Salsola stocksii*), *Hammada salicornica*, and *Cenchrus biflorus* (Snead and Tasnif, 1966; Ananda Rao and Meher-Homji, 1985). The coastal dunes of the Arabian Peninsula are dominated by communities of halophytic sand grasses and sedges like *Zaqiqah mucronata, Urochondra setulosa, Sporobolus* spp., *Dactyloctenium* spp., *Leptothrium senegalense, Halopyrum mucronatum, Cenchrus divisus, Panicum turgidum, Lasiurus scindicus, Coelachyrum piercei*, and *Cyperus* spp. as well as Chenopods like *Cornulaca monacantha and Atriplex stocksii* (Ghazanfar and Fisher, 1998; Brown and Mies, 2012).

2) *Halophytic C_4 vegetation* is highly diverse and affected by topography, water level, and local land use. Saline habitats range from marly and clayey hills with varying salt and gypsum composition to inland dry saline plains with clay, silt, and sandy soils in sabkhas, saline wetlands, and lakes, as well as coastal sabkhas and shorelines. They composed a wide range of communities from pure C_4 communities to C_4 patches occurring in microhabitats of C_3 dominated communities:

a) Dry marly or clayey hills with varying salt and gypsum composition are often dominated by xerohalophytic and gypsophytic C_4 Chenopods. In Iran such habitats are sparsely vegetated by xerohalophytic and gypsohalophytic shrubs like *Anabasis eugeniae* (**Figure 8F**), *Anabasis calcarea, Anabasis firouzii* (**Figure 10B**), *Halothamnus auriculus, Halothamnus lancifolius, Xylosalsola arbuscula, Noaea mucronata, Caroxylon verrucosum, Suaeda dendroides*, *Kaviria tomentosa, K. aucheri,* and *K. zedzadii* (Akhani, 2006; Pérez-García et al., 2018). In Israel similar saline chalk and marl slopes are dominated by communities of *Suaeda asphaltica, Hammada negevensis*, *and Caroxylon tetrandrum* (Danin and Orshan, 1999).

b) Saline plains, salt marshes, and depressions with varying water-table offer a habitat to pure or mixed C_4 Chenopod communities, that are typical for the Irano-Turanian and Saharo-Sindian regions. Such communities are divided into two main subgroups namely "the C_4 dominated shrubby communities (vegetation class Haloxylo-Kavirietea tomentosae)" and "the C_4 rich Irano-Turanian nitrophilous annual halophytic communities (vegetation class Caroxylo-Climacopteretea)." The C_4 shrubby dominated communities are composed of *Haloxylon ammodendron*, *Kaviria tomentosa*, *Hammada* spp., *Halothamnus subaphyllus*, *H. glaucus*, *Soda rosmarinus* (*Seidlitzia rosmarinus*), *Cornulaca monacantha*, *Anabasis aphylla, A. haussknechtii, A. iranica* spp., and *Suaeda fruticosa* (Akhani, 2004). The C_4 rich Irano-Turanian nitrophilous annual halophytic communities composed of a rich variety of annual chenopods like *Climacoptera* spp. (**Figures 8G** and **10I**), annual *Caroxylon* spp., *Halimocnemis* spp., *Petrosimonia* spp., *Cornulaca aucheri,* annual *Atriplex* spp., *Halocharis* spp., *Pyankovia brachiata, Bienertia* spp. (**Figures 8G** and **10J**), annual *Suaeda* spp. *(S. cochlearifolia, S. gracilis, S. microsperma, S. khalijefarsica, S. arcuata, S. aegyptiaca, S. altissima), Bassia*

FIGURE 8 | (A–K) Habitats dominated by C_4 vegetation: **(A)** habitat with high water table in Northeast (NE) Iran, near Caspian Sea, dominated by *Saccharum spontaneum L.* **(B)** euhalophytic vegetation of *Aeluropus litoralis* (Gouan) Parl. community, close to the SE Caspian shores, Turkman Sahra salt flats, Iran; **(C)** C_4 grassland in a temperate forest on rocky S-facing slopes of Golestan National Park, Iran; **(D)** *Desmostachya bipinnata* on temporarily flooded dry soils in Khuzestan, Southwest (SW) Iran; **(E)** ruderal vegetation dominated by *Sorghum halepense* (L.) Pers., SE Caspian coasts, Farahabad, Mazandaran, Iran; **(F)** gyspsiferous outcrop of *Anabasis eugeniae* Iljin and *Anabasis calcarea* (Charif & Aellen) Bokhari & Wendelbo community, NW Iran; **(G)** euhalophytic community in central Iran, 60 km W Tehran mainly dominated by *Bienertia cycloptera* Bunge ex Boiss. and *Climacoptera turcomanica* (Litv.) Botsch. **(H)** shores of the Makran coast (Pakistan) dominated by *Tecticornia* indica (Willd.) K. A. Sheph. & Paul G. Wilson, **(I)** *Hammada salicornica* (Moq.) Iljin community in Desert Lut, S. Iran; **(J)** *Calligonum amoenum* Rech. f. on moving dunes in Lut desert; **(K)** *Haloxylon persicum* Bunge shrubland near Mesr Village, c. 40 km E of Jandagh, Dasht-e-Kavir, Iran. All photos H. Akhani.

spp. (*B. hyssopifolia, B. erinatha, B. eriophora*) (**Figure 10A**), and *Soda florida* (*Seidlitzia florida*). (Ghazanfar and Fisher, 1998; Akhani, 2004; Akhani, 2006). This vegetation type develops on disturbed nitrified soils on salinized wastelands, ruderal habitats around the roads, and human settlements and as pioneer communities on exposed high saline soils of dried up salt lakes (Ghorbanalizadeh et al., 2020).

c) The vegetation of coastal saline flats resembles their inland counterparts. Such vegetation is present on the SE shores of the Caspian sea, along the coasts of the Persian Gulf, the Indian Ocean, and in lesser extent along the Red Sea. E.g., the coastal sabkha vegetation of the Persian Gulf is dominated by *Suaeda fruticosa, Caroxylon imbricatum, Atriplex leucoclada, Climacoptera* sp., *Soda rosmarinus* (*Salsola rosmarinus*), and *Bienertia sinuspersici*, while the Red Sea and Indian Ocean coastal flats harbor communities of *Suaeda monoica* (**Figure 10C**), *S. moschata, S. vermiculata/S. fruticosa, Atriplex leucoclada, and A. coriacea* (Ghazanfar and Fisher, 1998; Akhani and Deil, 2012; Akhani, 2015a). Similar communities occur along the Makran coast and in the saline flats of the Indus delta, these communities also include *Suaeda baluchestanica*, *Halopyrum mucronatum*, and the remarkable tropical *Tecticornia indica* (**Figure 8H**) (Snead and Tasnif, 1966; Voznesenskaya et al., 2008).

d) A remarkable type of C_4 halophytic vegetation are C_4-grass communities present on saline clayey soils with high water-table, formed by halophytic grasses of the genus *Aeluropus* (**Figure 8B**). Such communities occur e.g., on the eastern shores of the Caspian Sea, where *Aeluropus littoralis* forms almost pure communities on flooded plains in Iran and

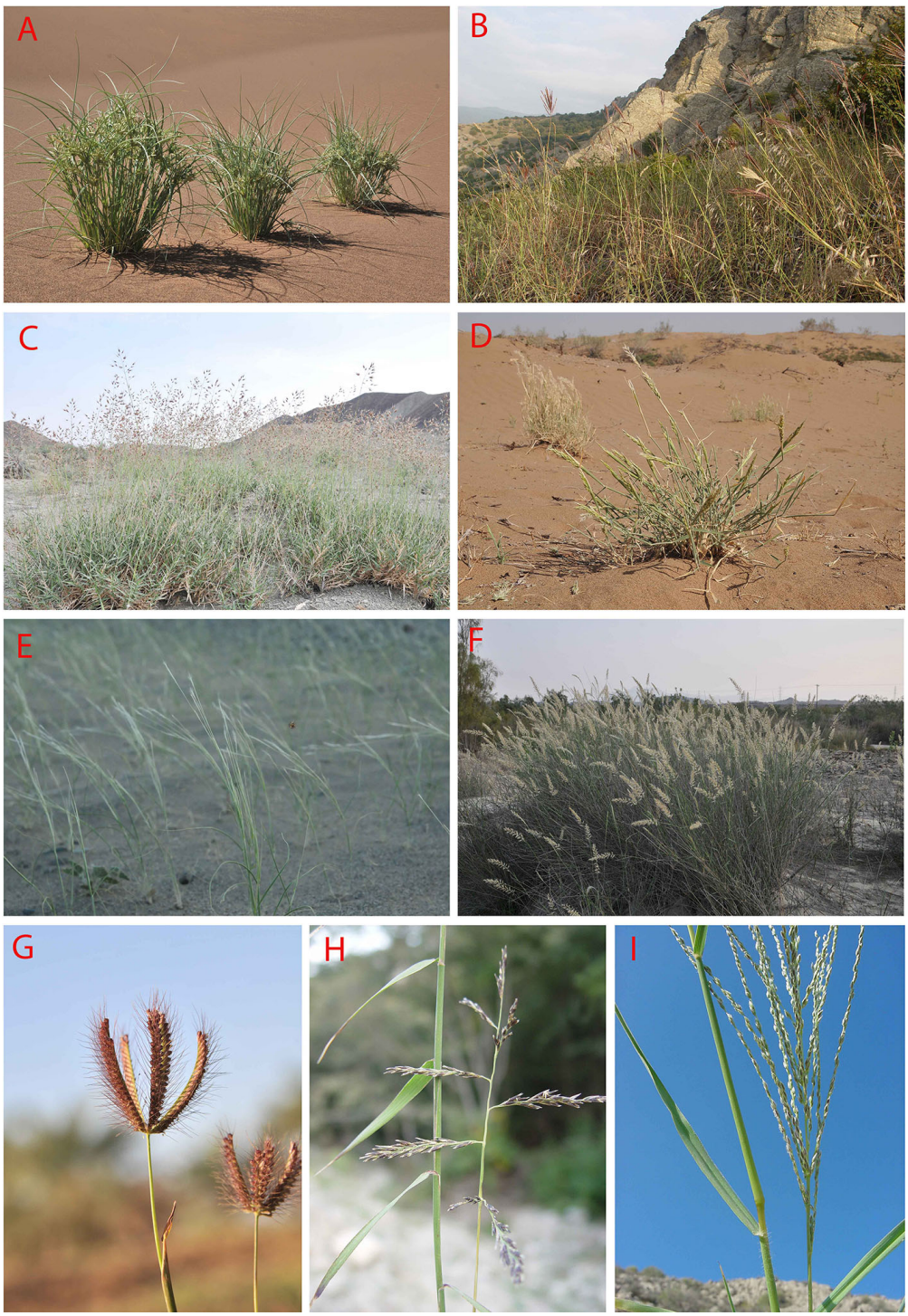

FIGURE 9 | (A–I): Some representatives of important C_4 Monocot lineages in Southwest (SW) Asia: **(A)** *Cyperus aucheri* Jaub. and Spach, sand dune in Desert Lut, Iran; **(B)** *Bothriochloa ischaemum* (L.) Keng-*B. bladhii* (Retz.) S.T. Blake-*Cleistogenes serotina* (L.) Keng community, S-facing rocky outcrop in Golestan National Park, Iran; **(C)** *Danthoniopsis stocksii* (Boiss.) C.E. Hubb., dry river bed, Baluchistan, Iran; **(D)** *Centropodia forskalii* (Vahl) Cope, Aran-Bidgol dunes in Esfahan Province, Iran, 6.6.2010; **(E)** *Stipagrostis multinervis* H. Scholz, Desert Lut, Iran, 1.4.2011; **(F)** *Cenchrus divisus* (J.F. Gmel.) Verloove, 19.2.2013, river side in Bahukalat, Baluchestan, Iran; **(G)** *Chloris barbata* Sw., 17.2.2013, ruderal places, in Zehkalut, Kerman, Iran; **(H)** *Cleistogenes serorina* (L.) Keng, 13.10.2003, limestone rocky outcrops in Golestan National Park, Iran; **(I)** *Digitaria nodosa* Parl., 18.12.2001, rocky shrubland, Kuhe Geno, Hormozgan, Iran (photos by H. Akhani).

FIGURE 10 | (A–K): Some representatives of important C_4 Eudicot lineages in Southwest (SW) Asia: **(A)** *Bassia eriantha* (Fisch. and C.A. Mey.) Kuntze, 10.6.2004, *Artemisia+Stipagrostis* semi-desert steppe, 50 km SE Esfahan, Iran; **(B)** *Anabasis firouzii* Akhani, 10.10.2012, marl slopes, Qorkhod Protected Area, N. Khorassan, Iran; **(C)** *Suaeda monoica* Forssk. ex J.F. Gmel. 18.6.2007, Red Sea coasts, Aqabeh, Jordan; **(D)** *Aerva javanica* (Burm.f.) Juss. ex Schult. 22.2.2013, rocky slopes, Hormoz Island, Persian Gulf; **(E)** *Euphorbia serpens* Kunth, 3.7.2014, ruderal soil, Khuzestan, Iran; **(F)** *Blepharis ciliaris* (L.) B.L. Burtt, 22.2.2013, gravely-sandy soils, Hormoz Island, Persian Gulf; **(G)** *Trianthema portulacastrum* L., 23.2.2013, ruderal places, Hormoz Island, Persian Gulf; **(H)** *Tribulus macropterus* Boiss. 18.9.2001, Khuzestan, Iran; **(I)** *Climacoptera chorassanica* Pratov, 1.9.2003, hypersaline soils, 40 km SE Birjand, Khorassan, Iran; **(J)** *Bienertia kavirense* Akhani, 5.8.2016, Esfahan/Semnan, Iran; **(K)** *Caroxylon abarghuense* (Assadi) Akhani and Roalson, 5.10.2001, Touran protected Area, Iran (photos by H. Akhani).

Turkmenistan (Nechayeva, 1992; Rukhlenko, 2001). Similar pure stands and mixed communities of *Aeluropus lagopoides* with *Tamarix* spp., *Desmostachya bipinnata*, and *Salvadora oleoides* have been reported from inland Sabkhas and salt pans of Iran, Syria, the Arabian Peninsula, and Pakistan (Dasti and Agnew, 1994; Ghazanfar and Fisher, 1998; Akhani, 2006; Al-Oudat and Qadir, 2011). Finally, *Aeluropus* spp. pure and mixed communities (in association with *Suaeda monoica, Fimbristylis* spp., etc.) as well as communities of other halophytic grasses, like *Sporobolus* spp. and *Halopyrum mucronatum* are also frequent on the shores of the Indian Ocean and the Persian Gulf (Snead and Tasnif, 1966; Ananda Rao and Meher-Homji, 1985; Ghazanfar and Fisher, 1998; Akhani, 2015a).

3) *C_4 grassland vegetation* is limited by high temperatures and the availability of water in the form of summer precipitation or high underground water table (Ghazanfar and Fisher, 1998). The following types of C_4 grassland communities have been so far described:

a) Oreophytic grasslands on the rocky highlands of S Arabia and Socotra may be distinguished by pure C_4 or mixed C_3 and C_4 grassland communities with the presence of local afromontane endemics. These are mainly secondary grasslands resulting from deforestation of local shrublands (Ghazanfar and Fisher, 1998). However, some examples of high-altitude primary grasslands exist in suitable habitats. Examples are high-altitude primary low growing mixed grasslands (up to 3,000 m) of the Arabian endemics *Festuca*

cryptantha (C_3), *Andropogon crossotos* (C_4), *Tripogon oliganthos* (C_4), and cold resistant *Stipa* spp. (C_3); grasslands dominated by *Andropogon bentii* on Socotra (600 m) and associations of the C_4 grass *Elionurus muticus* with the dwarf shrub *Macowania ericifolia* (Elionuro mutici-Macowanietum ericifoliae) on the mountainous basalt flats in Central Yemen (Deil and Müller-Hohenstein, 1985; Ghazanfar and Fisher, 1998; Brown and Mies, 2012). Examples of tall secondary grasslands in S Arabia are high altitude communities (2,000–3,000 m) dominated by *Themeda triandra, Heteropogon contortus, Andropogon distachyos, Bothriochloa insculpta, Cenchrus unisetus*, *Hyparrhenia hirta*, *Arthraxon lancifolius* and *A. hispidus* (Themedo-Hyparrhenietea); dry grasslands dominated by *Tetrapogon villosus* (C_4), *Elionurus muticus* (C_4), *Stipa tigrensis* (C_3), and *Cenchrus setaceus* (C_4) and tall monsoon facing grasslands (above 750 m) dominated by *Themeda quadrivalvis*, *Apluda mutica, Setaria pumila*, *Heteropogon contortus*, *Sporobolus spicatus*, and *Arthraxon junnarensis* (Ghazanfar and Fisher, 1998; Brown and Mies, 2012).

b) In appearance similar C_4-grass dominated communities exist on exposed S-facing slopes of Caspian forests (Akhani and Ziegler, 2002). A combination of edaphic and climatic factors, e.g., presence of summer rainfall in one hand and edaphic constraint as a result of rocky substrate with usually poor soil layer in another hand are advantageous for the formation of a C_4 grasslands in steep S-facing slopes, dominated by C_4 grasses like *Bothriochloa ischaemum* (**Figure 9B**), *B. bladhii, Cleistogenes serotina* (**Figure 9H**), *Heteropogon contortus, and Cenchrus orientalis* (**Figure 8C**). The arboreal vegetation is made by N temperate elements such as *Carpinus orientalis, Zelkova carpinifolia, Quercus castaneifolia, Colutea buhsei, Crataegus* spp., *Cotoneaster* spp., and *Rosa* spp. A comparable vegetation forms in olive orchards in Caspian lowlands as a result of artificial irrigation during summer time have been observed (Akhani, unpublished). The South Caspian C_4 grasslands are paralogues to the Mediterranean C_4 dominated grassland of the order Cymbopogono-Brachypodietalia ramosi (=Hyparrhenietalis hirtae) (Diez-Garretas and Asensi, 1999; Mucina et al., 2016). In the western Himalayas (E Afghanistan, N Pakistan), where monsoon provides sufficient rainfall, similar grasslands with *Cymbopogon, Chrysopogon, Heteropogon, Aristida*, etc. grow in association with *Acacia modesta* and *Olea cuspidata* woodlands. *Chrysopogon-Cymbopogon* grasslands are also typical for the uplands of Pakistani Balochistan (Suttie and Reynolds, 2003).

c) Xeromorphic grasslands occur mainly in lowlands of the Arabian Peninsula and in southern Iran. They occur rarely as separate communities but rather with open *Acacia* pseudo-savannas or mixed with shrubs like *Calligonum* sp., *Leptadenia pyrotechnica*, or *Lycium shawii* or with xeromorphic dwarf shrubs like *Hammada salicornica* and *Caroxylon cyclophyllum*. The dominating species is the C_4 grass *Panicum turgidum*, which mostly forms pure grass communities or co-occurs with other C_4 grasses, like *Lasiurus scindicus, Aristida mutabilis, Dichanthium* spp., *Cenchrus* spp. (**Figure 9F**), *Stipagrostis* spp., and *Aristida* spp. (Kürschner, 1986; Ghazanfar and Fisher, 1998).

4) *Meso- and hygrophytic communities* are known on river shores, irrigation channels, seasonal rivers, and wadis of Turkmenistan, Afghanistan, Pakistan, and Iran, as well as on non or slightly salty sandy soils with high water-table, close to Caspian shores, with intensive human disturbance, with patches and communities of tall C_4 grasses like *Imperata cylindrica*, *Saccharum spontaneum*, and *Tripidium ravennae*. *Saccharum spontaneum* (**Figure 8A**) communities often co-occur with *Tamarix* spp. shrubs or tall C_3 grasses, like *Phragmites* spp. and *Arundo donax*, and in transition zones, with more constant water-table these patches form sometimes transitions with more hygrophytic C_4 sedges, like *Cyperus* sp. or *Fimbristylis* sp. (Chaudhri et al., 1966; Nechayeva, 1992; Chandran, 2015). In Israel passes the northern distribution of another important C_4 hygrophyte, *Cyperus papyrus*. Once forming communities together with C_3 grasses like *Phragmites* and *Arundo*, as well as *Typha*, its communities unfortunately largely disappeared because of habitat destruction and pollution (Danin and Orshan, 1999). In drier regions of S Iran and Arabia more xerophytic grasses, like *Desmostachya bipinnata* (**Figure 8D**), *Cymbopogon* sp., etc. often replace this kind of vegetation in wadi beds and seasonally flooded plains (Kürschner, 1986; Ghazanfar and Fisher, 1998; Akhani, 2015a).

5) *Ruderal C_4 communities* can be classified as follows: C_4 ruderal and weedy communities in irrigated fields or adjacent disturbed places (**Figure 8E**): In most parts of the Irano-Turanian region and Caspian lowlands the weedy communities develop during summer times dominated by species of *Sorghum*, *Euphorbia* subgen. *Chamaesyce* (**Figure 10E**), *Bothriochola*, *Setaria*, *Eleusine*, *Echinochloa*, *Digitaria* (**Figure 9I**), *Paspalum*, *Cynodon*, *Imperata cylindrica*, *Cyperus* spp., *Atriplex* spp. (mostly *Atriplex tatarica*), *Portulaca oleracea*, *Bassia scoparia*, *Salsola tragus, Suaeda altissima*, *S. arcuata*, *Caroxylon* spp., *Tribulus terrestris*, and several *Amaranthus* species. The species combination varies based on the irrigation regime, soil, and management of weed control. In southern parts of SW Asia the diversity of more thermophilous ruderal grasses like *Cenchrus, Panicum, Eragrostis, Dactyloctenium, Chloris* (**Figure 9G**), *Sporobolus, Urochloa, Hyparrhenia hirta, Desmostachya bipinnata* (**Figure 8D**) as well as Eudicots like *Boerhavia repens*, *Euphorbia* subgen. *Chamaesyce, Zaleya pentandra, Trianthema portulacastrum* (**Figure 10G**), *Aerva javanica, Suaeda aegyptiaca* increases (Nasir and Ali, 1970-2003; Shmida and Aronson, 1986; El-Ghanim et al., 2010; Abdel Khalik et al., 2013; El-Sheikh, 2013).

C_4 Eudicots Are Related to Dominance of Continentality Index

Our findings show different tendencies of C_4 Monocot and C_4 Eudicot distributions in the study area. The usual pattern of increase of C_4 species along a latitudinal gradient is more or less

similar to global patterns (Pyankov et al., 2010, **Figure 6**) with deviations mostly due to its complex topography, edaphic factors, and the resulting presence of specific microclimates. The C_4 grasses increase along the southern and eastern edges of the region (Yemen, Oman, Pakistan) with higher summer precipitation due to monsoon and tropical climates (Ghazanfar and Fisher, 1998), **Figure 6D**. In fact, C_4-grasslands are known to depend on a dry and rain seasonality, where bushfires during dry seasons on one hand prevent forest growth while high temperatures on the other hand favor C_4 grasses (Skinner et al., 2002; Bond et al., 2005; Keeley and Rundel, 2005; Hoetzel et al., 2013). The savanna-like C_4 grasslands at the south-eastern corner of the Caspian forests are supported by a small peak of summer precipitation, while the spring flora is dominated by C_3 grasses and forbs (Akhani and Ziegler, 2002). The C_4 poor but rainfall rich Euro-Siberian portions of SW Asia also favor C_4 Monocots over C_4 Eudicots. For example, in the north-western edges of the study area the percentage of Monocot species in the local C_4 flora may reach 84% (Zaqatala, Republic of Azerbaijan) ('GBC').

The high proportion of C_4 Eudicots in Iran and Turkmenistan may be explained by large saline and sandy deserts and high continentality, which favor halophytic Chenopodiaceae species and psammophytic *Calligonum* spp. The increase of C_4 Eudicots in W-E direction even in similar longitudinal belts may be explained by a combination of edaphic and climatic conditions (compare the opposite tendency in China, Wang and Ma, 2016). Indeed, the continentality index clearly increases along a W-E direction over the SW Asia to Central Asia (Djamali et al., 2011; Djamali et al., 2012). The harsh summer times with scarcity of fresh water resources in deserts of Iran and Turkmenistan reduce the competitive advantages of C_3 species (Pyankov et al., 2010; Sage and Sultmanis, 2016). Additionally, absence of summer rainfall supresses C_4 grassland formation. Although higher continentality adversely affects general C_4 domination (**Figure 6C**), this is not the case for C_4 Chenopodiaceae, which are adapted to temperate deserts with continental climate (**Figure 6A**). This is explained by the phenology of chenopods with an estival active growing season (Toderich et al., 2007). Many chenopods and species of *Calligonum* are highly specialized by their morpho-anatomical and physiological traits to live under harsh conditions (e.g., their long root systems have access to the underground and subsurface water-table) (Gintzburger et al., 2003; Soskov, 2011). The negative correlation of C_4 Cyperaceae with continentality (**Figure 6B**) and positive correlation with average annual temperature relate to their sensitivity to low winter temperatures (Wang and Ma, 2016). C_4 Cyperaceae however do not seem to be unaffected by precipitation. The later may be explained by a high variety of ecotypes within C_4 *Cyperus* (the main bulk of C_4 Cyperaceae), ranging from psammophytic xerophytes to hygrophytes limited to permanent wetlands and the consequent species shift in relation to various ecological conditions.

Finally, it is interesting to note, that while different metabolic subtypes often indicate the adaptation of C_4 Monocots to specific ecological conditions, e.g., NADP-ME Monocots are more likely distributed in areas with high rainfall, while NAD-ME Monocots grow in conditions with lower rainfall (Schulze et al., 1996), this feature doesn't seem indicative in C_4 Eudicots E.g., xerophytic C_4 chenopods show both NAD-ME or NADP-ME metabolisms (see **Supplementary Appendix Table 1**).

Southwest Asia—Center of Biodiversity of Single-Cell C_4 and C_3–C_4 Switching Plants

One of the fascinating aspects of C_4 photosynthesis is the discovery of Single-Cell functioning C_4 photosynthesis in two C_4 lineages of the Chenopodiaceae (Voznesenskaya et al., 2001; Edwards et al., 2004; Akhani et al., 2005). This photosynthetic type was described for the first time in *Suaeda aralocaspica* (Bunge) Freitag and Schütze (= *Borszczowia aralocaspica* Bunge), a hygrohalophyte from the saline depressions of Central Asian semideserts (Freitag and Stichler, 2000; Voznesenskaya et al., 2001). *Bienertia* as a monophyletic lineage in which all species perform single-cell functioning C_4 was discovered shortly after *S. aralocaspica* with some new species (Voznesenskaya et al., 2002; Akhani et al., 2003; Akhani et al., 2005; Kapralov et al., 2006; Akhani et al., 2012). The genus *Bienertia* is diversified mostly in Iran and some surrounding areas often on moist and highly saline soils in association with several annual C_4 chenopods belonging to Caroxylo-Climacopteretea class in the interior Iran or on open habitats of saline shrublands on the lowlands around the Persian Gulf between *Tamarix* species or on tidal shores (Akhani et al., 2003; Akhani et al., 2009).

Instead of a conventional system of C_4 terrestrial species, having a dual-cell compartment consisting of mesophyll and bundle sheath cells, in both single-cell C_4 lineages, this achieved by localizing photosynthetic machinery in a single-cell type. In *Suaeda aralocapica* dimorphic chloroplasts are polarized in a single layer mesophyll cell, in which the proximal chloroplasts fix CO_2 using PEPC into a C_4 acid which moves to distal chloroplasts *via* a cytoskeleton network (Edwards et al., 2004; Chuong et al., 2006). Similarly, the *Bienertia* species single-cell system has a unique form in which lateral chloroplasts function as mesophyll cells and a bubble-like central chloroplast compartment (CCC) acts as Kranz-cells in usual C_4 species. This discovery stimulated scientists to deeply investigate the biology and genomics of this simplified system, which might have advantages for those looking for genetic engineering of C_4 photosynthesis in C_3 crop plants (Schuler et al., 2016).

Another peculiarity within the C_4 plants of SW and Central Asia, is the presence of two types of photosynthesis within the life cycles of particular lineages and species of Chenopodiaceae. In these species C_3 cotyledon leaves are replaced by C_4 shoots. This characteristic is widespread in the subfamily Salsoloideae and rarely in Suadedoideae (Pyankov et al., 1999; Pyankov et al., 2000b; Pyankov et al., 2001; Akhani and Ghasemkhani, 2007). In both tribes of Salsoleae and Caroxyleae several genera, such as *Haloxylon, Halothamnus, Hammada, Girgensohnia, Noaea* and *Soda inermis* (*Salsola soda*), *Climacoptera, Halimocnemis, Petrosimonia, Kaviria*, and *Halocharis* are known to have this switching mechanism. In Suaedoideae, this type was known in

Suaeda microphylla evidenced by carbon isotope values (Akhani and Ghasemkhani, 2007) or anatomy (Khoshravesh and Akhani, unpublished data). As this characteristic is of interest for gene engineering, the transcriptomes of *Haloxylon ammodendron* and *Soda inermis* (*Salsola soda)* have been studied (Li et al., 2015; Lauterbach et al., 2017).

Ecologically, the development of a switching mechanism from a C_3 to a C_4 photosynthetic metabolism hasn't however received much attention. Switching chenopods are mainly halophytes and xerohalophytes of continental temperate saline ecosystems (*Climacoptera, Petrosimonia, Halimocnemis, Soda inermis*), gypsiferous (*Halothamnus*), and sandy (*Haloxylon*) habitats of the Irano-Turanian floristic region and are taxonomically among the main and most biodiverse taxa of SW Asian C_4 Eudicots. Switching plants are also among the main biomass producers in the Irano-Turanian deserts, suggesting that the switching mechanism may imply an evolutionary advantage to those species. The continental climates of their habitats may probably favor a switching mechanism and the presence of C_3 cotyledons at early developmental stages, when germination at low temperatures favors the presence of C_3 cotyledons while increasing temperatures during the growth period favor C_4 leaves (Akhani and Ghasemkhani, 2007). Further investigations however are needed to comprehend better the ecological advantages in comparison with tropical deserts.

Palaeoclimatic Implications

The distribution pattern of C_4 plants in SW Asia is at least partly related to the palaeoclimatic conditions which have prevailed in the region during the Neogene. Today, the region is dominated by the summertime subtropical anticyclones (Zarrin et al., 2009) which induce a long summer drought in most parts of SW Asia. The subtropical anticyclonic system is particularly intensified and maintained by the high elevations in SW Asia (Zarrin et al., 2011) which are mostly present since at least 7 million years ago (Djamali et al., 2012). During the late Neogene, a long summer drought has thus dominated over the region impeding the penetration of moisture-bearing westerlies into the Irano-Anatolian inlands and Central Asia. The continental inlands of SW Asia, although close to the Indian Ocean, receive no monsoon precipitation during the summertime because of the complex monsoon-desert mechanism described by Rodwell and Hoskins (1996). The palaeoclimatic archives suggest that excepting the Arabian Peninsula, most of the continental interior of SW Asia has not received summer monsoon rainfall during the intensification phase of the latter at the beginning of the Holocene (Djamali et al., 2010). Only SE Iran might have received some direct summer precipitation from the summer monsoons some 11,400 to 6,500 years ago (Vaezi et al., 2019). Some of the C_4 plant communities found in currently dry areas of S Iran and Arabian Peninsula (see above) may be the relicts of formerly widespread C_4 communities when the area received more summer rainfall. Relatively higher values of $\delta^{13}C$ of organic matter in the Jazmurian playa sediments during the early Holocene (**Figure 8A** in Vaezi et al., 2019) may indeed reflect the important contribution of more abundant C_4 grasses during the Indian Monsoon intensification phase in SE Iran. With the exception of increasing summer rains in SE Iran and Arabia, it seems thus that most of SW Asia has always been subjected to long summer droughts and high continentality since several million years (Djamali et al., 2012). Such long-lasting geo-climatic conditions are characterized by strong continentality, long summer droughts and presence of intracontinental endorheic basins which support the formation of a broad range of saline environments suitable for the diversification and specialization of C_4 Eudicots in particular the halophytic chenopods.

Human Utilization of C_4 Plants in Southwest Asia

SW Asia including the Fertile Crescent had a long history of plant domestication and land use (Zohary and Hopf, 2000). Deserts and steppe populations utilize many C_4 species in a variety of ways, as source for food, fire wood, for grazing, construction, greening of their surroundings, medicine, and in recent times for desert reclamation programs and afforestation. They have additional potentials such as usage as biofuel, genetic engineering practices and even invention of new crops. C_4 crops exploited in SW Asia either natively originated or widely distributed or imported from other parts of the world together with their main applications are listed in the **Supplementary Table 4**. It has to be noted, that although the wild forms of many C_4 crops, like *Eragrostis tef, Echinochloa frumentacea, Panicum miliaceum, Eleusine coracana, Setaria italica, Soda inermis*, etc. are distributed throughout SW Asia, they are mostly cultivated outside of this region.

The most important and most species rich group of C_4 crops are millets and millet-like cereals cultivated traditionally for their grain in arid areas of S Asia and Africa. Among them several major millet crops like sorghum (*Sorghum bicolor*) (Sanjana Reddy, 2017a), proso millet (*Panicum miliaceum*) (Gomashe, 2017), foxtail millet (*Setaria italica*) (Hariprasanna, 2017), and pearl millet (*Cenchrus americanus*) (Sanjana Reddy, 2017b) made it to fame out of their region of domestication and have been introduced not only to SW Asia but are extensively cultivated worldwide for their grains. Sweet sorghum (*Sorghum bicolor*) is also an alternative source of syrup and sugar (Sanjana Reddy, 2017a), although the main pantropical sugar crop remains the extensively cultivated sugarcane (*Saccharum officinarum*) (James, 2014). Although the wild forms of small millets, are distributed in SW Asia, they are mainly cultivated as traditional cereals of cultural significance outside of this area (Seetharam and Riley, 1986). In fact, they form important grain crops in the traditional communities of S Asia and Subsaharan Africa. An interesting example is cultivation of teff (*Eragrostis tef*) concentrated in Ethiopia and Eritrea, where it is of the most important crop plants and is used mainly for the production of traditional Injera flat bread (Seetharam and Riley, 1986). Teff recently however is wining fame as a healthy alternative cereal outside of E Africa. Molecular studies have shown that this allotetraploid is closely related to *Eragrostis pilosa*, growing in SW Asia (Ingram and Doyle, 2003; Assefa et al., 2017).

Millets cultivated on smaller scales in SW Asia include finger millet (*Eleusine coracana*), Indian barnyard millet (*Echinochloa frumentacea*), Japanese millet (*E. esculenta*), and adlay millet (*Coix lacryma-jobi*). Their cultivation in SW Asia is mainly limited to regions where they bear cultural significance, such as the plains and hills of Afghanistan and Pakistan (Breckle and Rafiqpoor, 2010; Breckle et al., 2013). Corn (*Zea mays*) is the world's most important C_4 grain crop and SW Asia is not an exception, where it is extensively cultivated (Staller, 2010).

Many wild SW Asian C_4 grasses are important fodder and pasture crops, for biomass production or used for landscape greening on large and small scales in arid areas of N America, Australia, S Europe, Central Asia, India, and Subsaharan Africa, namely *Bouteloua curtipendula*, *Cenchrus* spp., *Chloris gayana*, *Cynodon dactylon*, *Diplachne fusca*, *Lasiurus scindicus*, *Panicum antidotale*, *Setaria viridis*, *Sorghum halepense* and *Sporobolus* spp. They may present a source for further millet and forage grass breeding and cultivation for forage and erosion control in disturbed and desertifying areas of SW Asia.

The C_4 Monocots are of high importance for summer grazing of wildlife such as Persian Ibex or livestock on steep rocky outcrops and disturbed or degraded South Caspian forests (Akhani and Ziegler, 2002). Grazing on salt marsh grasslands dominated by *Aeluropus* is common in most parts of the region (Whigham et al., 1993).

The Saharo-Sindian and Somali-Masai vegetations of SW Asia are rich in highly productive high biomass Monocots. Genera like *Panicum, Cenchrus, Desmostachys bipinnata, Saccharum, Tripidium, Mischanthus, Paspalum, Pogonatherum*, and *Cyperus* are high biomass producers that both have a value for grazing and industrial applications (Serag, 2003; Tubeileh et al., 2016; Mullet, 2017).

Haloxylon persicum, H. ammodendron, Xylosalosla richteri, and *Calligonum* spp. may yield up to 1.2 t, 3.0 t, 1.3, and 1.2 t green biomass per hectare respectively, depending on habitat type and population density. *Haloxylon ammodendron* is definitely the largest species by biomass and can reach a height of up to 9 m and an age of up to 100 years (Fet and Atamuradov, 1994; Gintzburger et al., 2003) (**Figure 8K**).

Additionally, psammophytic *Xylosalsola* sp. in Turkmenistan and Central Asia and *Soda stocksii* (*Salsola stocksii*) and *Hammada salicornica* in Pakistan and India are grown for the same purpose. *Haloxylon ammodendron* is mainly cultivated in Central Asia, SE Europe, NW China, and Iran for as erosion control and forage on salt and clay deserts, saline flats, and saline sands. From the Mediterranean toward Iran *Atriplex* spp. like *A. halimus* and *A. canescens* are cultivated for the same purpose on saline clayey soils (Hanelt, 2001; Danin, 2007; Walker et al., 2014). *Bassia prostrata* has been cultivated in Turkmenistan, Central Asia, Europe, and the USA as forage and erosion control on clayey, slightly saline, sandy, and rocky soils (Dzyubenko and Soskov, 2014).

Several C_4 crops are gaining importance as healthy food plants. To mention are seeds of *Amaranthus caudatus*, *A. cruentus*, and *A. hypochondriacus*, used as gluten-free pseudocereals with increasing cultivation worldwide (Sauer, 1967). Furthermore, several C_4 Eudicots are cultivated as leaf vegetables. E.g., *Portulaca oleracea* is cultivated around the Mediterranean, in Iran, Turkey, and Transcaucasia (Gonnella et al., 2010; Uddin et al., 2014). *Suaeda aegyptiaca* is cultivated in southern Iran and *Amaranthus tricolor* in Pakistan, India, and E Asia for food (Akhani, 2006). *Soda inermis*, growing on saline soils throughout Armenia, Iran, Turkey, and Turkmenistan, is cultivated and highly prized as a leaf vegetable (agretti) in the Mediterranean region (Centofanti and Bañuelos, 2015). Another chenopod, the common ruderal and ornamental *Bassia scoparia*, is cultivated for the production of a caviar substitute in E Asia ("Useful Temperate Plants Database"; Han et al., 2006; Nedelcheva et al., 2007). Recent efforts have also been done to introduce locally collected leaf vegetables (*Cleome gynandra*, *Boerhavia* sp. and *Sesuvium sesuvioides*) and medicinal plants [*Blepharis* sp. (**Figure 10F**) and *Tribulus terrestris*] into cultivation (Hanelt, 2001; Cheikhyoussef et al., 2011; Mahesh et al., 2012; Boteva et al., 2014; Kripa and Vijayalakshmi, 2016; Salamon et al., 2016; 'PROTA4U web database', 2018). Furthermore, tuber bearing sedges, like *Cyperus esculentus* are interesting candidates as food crops, already cultivated throughout the world (Pascual et al., 2000; Arafat et al., 2009). Several C_4 Eudicots [e.g., *Gisekia pharnacoides*, *Hypertelis cerviana*, *Suaeda* spp., *Tecticornia indica*, *Atriplex* spp., and *Soda* and *Salosla* spp.], are used locally as famine foods or collected as local leaf vegetables and may be interesting for further investigations as crop plants (Hanelt, 2001; "PROTA4U web database", 2018; "Useful Temperate Plants Database"). Some C_4 plants are cultivated worldwide for their essential oils, such as *Cymbopogon* spp. (citronella oil, palmarosa oil, etc.), *Chrysopogon gryllus*, *Chrysopogon zizanioides* (vetiver oil), and *Cyperus articulatus* (Duke, 1993; Hanelt, 2001; Simpson and Inglis, 2001; Bertea and Maffei, 2009; Atala, 2012; Pareek and Kumar, 2013). Some local species of *Cymbopogon* are already used in local medicine for their aromatic properties and may be of interest for introduction into cultivation (El-Kamali et al., 2005; Bertea and Maffei, 2009; Prasad et al., 2014). A number of C_4 species are also favored for their biomass and as a source of fiber, mat, basket weaving, broom, and construction materials, and could be interesting candidates for the diversification of local agriculture. E.g., *Bassia scoparia*, *Cyperus corymbosus*, *C. malaccensis*, *C. pangorei*, *Desmostachya bipinnata* and *Eulaliopsis binata*, and *Sorghum bicolor* are already cultivated for fibers and high quality mat weaving in Iran, India, and Africa (Wendrich, 1989; Ravichandran et al., 2005; Sahu et al., 2010; Jana and Puste, 2014; Khyade et al., 2018; Shioya et al., 2019).

Finally, some C_4 plants are frequently grown ornamentals (e.g., *Cyperus alternifolius*, *C. papyrus*, *Bouteloua gracilis*, *B. curtipendula*, *Stenotaphrum secundatum*, *Coix lacryma-jobi*, *Imperata cylindrica*, *Miscanthus nepalensis*, *Miscanthus sinensis*, *Schizachyrium scoparium*, *Setaria palmifolia*, *Tripiduium ravennae*, *Amranthus caudatus*, *A. tricolor*, *Gomphrena globosa*, *G. haageana*, *B. scoparia*, *Portulaca pilosa* and *P. grandiflora*).

With climate change, overpopulation and resource mismanagement, SW Asia is highly in need of alternative,

drought, and salt resistant crops to make agriculture more sustainable. A series of C_4 crops, in addition to those already introduced and cultivated in SW Asia, are highly interesting for further introduction into cultivation, since their wild forms are already distributed and adapted to SW Asian climate conditions.

Conservation

Being mainly part of the so-called MENA Region (Middle East and North Africa), SW Asia with its xeric climates is highly susceptible to climate change (Pal and Eltahir, 2016). Although future scenarios vary, concerning the degree of climatic changes, a general consensus on the increase of mean temperatures and heat extremes exist (Evans, 2009; Waha et al., 2017). The same is regarding the decrease of precipitation and increase of drought and aridity, with the exceptions of the southern shores of SW Asia, where the increase of monsoon precipitation due to a shift of the inter-tropical convergence zone is expected to occur according to some predictions (Waha et al., 2017; Byrne et al., 2018). The effects of climate change are enhanced by growing population and an aggressive mismanagement of water and land resources. For example, in Iran an ineffective irrigation agriculture, extensive dam construction and groundwater over-use has led to a significant decrease of groundwater levels and the drying and destruction of the main lake, river, and wetland ecosystems (Motagh et al., 2008; Madani, 2014; Akhani, 2015b; Motagh et al., 2017). On the other hand, the SW Asian C_4 flora, although highly specialized, is very susceptible to minor changes in many extreme habitats (groundwater levels, period, and amount of precipitation, etc.).

Of the 923 C_4 species of SW Asia, 141 (105 Eudicot and 36 Monocots—15.3%) are endemic to SW Asia, while 70 species (50 Eudicots and 20 Monocots—7.6%) are strict country endemics with very limited habitats. However, even some species distributed beyond SW Asia (e.g., several *Cyperus* species described from limited areas of the Somali-Masai floristic region) show very restricted distributions. The strict country endemics can be grouped mainly in two subgroups: a) C_4 Eudicots (mainly chenopods) mainly endemic to habitats of the Irano-Turanian floristic region; b) C_4 Monocots (mainly Poaceae) endemic to the Somali-Masai floristic region (of those the half of the species are endemic to the island of Socotra). Under high climatic and anthropogenic pressure on the narrow habitats of SW Asian C_4 endemics, many such species are critically endangered.

As an example, the recently discovered *Bienertia kavirense* Akhani (**Figure 10J**), restricted to a narrow region within Iran's central saline desert, has been declared critically endangered from its discovery (Akhani et al., 2012). According to our own documentation in a saline flat located 60 km W of Tehran near Rude Shur (saline river), a very dense subpopulation of *Bienertia cycloptera* in 2003 completely disappeared in 2009 (Akhani et al., 2003; Akhani, 2016; Akhani and Rudov, 2018) apparently due to dropping of underground water levels. This tragic situation is observed in many similar habitats, where dropping of underground and subsurface water levels affects soil moisture and is consequently a threat for existence of many C_4 annuals. Recent field trips of the authors discovered, that the narrow habitats of *Halimocnemis alaeflava* and *Halimocnemis azerbaijanensis* have been fragmented and partly destroyed by factory construction, complete removal of upper soil layers, as well as road and dam construction. In fact, if further localities of *H. alaeflava* are not discovered in future, there is a probability of its complete extinction as its type habitat is on the way toward complete destruction. Another critical example is a subpopulation of the local endemic *Caroxylon abarghuense* (**Figure 10K**) in Touran Biosphere Reserve, located in Central East of Iran. In a small valley dominated by *Tamarix* shrubs, only 16 living individuals of *C. abarghuense* have been found. Their seeds do not seem germinable probably because allee effect resulted from small size population. A subpopulation of *Piptoptera turkestana*, discovered in 1989 on sandy dunes of central Iran, ca. 30 km ESE of Kashan, could not be recollected in the same place after extensive searches and apparently disappeared from the locality probably due to habitat disturbance and oil mulching (**Figure 11A**). Several taxa (e.g., *Climacoptera zenobiae*) lack proper assessment and are known from very limited collection samples. Species like *Climacoptera czelekenica* Pratov, being island endemics, depend on the changing water level of the Caspian Sea. On the island of Socotra, the high number of endemic plants is threatened by both climate change and overgrazing (Attorre et al., 2007; Rejzek et al., 2016). This directly affects also the C_4 endemics of Socotra, being all C_4 grasses. Conservation of endangered C_4 endemics is further complicated by the lack of proper population assessment in many regions of SW Asia because of lacks of interest and founding and specially because of inaccessibility due to long lasting military conflicts. The two of the four most endemic rich countries (Afghanistan and Yemen) are both long time battle grounds.

Additional threats to the local C_4 flora are introduced and invasive C_4 plants. The introduced C_4 flora of SW Asia (68 species) is mainly composed of Poaceae (34 species), Amaranthaceae *sensu stricto* (20 species), and Euphorbiaceae (7 species) (Pahlevani et al., 2020). Several C_4 lineages not typical for the region have been introduced to SW Asia (C_4 *Flaveria* clade A, C_4 *Alternanthera*, and *Gomphrena*). The genus *Amaranthus*, although a neophytic genus in most areas of SW Asia, includes not only recently introduced species but also apparently old neophytes (e.g., *A. blitum*) and local species (e.g., *A. graecizans*, *A. tenuifolius*, *A. sparganicephalus*). Of the 67 introduced C_4 species at least 48 have been reported to be invasive in various regions of the world ("CABI—Invasive Species Compendium. Wallingford, UK: CAB International"; Elmore and Paul, 1983). These species may form a major threat to local floras and economic burdens for agriculture and livestock. An important aspect of invasiveness in the area is introducing C_3 invasive species into C_4 habitats. This happened in S Iran and Pakistan where introduction of *Prosopis juliflora* occupied many of the habitats of drought resident species including native C_4 species.

The saline areas, sandy dunes, and marl or gypsum habitats, where the majority of C_4 Eudicots grow, have no protection

FIGURE 11 | (A, B): Mismanagement of C_4 habitats in Iran. **(A)** Oil mulching in sand dunes with rich C_4 plant species in Abuzeidabad, Kashan, Central deserts of Iran, 28.5.2020. **(B)** Overgrazing in gypsum and salty ground between Zanjan and Mianeh, Northeast (NE) Iran, 14.4.2018. (Photos by H. Akhani).

priority in most of the countries of SW Asia. Mostly, these habitats are considered as badlands with poor biodiversity. Grazing and shortage of water are big problems affecting the vegetation in such habitats. Therefore, most of these lands with poor vegetation cover are converted into dust emission centers with a huge environmental concern (Akhani, 2015b). An example of mismanagement of such habitats we refer to intensive oil mulching in many sand dunes in Iran with dense C_4 plant vegetation. Such activities which aim to control dune movement result in destroying natural flora with many C_4 annuals and shrubs (**Figure 11A**). Sadly, improper managements threatened many gypsophytic, halogypsophytic, and xerophytic hill habitats, which are the main habitats of endemic C_4 chenopods (Akhani, 2006; Ghorbanalizadeh et al., 2020) (**Figure 11B**). We strongly recommend re-evaluation of protection policies to restore and protect C_4-rich habitats that are of high advantage in desert areas because of their low water requirements and vegetation cover provision during harsh seasons.

CONCLUSIONS

SW Asia is not only an area of origin and diversification of most interesting and highly adapted C_4 Eudicot lineages, but also provide diverse and vast habitats for growing C_4 Monocots. The evolution of various eco-morphological traits among C_4 Eudicots lineages (notably single-cell functioning C_4 plants and switching C_3–C_4 species) and the presence of many endemic species are indicative of long-lasting ecological pressure that supports speciation and specialization among different families in particular Chenopodiaceae and *Calligonum* (Polygonaceae).

The C_4 Eudicots are known to dominate vast Irano-Turanian deserts. Our data suggest, that this is mainly related to the adaptation of C_4 Eudicots to the continental climatic conditions of the Irano-Turanian deserts in contrast with the C_4 Poaceae, that dominate areas with the presence of summer rainfall in the southern and southeastern parts of the area influenced by monsoon summer rains.

Unfortunately, the SW Asian C_4 plant diversity is threatened by the impact of intensive land use synergized by global warming and rapid desertification. In spite of our knowledge on the taxonomy and phylogeny of many C_4 lineages, many questions regarding SW Asian C_4 plants are however still unresolved. For example, the taxonomy of the genera *Calligonum*, *Climacoptera, Hammada*, *Kaviria*, *Caroxylon*, and *Tribulus* needs still to be clarified. More studies are necessary to understand the phylogenetic relationships of C_3 and C_4 species of the genera *Fimbristylis*, *Polycarpaea* and clarify ambiguities in presence of some C_3 and C_3–C_4 intermediate lineages within prevalently C_4 Salsoloideae (Chenopodiaceae). The study of the vegetation of some neglected areas and the description of some specific C_4 plant dominated communities as well as the compilation of some country floras in the region (e.g., Syria, Jordan) would be highly informative.

Due to political instability and low interest in the conservation of desert areas with lower biodiversity, several rare C_4 endemics are under the threat of extinction. Being part of the ecologically and climatically vulnerable MENA region, SW Asia is also in need of a sustainable management of water resources and agriculture. The C_4 dominated habitats requires protection priority and monitoring of highly adapted plants to harsh environments. We re-emphasize the importance of regional C_4 crops and the selection of new C_4 crop candidates with lesser ecological impact than genetic engineering as a much more sustainable approach to guarantee the food security facing the future global change.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

AR provided the data, wrote the first draft, prepared all figures and tables. HA planned and supervised the research, contributed to the preparation of the data and text, provided all photos, edited the text and jointly worked together in writing of first

draft. MM analyzed all the carbon isotopes in the paper, read, and edited the manuscript. MD contributed to the writing of the palaeoecology part of the paper.

FUNDING

This research was funded by postdoctoral support for the first author by the International Office of the University of Tehran and the Erasmus Mundus Marhaba Program.

ACKNOWLEDGMENTS

The authors thank the curators of the Royal Botanic Garden Edinburgh Herbarium (E) and the Herbarium of Russian Academy of Sciences - V. L. Komarov Botanical Institute (LE) for the provided plant samples, the Iranian National Science Foundation for a travel support to St. Petersburg Herbarium (AR), Sibald Trust for a herbarium visit to Edinburgh Botanical Garden (HA) and LIA HAOMA CNRS Project for funding the isotopic analyses. We would like to thank Mr. Denis Fiorillo at the SSMIM/MNHN for the isotopic measurements and Dr Marie Balasse who authorised us to use the facilities of this Laboratory. We also appreciate the help of Dr. Claudia Deigele (Munich) for providing us the unpublished carbon isotope database of late Prof. Hubert Ziegler. Finally, we would like to thank Dr. David A. Simpson (Royal Botanical Gardens Kew) for his precious personal communication on the Cyperaceae of the Arabian Peninsula and Socotra, Targol Chatrenoor for her advices on the list of *Climacoptera* species and Roxana Khoshravesh for her critical suggestion to improve this paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.546518/full#supplementary-material

SUPPLEMENTARY APPENDIX TABLE 1 | Complete List of C4 plants in Southwest Asia showing their C_4 lineage, life form, chorotype, endemism, ecotype, $\delta^{13}C$ and C_4 subtype and respective leaf or shoot anatomical type.

SUPPLEMENTARY FIGURE 1 | Climatic diagrams of selected 29 stations in different parts of SW Asian countries obtained from of the Iranian Meterorological Organization (IRIMO), Scholte and De Geest (2010) and Raza et al. (2015).

SUPPLEMENTARY TABLE 2 | List of *Polycarpaea* (Caryophyllaceeae) species known from SW Asia which we could not test their photosynthetic type because of absence of material.

SUPPLEMENTARY TABLE 3 | The List of *Polycarpaea* (Caryophyllaeceae), *Cyperus* and *Fimbristylis* (Cyperceae) species with C_3 type carbon isotope values.

SUPPLEMENTARY TABLE 4 | List of C_4 plants with economic importance in SW Asia including non-native cultivated or introduced species.

SUPPLEMENTARY DATA SHEET 1 | References of floristic database.

REFERENCES

Abdel Khalik, K., El-Sheikh, M., and El-Aidarous, A. (2013). Floristic diversity and vegetation analysis of Wadi Al-Noman, Mecca, Saudi Arabia. *Turkish J. Bot.* 37, 894–907. doi: 10.3906/bot-1209-56

Akhani, H., and Deil, U. (2012). First observations of the flora and vegetation of three Islands in the NW Persian Gulf (Iran). *Phyton* 52, 73–99.

Akhani, H., and Ghasemkhani, M. (2007). Diversity of photosynthetic organs in Chenopodiaceae from Golestan National Park (NE Iran) based on carbon isotope composition and anatomy of leaves and cotyledons. *Nova Hedwigia* 131, 265–277.

Akhani, H., and Rudov, A. (2018). Biodiversity Conservation or Biotechnology: What is Iran's priority today? *Strategic Res. J. Agric. Sci. Natural Resour.* 3 (2), 181–194. doi: 10.1007/s11120-008-9376-0

Akhani, H., Trimborn, P., and Ziegler, H. (1997). Photosynthetic pathways in Chenopodiaceae from Africa, Asia and Europe with their ecological, phytogeographical and taxonomical importance. *Plant Syst. Evol.* 206, 187–221. doi: 10.1007/BF00987948

Akhani, H., Ghobadnejhad, M., and Hashemi, S. (2003). Ecology, Biogeography and Pollen Morphology of *Bienertia cycloptera* Bunge ex Boiss. (Chenopodiaceae), an Enigmatic C_4 Plant without Kranz Anatomy. *Plant Biol.* 5, 167–178. doi: 10.1055/s-2003-40724

Akhani, H., Barroca, J., Koteyeva, N., Voznesenskaya, E., Franceschi, V., Edwards, G., et al. (2005). *Bienertia sinuspersici* (Chenopodiaceae): A New Species from Southwest Asia and Discovery of a Third Terrestrial C4 Plant Without Kranz Anatomy. *Syst. Bot.* 30, 290–301. doi: 10.1600/0363644054223684

Akhani, H., Edwards, G., and Roalson, E. (2007). Diversification of the Old World Salsoleae s.l. (Chenopodiaceae): Molecular Phylogenetic Analysis of Nuclear and Chloroplast Data Sets and a Revised Classification. *Int. J. Plant Sci.* 168, 931–956. doi: 10.1086/518263

Akhani, H., Lara, M. V., Ghasemkhani, M., Ziegler, H., and Edwards, G. E. (2009). Does *Bienertia cycloptera* with the single-cell system of C-4 photosynthesis exhibit a seasonal pattern of delta C-13 values in nature similar to co-existing C-4 Chenopodiaceae having the dual-cell (Kranz) system? *Photosynth. Res.* 99, 23–36. doi: 10.1007/s11120-008-9376-0

Akhani, H., Chatrenoor, T., Dehghani, M., Khoshravesh, R., Mahdavi, P., and Matinzadeh, Z. (2012). A new species of *Bienertia* (Chenopodiaceae) from Iranian salt deserts: A third species of the genus and discovery of a fourth terrestrial C_4 plant without Kranz anatomy. *Plant Biosyst.* 146, 1–10. doi: 10.1080/11263504.2012.662921

Akhani, H., Greuter, W., and Roalson, E. H. (2014). Notes on the typification and nomenclature of Salsola and *Kali* (Chenopodiaceae). *Taxon.* 63, 647–650. doi: 10.12705/633.1

Akhani, H., and Ziegler, H. (2002). Photosynthetic pathways and habitats of grasses in Golestan National Park (NE Iran), with an emphasis on the C 4 -grass dominated rock communities. *Phytocoenologia* 32, 455–501. doi: 10.1127/0340-269X/2002/0032-0455

Akhani, H. (2004). Halophytic vegetation of Iran: Towards a syntaxonomical classification. *Annali di Botanica* 4, 65–82.

Akhani, H. (2006). "Biodiversity of halophytic and sabkha ecosystems in Iran," in *Tasks for Vegetation Science. Sabkha Ecosystems Volume II* (West and Central Asia: Springer), 71–88. doi: 10.1007/978-1-4020-5072-5_6

Akhani, H. (2015a). *Plants and Vegetation of North-West Persian Gulf: The coasts and Islands of Khore Musa* (Mahshahr and Adjacent Areas: University of Tehran Press).

Akhani, H. (2015b). Iran's environment under siege. *Science* 350, 392–392. doi: 10.1126/science.350.6259.392-a

Akhani, H. (2016). "Plant diversity of saline wetlands and salt marshes of Iran," in *U.S.-Iran Symposium on Wetlands*. Eds. K. Lansey, H. Vafai and D. Quanrud (Irvine, California: University of Arizone), 38–45.

Al-Hemaid, F., and Thomas, J. (1996). Review of the genus *Tribulus* L. @ in Saudi Arabia. *Arab Gulf J. Sci. Res.* 14, 415–443.

Al-Oudat, M., and Qadir, M. (2011). *The Halophytic Flora of Syria* (Aleppo, Syria: International Center for Agricultural Research in the Dry Areas).

Ananda Rao, T., and Meher-Homji, V. M. (1985). Strand plant communities of the Indian sub-continent. *Proc. Indian Acad. Sci. (Plant Sci.)* 94, 505–523. doi: 10.1007/BF03053163

APG III (The Angiosperm Phylogeny Group) (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121. doi: 10.1111/j.1095-8339.2009.00996.x

APG IV (The Angiosperm Phylogeny Group) (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181, 1–20. doi: 10.1111/boj.12385

Applequist, W. L. (2016). Report of the Nomenclature Committee for Vascular Plants: 68. *Taxon* 65, 1153–1165. doi: 10.12705/655.16

Arafat, S., Gaafar, A., Basuny, A., and Nassef, S. (2009). Chufa Tubers (*Cyperus esculentus* L.): As a New Source of Food. *World Appl. Sci. J.* 7, 151–156.

Assefa, K., Chanyalew, S., and Tadele, Z. (2017). "Tef, Eragrostis tef (Zucc.) Trotter," in *Millets and Sorghum: Biology and Genetic Improvement*. Ed. J. V. Patil (Chennai: John Wiley and Sons Ltd), 226–256. doi: 10.1002/9781119130765.ch9

Atala, A. (2012). A new ingredient: The introduction of priprioca in gastronomy. *Int. J. Gastronomy Food Sci.* 1, 61–63. doi: 10.1016/j.ijgfs.2011.11.001

Attorre, F., Francesconi, F., Taleb, N., Scholte, P., Saed, A., Alfo, M., et al. (2007). Will dragonblood survive the next period of climate change? Current and future potential distribution of *Dracaena cinnabari* (Socotra, Yemen). *Biol. Conserv.* 138, 430–439. doi: 10.1016/j.biocon.2007.05.009

Batanouny, K., Stichler, W., and Ziegler, H. (1988). Photosynthetic pathways, distribution, and ecological characteristics of grass species in egypt. *Oecologia* 75, 539–548. doi: 10.1007/BF00776418

Bertea, M., and Maffei, M. (2009). Botany, including anatomy, physiology, biochemistry and molecular biology. *Essential Oil-Bearing Grasses: The genus Cymbopogon*,. Ed. A. Akhila (London: Taylor and Francis), 1–24. doi: 10.1201/9780849378584

Bond, W., Woodward, F., and Midgley, G. (2005). The global distribution of ecosystems in a world without fire. *New Phytol.* 165, 525–538. doi: 10.1111/j.1469-8137.2004.01252.x

Boteva, H., Dintcheva, T., Masheva, S., Yankova, V., and Markova, D. (2014). Opportunities for Growing *Tribulus terrestris* L. as Semi-Culture. *Biotechnol. Biotechnol. Equip.* 25, 2388–2390. doi: 10.5504/BBEQ.2011.0029

Bräutigam, A., and Gowik, U. (2016). Photorespiration connects C_3 and C_4 photosynthesis. *J. Exp. Bot.* 67, 2953–2962. doi: 10.1093/jxb/erw056

Bräutigam, A., Schliesky, S., Külahoglu, C., Osborne, C., and Weber, A. (2014). Towards an integrative model of C4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C4 species. *J. Exp. Bot.* 65, 3579–3593. doi: 10.1093/jxb/eru100

Breckle, S., and Rafiqpoor, M. (2010). "Field Guide Afghanistan - Flora and Vegetation. Deutsche Nationalbibliothek," (Bonn: Scientia Bonnensis).

Breckle, S., Hedge, I., and Rafiqpoor, M. (2013). "Vascular Plants of Afghanistan: An Augmented Checklist," in *Scientia Bonnensis, Bonn, Manama, New York, Florianapólis*.

Brown, G., and Mies, B. (2012). The vegetation ecology of Socotra. (Dordrecht: Springer, Netherland). doi: 10.1007/978-94-007-4141-6

Brullo, C., Brullo, S., Gaskin, J., Galdo, G., Hrusa, G., and Salmeri, C. (2015). A new species of *Kali* (Salsoloideae, Chenopodiaceae) from Sicily, supported by molecular analysis. *Phytotaxa* 201, 256–277. doi: 10.11646/phytotaxa.201.4.2

Byrne, M., Pendergrass, A., Rapp, A., and Wodzicki, K. (2018). Response of the Intertropical Convergence Zone to Climate Change: Location, Width, and Strength. *Curr. Climate Change Rep.* 4, 355–370. doi: 10.1007/s40641-018-0110-5

CABI *Invasive Species Compendium* (Wallingford, UK: CAB International). Available at: https://www.cabi.org/ISC/ (Accessed Accessed: 14.01.2020).

Cabido, M., Pons, E., Cantero, J., Lewis, J., and Anton, A. (2008). Photosynthetic pathway variation among C_4 grasses along a precipitation gradient in Argentina. *J. Biogeogr.* 35, 131–140. doi: 10.1111/j.1365-2699.2007.01760.x

Centofanti, T., and Bañuelos, G. (2015). Evaluation of the halophyte *Salsola soda* as an alternative crop for saline soils high in selenium and boron. *J. Environ. Manage.* 157, 96–102. doi: 10.1016/j.jenvman.2015.04.005

Chandran, M. (2015). Grassland vegetation of India: An update. Ecology and Management of Grassland Ecology in India -. *Envis Bull.* 17. Wildlife Institute of India, Dehradun.

Chaudhri, I. I., Hussain, S. I., and Shah, B. H. (1966). The vegetation of riverian tract of the Indus around Ghulam Mohammed Barrage, West Pakistan. *Vegetatio* 13, 319–338. doi: 10.1007/BF00242772

Cheikhyoussef, A., Mapaure, I., and Shapi, M. (2011). The use of some Indigenous Plants for Medicinal and other Purposes by Local Communities in Namibia with Emphasis on Oshikoto Region: A Review. *Res. J. Med. Plant* 5, 406–419. doi: 10.3923/rjmp.2011.406.419

Christin, P., Besnard, G., Samaritani, E., Duvall, M., Hodkinson, T., Savolainen, V., et al. (2008). Oligocene CO_2 Decline Promoted C_4 Photosynthesis in Grasses. *Curr. Biol.* 18, 37–43. doi: 10.1016/j.cub.2007.11.058

Chuong, S. D. X., Franceschi, V. R., and Edwards, G. E. (2006). The cytoskeleton maintains organelle partitioning required for single-cell C_4 photosynthesis in Chenopodiaceae species. *Plant Cell* 18, 2207–2223. doi: 10.1105/tpc.105.036186

Danin, A., and Orshan, G. (1999). Vegetation of Israel I. Desert and coastal vegetation. (Leiden: Backuys Publishers).

Danin, A. (2007). "Edible Leaves and Fruits," in *Plant Stories*. Available at: https://flora.org.il/en/books/plant-stories-2/.

Dasti, A., and Agnew, A. D. Q. (1994). The vegetation of Cholistan and Thal deserts, Pakistan. *J. Arid Environ.* 27, 193–208. doi: 10.1006/jare.1994.1058

Davis, P. (1966-2001). *Flora of Turkey and the East Aegean Islands* (Edinburgh: Edinburgh University Press).

Deil, U., and Müller-Hohenstein, K. (1985). Beiträge zur Vegetation des Jemen I Pflanzengesellschaften und Ökotopgefüge der Gebirgstihamah am Beispiel des Beckens von At Tur (J.A.R.). *Phytocoenologia* 13, 1–102. doi: 10.1127/phyto/13/1985/1

Diez-Garretas, B., and Asensi, A. (1999). Syntaxonomic analysis of the andropogon-rich grasslands (Hyparrhenietalia hirtae) in the western mediterranean region. *Folia Geobot.* 34, 307–320. doi: 10.1007/BF02912817

Djamali, M., Akhani, H., Andrieu-Ponel, V., Bracconnot, P., Brewer, S., de Beaulieu, J.-L., et al. (2010). Indian Summer Monsoon variations could have affected the early Holocene woodland expansion in the Near East. *Holocene* 20, 813–820. doi: 10.1177/0959683610362813

Djamali, M., Akhani, H., Khoshravesh, R., Andrieu-Ponel, V., Ponel, P., and Brewer, S. (2011). Application of the Global Bioclimatic Classification to Iran: implications for understanding the modern vegetation and biogeography. *Int. J. Mediterranean Ecol.* 37, 91–114. doi: 10.3406/ecmed.2011.1350

Djamali, M., Brewer, S., Breckle, S., and Jackson, S. (2012). Climatic determinism in phytogeographic regionalization: A test from the Irano-Turanian region, SW and Central Asia. *Flora - Morphol. Distribution Funct. Ecol. Plant* 207, 237–249. doi: 10.1016/j.flora.2012.01.009

Doostmohammadi, M., Malekmohammadi, M., Djamali, M., and Akhani, H. (2020). Is *Pteropyrum* a pathway to C_4 evolution in Polygonaceae? An integrative approach to the taxonomy and anatomy of *Pteropyrum* (C_3), an immediate relative to *Calligonum* (C_4). *Bot. J. Linn. Soc.* 192, 369–400. doi: 10.1093/botlinnean/boz079

Duke, J. (1993). *CRC handbook of alternative cash crops* (CRC Press).

Dzyubenko, N., and Soskov, Y. (2014). *Geneticheskie resursy Kokhii prostyortoy - Kochia prostrata (L.) Schrad* (Sankt-Petersburg: SPB:VIR).

Edwards, G., Franceschi, V., and Voznesenskaya, E. (2004). Single-Cell C_4 Photosynthesis Versus the Dual-Cell (kranz) Paradigm. *Annu. Rev. Plant Biol.* 55, 173–196. doi: 10.1146/annurev.arplant.55.031903.141725

Ehleringer, J., Lin, Z., Field, C., Sun, G., and Kuo, C. (1987). Leaf carbon isotope ratios of plants from a subtropical monsoon forest. *Oecologia* 72, 109–114. doi: 10.1007/BF00385053

El-Ghanim, W. M., Hassanb, L. M., Galal, L. M., and Badr, A. (2010). Floristic composition and vegetation analysis in Hail region north of central Saudi Arabia. *Saudi J. Biol. Sci.* 17, 119–128. doi: 10.1016/j.sjbs.2010.02.004

El-Kamali, H., Hamza, M., and El-Amir, M. (2005). Antibacterial activity of the essential oil from Cymbopogon nervatus inflorescence. *Fitoterapia* 76, 446–449. doi: 10.1016/j.fitote.2005.03.001

Elmore, C., and Paul, R. (1983). Composite List of C_4 Weeds. *Weed Sci.* 31, 686–692. doi: 10.1017/S0043174500070193

El-Sheikh, M. A. (2013). Weed vegetation ecology of arable land in Salalah, Southern Oman. *Saudi J. Biol. Sci.* 20, 291–304. doi: 10.1016/j.sjbs.2013.03.001

Evans, J. (2009). 21st century climate change in the Middle East. *Clim. Change* 92, 417–432. doi: 10.1007/s10584-008-9438-5

Fayvush, G. M., and Aleksanyan, A. S. (2016). Habitats of Armenia. *Bot. Institute Yerevan*. doi: 10.13140/RG.2.1.1695.9601

Fern, K. (2016). *Useful Temperate Plants Database*. Available at: http://temperate.theferns.info/ (Accessed 22.06.2019).

Fet, V., and Atamuradov, K. (1994). *Biogeography and Ecology of Turkmenistan* (Dordrecht: Springer). doi: 10.1007/978-94-011-1116-4

Freitag, H., and Stichler, W. (2000). A Remarkable New Leaf Type With Unusual Photosynthetic Tissue in a Central Asiatic Genus of Chenopodiaceae. *Plant Biol.* 2, 154–160. doi: 10.1055/s-2000-9462

GBC *(Worldwide Bioclimatic Classification System)*. Available at: http://www.globalbioclimatics.org/ (Accessed 13.06.2019).

Ghasemkhani, M., Akhani, H., Sahebi, J., and Scholz, H. (2008). The genera Aristida and Stipagrostis (Poaceae) in Iran. *Willdenowia* 38, 135–148. doi: 10.3372/wi.38.38108

Ghazanfar, S. A., and Fisher, M. (1998). *Vegetation of the Arabian Peninsula* (Dordrecht: Springer). doi: 10.1007/978-94-017-3637-4

Ghorbanalizadeh, A., Akhani, H., and Bergmeier, E. (2020). Vegetation patterns of a rapidly drying up salt lake ecosystem: Lake Urmia, NW Iran. *Phytocenologia* 50, 1–46. doi: 10.1127/phyto/2019/0338

Gintzburger, G., Toderich, K., Mardonov, B., and Mahmudov, M. (2003). Rangelands of the Arid and Semi-Arid zones in Uzbekistan. *ClRAD ICARDA*.

Gomashe, S. (2017). "Proso Millet, Panicum miliaceum (L.): Genetic Improvement and Research Needs," in *Millets and Sorghum: Biology and Genetic Improvement*. Ed. J. V. Patil (Chennai: John Wiley and Sons Ltd), 150–169. doi: 10.1002/9781119130765.ch5

Gonnella, M., Charfeddine, M., Conversa, G., and Santamaria, P. (2010). Purslane: A Review of its Potential for Health and Agricultural Aspects. *Eur. J. Plant Sci. Biotechnol.* 4, 131–136.

Guest, E., and Ghazanfar, S. (1966-2013). "Flora of Iraq," vols. 1-9, (Kew: Royal Botanic Gardens, Bentham-Moxon Trust).

Gutermann, W. (2011). Notulae nomenclaturales 41-45. *Phyton-Annales Rei Botanicae* 51, 95–102.

Haghshenas-Haghighi, M., and Motagh, M. (2019). Ground surface response to continuous compaction of aquifer system in Tehran, Iran: Results from a long-term multi-sensor InSAR analysis. *Remote Sens. Environ.* 221, 534–550. doi: 10.1016/j.rse.2018.11.003

Hakobyan, J. (2011). The genus *Salsola* sensu lato (Chenopodiaceae) in southern Transcaucasia. *Takhtajania* 1, 124–132.

Han, L., Nose, R., Li, W., Gong, X., Zheng, Y., Yoshikawa, M., et al. (2006). Reduction of fat storage in mice fed a high-fat diet long term by treatment with saponins prepared from Kochia scoparia fruit. *Phytother. Res.* 20, 877–882. doi: 10.1002/ptr.1981

Hanelt, P. (2001). *Mansfeld's Encyclopedia of Agricultural and Horticultural Crops*. (Springer Verlag Berlin Heidelberg).

Hariprasanna, K. (2017). "Foxtail Millet, Setaria italica (L.) P. Beauv," in *Millets and Sorghum - Wiley Online Library. Millets and Sorghum: Biology and Genetic Improvement*. Ed. J. V. Patil (Chennai: John Wiley and Sons Ltd), 112–149. doi: 10.1002/9781119130765.ch4

Hatch, M. D. (1999). "C4 photosynthesis: A historical overview," in *C4 Plant Biology*. Eds. R. Sage and R. Monson (San Diego, California: Academic Press), 17–46.

Hattersley, P. (1983). The distribution of C_3 and C_4 grasses in Australia in relation to climate. *Oecologia* 57, 113–128. doi: 10.1007/BF00379569

Hernandez-Ledesma, P., Berendsohn, W., Borsch, T., Von Mering, S., Akhani, H., Arias, S., et al. (2015). A taxonomic backbone for the global synthesis of species diversity in the angiosperm order Caryophyllales. *Willdenowia* 45, 281–384. doi: 10.3372/wi.45.45301

Hodkinson, T. R., Chase, M. W., Lledo, D. M., Salamin, N., and Renvoize, S. A. (2002). Phylogenetics of *Miscanthus, Saccharum* and related genera (Saccharinae, Andropogoneae, Poaceae) based on DNA sequences from ITS nuclear ribosomal DNA and plastid trnL intron and trnL-F intergenic spacers. *J. Plant Res.* 115, 381–392. doi: 10.1007/s10265-002-0049-3

Hoetzel, S., Dupont, L., Schefuß, E., Rommerskirchen, F., and Wefer, G. (2013). The role of fire in Miocene to Pliocene C_4 grassland and ecosystem evolution. *Nat. Geosci.* 6, 1027–1030. doi: 10.1038/ngeo1984

Ingram, A. L., and Doyle, J. J. (2003). The origin and evolution of *Eragrostis tef* (Poaceae) and related polyploids: evidence from nuclear waxy and plastid rps16. *Am. J. Bot.* 90, 116–122. doi: 10.3732/ajb.90.1.116

IPNI (2019). *International Plant Names Index* (The Royal Botanic Gardens, Kew, Harvard University Herbaria & Libraries and Australian National Botanic Gardens). Available at: http://www.ipni.org.

IRIMO *Data Base of Iranian Meteorological Organization*. Available at: http://data.irimo.ir (Accessed 12.12.2018).

James, G. (2014). "An Introduction to Sugarcane," in *Sugarcane* (Oxford: Blackwell Publishing Ltd), 1–19.

Jana, K., and Puste, A. (2014). Madur Kathi – An Important Economic Non-food Crop of West Bengal. *Asian Agri-History* 18, 145–151.

Kadereit, G., and Freitag, H. (2011). Molecular phylogeny of Camphorosmeae (Camphorosmoideae, Chenopodiaceae): Implications for biogeography, evolution of C4-photosynthesis and taxonomy. *Taxon* 60, 51–78. doi: 10.1002/tax.601006

Kadereit, G., Borsch, T., Weising, K., and Freitag, H. (2003). Phylogeny of Amaranthaceae and Chenopodiaceae and the Evolution of C_4 Photosynthesis. *Int. J. Plant Sci.* 164, 959–986. doi: 10.1086/378649

Kadereit, G., Ackerly, D., and Pirie, M. (2012). A broader model for C_4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae s.s.). *Proc. R. Soc. B.: Biol. Sci.* 279, 3304–3311. doi: 10.1098/rspb.2012.0440

Kapralov, M. V., Akhani, H., Voznesenskaya, E. V., Edwards, G., Franceschi, V., and Roalson, E. H. (2006). Phylogenetic relationships in the Salicornioideae/Suaedoideae/Salsoloideae s.l. (Chenopodiaceae) Clade and a clarification of the phylogenetic position of *Bienertia* and *Alexandra* using multiple DNA sequence datasets. *Syst. Bot.* 31, 571–585. doi: 10.1600/036364406778388674

Keeley, J. E., and Rundel, P. (2005). Fire and the Miocene expansion of C_4 grasslands. *Ecol. Lett.* 8, 683–690. doi: 10.1111/j.1461-0248.2005.00767.x

Kellogg, E. A. (2015). *The Families and Genera of Vascular Plants - Flowering Plants* (Monocots. Poaceae: Springer, Cham). doi: 10.1007/978-3-319-15332-2

Khyade, V., Pawar, S., and Sarwade, J. (2018). Novel Sacrificial Medicinal Repositories: Halfa grass, *Desmostachya bipinnata* (L.) and Cogon grass, *Imperata cylindrica* (L.). *World Sci. News* 100, 35–50.

Kool, A. (2012). *Desert Plants and Deserted Islands: Systematics and Ethnobotany in Caryophyllaceae* (Uppsala: Acta Universitatis Upsaliensis).

Kripa, K., and Vijayalakshmi, S. (2016). Therapeutic uses of plants of genus Blepharis - a systematic review. *Int. J. Pharma Bio Sci.* 7, 236–243. doi: 10.22376/ijpbs.2016.7.4.b236-243

Kürschner, H. (1986). *Contributions to the Vegetation of Southwest Asia. Tübinger Atlas des Vorderen Orients (TAVO) 24* (Tübingen: Reichert Verlag).

Lauterbach, M., Billakurthi, K., Kadereit, G., Ludwig, M., Westhoff, P., and Gowik, U. (2017). C_3 cotyledons are followed by C_4 leaves: intra-individual transcriptome analysis of *Salsola soda* (Chenopodiaceae). *J. Exp. Bot.* 68, 161–176. doi: 10.1093/jxb/erw343

Léonard, J. (1988). Contribution a l'etude de la Flore et de la Vegetation des Deserts d'Iran (Dasht-E-Kavir, Dasht-E-Lut, Jaz Murian). Fasc. 8: Etude des aires de distribution, Les phytochories, Les chorotypes. (Meise: Jardin botanique Natl. Belgique).

Li, Y., Feng, Y., Wang, X., Liu, B., and Lv, G. (2014). Failure of DNA barcoding in discriminating *Calligonum* species. *Nordic J. Bot.* 32, 511–517. doi: 10.1111/njb.00423

Li, Y., Ma, X., Zhao, J., Xu, J., Shi, J., Zhu, X., et al. (2015). Developmental Genetic Mechanisms of C4 Syndrome Based on Transcriptome Analysis of C_3 Cotyledons and C_4 Assimilating Shoots in *Haloxylon ammodendron*. *PloS One* 10, e0117175. doi: 10.1371/journal.pone.0117175

Madani, K. (2014). Water management in Iran: what is causing the looming crisis? *J. Environ. Stud. Sci.* 4, 315–328. doi: 10.1007/s13412-014-0182-z

Mahesh, A., Kumar, H., Ranganath, M., and Devkar, R. (2012). Detail Study on Boerhaavia Diffusa Plant for its Medicinal Importance-A Review. *Res. J. Pharm. Sci.* 1, 28–36.

Mantlana, K., Arneth, A., Veenendaal, E., Wohland, P., Wolski, P., Kolle, O., et al. (2008). Photosynthetic properties of C_4 plants growing in an African savanna/wetland mosaic. *J. Exp. Bot.* 59, 3941–3952. doi: 10.1093/jxb/ern237

Medina, E., Cram, W., Lee, H., Luettge, U., Popp, M., Smith, J., et al. (1989). Ecophysiology of xerophytic and halophytic vegetation of a coastal alluvial plain in northern Venezuela. *New Phytol.* 111, 233–243. doi: 10.1111/j.1469-8137.1989.tb00688.x

Miller, A., and Cope, T. (1996). *Flora of the Arabian Peninsula and Socotra* (Edinburgh: Edinburgh University Press).

Mosyakin, S. L., Rilke, S., and Freitag, H. (2014). Proposal to conserve the name *Salsola* (Chenopodiaceae s.str.; Amaranthaceae sensu APG) with a conserved type. *Taxon* 63, 1134–1135. doi: 10.12705/635.15

Mosyakin, S. L., Freitag, H., and Rilke, S. (2017). *Kali* versus *Salsola*: the instructive story of a questionable nomenclatural resurrection. *Israel J. Plant Sci.* 64, 18–30. doi: 10.1080/07929978.2016.1256135

Motagh, M., Walter, T. R., Sharifi, M. A., Fielding, E., Schenk, A., Anderssohn, J., et al. (2008). Land subsidence in Iran caused by widespread water reservoir overexploitation. *Geophys. Res. Lett.* 35. doi: 10.1029/2008GL033814

Motagh, M., Shamshiri, R., Haghshenas-Haghighi, M., Wetzel, H. U., Akbari, B., Nahavandchi, H., et al. (2017). Quantifying groundwater exploitation induced subsidence in the Rafsanjan plain, southeastern Iran, using InSAR time-series and in situ measurements. *Eng. Geol.* 218, 134–151. doi: 10.1016/j.enggeo.2017.01.011

Mucina, L., Bültmann, H., Dierßen, K., Theurillat, J. P., Raus, T., Čarni, A., et al. (2016). Vegetation of europe: Hierarchical floristic classification system of vascular plant, bryophyte, lichen, and algal communities. *App. Veg. Sci.* 19, 3–264. doi: 10.1111/avsc.12257

Mullet, J. (2017). High-biomass C_4 grasses-Filling the yield gap. *Plant Sci.: Int. J. Exp. Plant Biol.* 261, 10–17. doi: 10.1016/j.plantsci.2017.05.003

Nasir, E., and Ali, S. I. (1970-2003). *Edit. Flora of Pakistan* (Karachi: University of Karachi), 1–202.

Nechayeva, N. T. (1992). *Rastitelnost Turkmenistana* (*Ashgabat, Ylym*).

Nedelcheva, A., Dogan, Y., and Guarrera, P. (2007). Plants traditionally used to make brooms in several European countries. *J. Ethnobiol. Ethnomed.* 3, 20. doi: 10.1186/1746-4269-3-20

Nikitin, V., and Geldikhanov, A. (1988). *Opredelitel rastenij Turkmenistana* (Ashgabat: Academy of sciences of the Turkmen SSR).

Origin Pro, Version Number (2017). *OriginLab Corporation* (Northampton, MA, USA).

O'Leary, M. (1988). Carbon Isotopes in Photosynthesis. *BioScience* 38, 328–336. doi: 10.2307/1310735

Osborne, C., and Beerling, D. (2006). Nature's green revolution: the remarkable evolutionary rise of C_4 plants. *Philos. Trans. R. Soc. London. Ser. B. Biol. Sci.* 361, 173–194. doi: 10.1098/rstb.2005.1737

Osborne, C. (2010). "The Geologic History of C4 Plants. Advances in Photosynthesis and Respiration," in *C4 Photosynthesis and Related CO2 Concentrating Mechanisms* (Dordrecht: Springer), 339–357. doi: 10.1007/978-90-481-9407-0

Osmond, C., Ziegler, H., Stichler, W., and Trimborn, P. (1975). Carbon isotope discrimination in alpine succulent plants supposed to be capable of Crassulacean acid metabolism (CAM). *Oecologia* 28, 323–328. doi: 10.1007/BF00345423

Pahlevani, A. H., Liede-Schumann, S., and Akhani, H. (2020). Diversity, distribution, endemism and conservation status of Euphorbia (Euphorbiaceae) in SW Asia and adjacent countries. *Plant Syst. Evol.* 306, 80. doi: 10.1007/s00606-020-01705-4

Pal, J. S., and Eltahir, E. A. B. (2016). Future temperature in southwest Asia projected to exceed a threshold for human adaptability. *Nat. Climate Change* 6, 197–200. doi: 10.1038/nclimate2833

Pareek, A., and Kumar, A. (2013). Ethnobotanical and pharmaceutical uses of *Vetiveria zizanioides* (Linn) Nash: a medicinal plant of Rajasthan. *Int. J. Life Sci. Pharma Res.* 3, L12–l18.

Pascual, B., Maroto, J., Lopez-Galarza, S., and Sanbautista, A. (2000). Chufa (*Cyperus esculentus* L. var. *sativus* Boeck.): Unconventional crop. Studies related to applications and cultivation. *Economic Bot.* 54, 439–448. doi: 10.1007/BF02866543

Pérez-García, F., Akhani, H., Parsons, R., Silcock, J. L., Kurt, L., Özdeniz, E., et al. (2018). A first inventory of gypsum flora in the Palearctic and Australia. *Mediterranean Bot.* 39, 35–49. doi: 10.5209/MBOT.59428

POWO (2019). *"Plants of the World Online* (Facilitated by the Royal Botanic Gardens, Kew). Available at: http://www.plantsoftheworldonline.org/.

Prasad, C., Singh, D., Shkula, O., and Singh, U. (2014). *Cymbopogon jwarancusa* - An important medicinal plant: A review. *Pharma Innovation J.* 3, 13–19.

PROTA4U web database (2018). Available at: https://www.prota4u.org/database/ (Accessed 23.06.2019).

Pyankov, V., Black, C., Artyusheva, E., Voznesenskaya, E., Ku, M., and Edwards, G. (1999). Features of Photosynthesis in *Haloxylon* species of Chenopodiaceae that are Dominant Plants in Central Asian Deserts. *Plant Cell Physiol.* 40, 125–134. doi: 10.1093/oxfordjournals.pcp.a029519

Pyankov, V., Gunin, P., Tsoog, S., and Black, C. (2000a). C_4 plants in the vegetation of Mongolia: their natural occurrence and geographical distribution in relation to climate. *Oecologia* 123, 15–31. doi: 10.1007/s004420050985

Pyankov, V., Voznesenskaya, E., Kuz'min, A., Ku, M., Ganko, E., Franceschi, V., et al. (2000b). Occurrence of C_3 and C_4 photosynthesis in cotyledons and leaves of *Salsola* species (Chenopodiaceae). *Photosynth. Res.* 63, 69–84. doi: 10.1023/A:1006377708156

Pyankov, V., Ziegler, H., Kuz'min, A., and Edwards, G. (2001). Origin and evolution of C_4 photosynthesis in the tribe Salsoleae (Chenopodiaceae) based on anatomical and biochemical types in leaves and cotyledons. *Plant Syst. Evol.* 230, 43–74. doi: 10.1007/s006060170004

Pyankov, V., Ziegler, H., Akhani, H., Deigele, C., and Lüttge, U. (2010). European plants with C_4 photosynthesis: geographical and taxonomic distribution and relations to climate parameters. *Bot. J. Linn. Soc.* 163, 283–304. doi: 10.1111/j.1095-8339.2010.01062.x

Raghavendra, A., and Das, V. (1976). Distribution of the C_4 dicarboxylic acid pathway of photosynthesis in local Monocotyledonous plants and its taxonomic significance. *New Phytol.* 76, 301–305. doi: 10.1111/j.1469-8137.1976.tb01465.x

Rao, X., and Dixon, R. (2016). The Differences between NAD-ME and NADP-ME Subtypes of C_4 Photosynthesis: More than Decarboxylating Enzymes. *Front. Plant Sci.* 7, 1525. doi: 10.3389/fpls.2016.01525

Ravichandran, P., Thumilan, M., Benazir, J., and Manimekalai, V. (2005). Anatomy and Vascular Bundle Diversity in Mat sedges. *Phytomorphology* 55, 75–83.

Raza, M., Hussain, D., Rasul, G., Akbar, M., and Raza, G. (2015). Variations of surface temperature and precipitation in Gilgit-Baltistan (GB), Pakistan from 1955 to 2010. *J. Biodivers. Environ. Sci.* 6, 67–73.

Rechinger, K. H. (1963-2015). *Flora Iranica: Flora des iranischen Hochlandes und der umrahmenden Gebirge, , Fas. 1-181* (Graz, Austria: Akademische Druck - u. Verlagsanstalt).

Rejzek, M., Svatek, M., Sebesta, J., Adolt, R., Madera, P., and Matula, R. (2016). Loss of a single tree species will lead to an overall decline in plant diversity: Effect of *Dracaena cinnabari* Balf. f. on the vegetation of Socotra Island. *Biol. Conserv.* 196, 165–172. doi: 10.1016/j.biocon.2016.02.016

Rivas-Martinez, S., and Rivas-Saenz, S. (1996-2019). *Worldwide Bioclimatic Classification System Phytosociological Research Center*. Available at: https://www.globalbioclimatics.org/

Rodwell, M., and Hoskins, B. J. (1996). Monsoons and the dynamics of deserts. *Quartly J. R. Meteorol. Soc.* 122, 1385–1404. doi: 10.1002/qj.49712253408

Rukhlenko, I. A. (2001). The order Aeluropodetalia littoralis in the flood plain of Atrek River (South West Turkmenia). *Feddes Repertorium* 112, 107–125. doi: 10.1002/fedr.20011120116

Rundel, P., Esler, K., and Cowling, R. (1999). Ecological and phylogenetic patterns of carbon isotope discrimination in the winter-rainfall flora of the Richtersveld, South Africa. *Plant Ecol.* 142, 133–148. doi: 10.1023/A:1009878429455

Sage, R., and Sultmanis, S. (2016). Why Are There No C_4 Forests? *J. Plant Physiol.* 203, 55–68. doi: 10.1016/j.jplph.2016.06.009

Sage, R., Sage, T., and Kocacinar, F. (2012). Photorespiration and the evolution of C_4 photosynthesis. *Annu. Rev. Plant Biol.* 63, 19–47. doi: 10.1146/annurev-arplant-042811-105511

Sage, R. (2001). Environmental and Evolutionary Preconditions for the Origin and Diversification of the C_4 Photosynthetic Syndrome. *Plant Biol.* 3, 202–213. doi: 10.1055/s-2001-15206

Sage, R. (2004). The evolution of C4 photosynthesis. *New Phytol.* 161, 341–370. doi: 10.1111/j.1469-8137.2004.00974.x

Sage, R. (2011). The C_4 plant lineages of planet Earth. *J. Exp. Bot.* 62, 3155–3169. doi: 10.1093/jxb/err048

Sage, R. (2016). A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *J. Exp. Bot.* 67, 4039–4056. doi: 10.1093/jxb/erw156

Sahu, S., Rout, N., and Dhal, N. (2010). Ethnobotany of *Eulaliopsis binata* (Retz.) Hubbard- Poaceae, in Odisha, Eastern India: Cultivation Practice, Economics and Prospects. *J. Adv. Dev. Res.* 1, 155–160.

Salamon, I., Grulova, D., and De Feo, V. (2016). Comparison of two methods for field grow of puncture vine (*Tribulus terrestris* L.) in Slovakia. *Acta Agric. Scand. Sect. B. - Soil Plant Sci.* 66, 267–271. doi: 10.1080/09064710.2015.1093652

Sanjana Reddy, P. (2017a). "Sorghum, Sorghum bicolor (L.) Moench. Millets and Sorghum - Wiley Online Library," in *Millets and Sorghum: Biology and Genetic Improvement*. Ed. J. V. Patil (Chennai: John Wiley and Sons Ltd), 1–48. doi: 10.1002/9781119130765.ch1

Sanjana Reddy, P. (2017b). "Pearl Millet, *Pennisetum glaucum* (L.) R. Br," in *Millets and Sorghum: Biology and Genetic Improvement Editor*. Ed. J. V. Patil (Chennai: John Wiley and Sons Ltd), 49–86. doi: 10.1002/9781119130765.ch2

Sauer, J. (1967). The grain amaranths and their relatives: a revised taxonomic and geographic survey. *Ann. Missouri Bot. Garden* 54, 103–137. doi: 10.2307/2394998

Scholte, P., and De Geest, P. (2010). The climate of Socotra Island (Yemen): A first-time assessment of the timing of the monsoon wind reversal and its influence on precipitation and vegetation patterns. *J. Arid Environ.* 74, 1507–1515. doi: 10.1016/j.jaridenv.2010.05.017

Schuler, M. L., Mantegazza, O., and Weber, A. P. (2016). Engineering C_4 photosynthesis into C_3 chassis in the synthetic biology age. *Plant J.* 87, 51–65. doi: 10.1111/tpj.13155

Schulze, E., Ellis, R., Schulze, W., Trimborn, P., and Ziegler, H. (1996). Diversity, metabolic types and $\delta^{13}C$ carbon isotope ratios in the grass flora of Namibia in relation to growth form, precipitation and habitat conditions. *Oecologia* 106, 352–369. doi: 10.1007/BF00334563

Schüssler, C., Freitag, H., Koteyeva, N., Schmidt, D., Edwards, G., Voznesenskaya, E., et al. (2017). Molecular phylogeny and forms of photosynthesis in tribe Salsoleae (Chenopodiaceae). *J. Exp. Bot.* 68, 207–223. doi: 10.1093/jxb/erw432

Seetharam, A., and Riley, K. (1986). "Small Millets in Global Agriculture," in *Proceedings of the First International Small Millets Workshop*, October 29-November 2, 1986 (Bangalore, India: Oxford and IBH Publishing Company).

Serag, M. (2003). Ecology and biomass production of *Cyperus papyrus* L. on the Nile bank at Damietta, Egypt. *J. Mediterranean Ecol.* 4, 15–24.

Shi, W., Wen, J., and Pan, B. (2016). A comparison of ITS sequence data and morphology for *Calligonum pumilum* and *C. mongolicum* (Polygonaceae) and its taxonomic implications. *Phytotaxa* 261, 157–167. doi: 10.11646/phytotaxa.261.2.5

Shioya, M., Myoga, A., Kitagawa, A., Tokunaga, Y., Hayashi, H., Kogo, Y., et al. (2019). Analysis of deflection and dynamic plant characteristics of *Cyperus malaccensis* Lam. *Plant Prod. Sci.* 22, 1–8. doi: 10.1080/1343943X.2019.1588075

Shmida, A., and Aronson, J. A. (1986). Sudanian Elements in the Flora of Israel. *Ann. Missouri Bot. Garden* 73, 1–28. doi: 10.2307/2399138

Simpson, D., and Inglis, C. (2001). Cyperaceae of Economic, Ethnobotanical and Horticultural Importance: A Checklist. *Kew Bull.* 56, 257–360. doi: 10.2307/4110962

Skinner, R. H., Hanson, J. D., Hutchinson, G. L., and Schuman, G. E. (2002). Response of C_3 and C_4 grasses to supplemental summer precipitation. *J. Range Manage.* 55, 517–522. doi: 10.2458/azu_jrm_v55i5_skinner

Snead, R. E., and Tasnif, M. (1966). Vegetation Types in the Las Bela Region of West Pakistan. *Ecology* 47, 494–499. doi: 10.2307/1932993

Soreng, R. J., Peterson, P. M., Romaschenko, K., Davidse, G., Zuloaga, F. O., Judziewicz, E. J., et al. (2015). A worldwide phylogenetic classification of the Poaceae (Gramineae). *J. Syst. Evol.* 53, 117–137. doi: 10.1111/jse.12262

Soreng, R. J., Peterson, P. M., Romaschenko, K., Davidse, G., Teisher, J. K., Clark, L. G., et al. (2017). A worldwide phylogenetic classification of the Poaceae (Gramineae) II: An update and a comparison of two 2015 classifications. *J. Syst. Evol.* 55, 259–290. doi: 10.1111/jse.12262

Soskov, Y. (2011). *Rod Calligonum L. — zhuzgun (sistematika, geografiya, evolyutsiya, introduktsiya) [Genus Calligonum L. — zhuzgun (taxonomy, geography, evolution, introduction)* (Novosibirsk: Siberian Sci. Agric. Library).

Staller, J. (2010). *Maize Cobs and Cultures: History of Zea mays L* (Berlin Heidelberg: Springer-Verlag), 262. doi: 10.1007/978-3-642-04506-6

Stock, W., Chuba, D., and Verboom, G. (2004). Distribution of South African C_3 and C_4 species of Cyperaceae in relation to climate and phylogeny. *Austral Ecol.* 29, 313–319. doi: 10.1111/j.1442-9993.2004.01368.x

Stowe, L., and Teeri, J. (1978). The Geographic Distribution of C_4 Species of the Eudicotyledonae in Relation to Climate. *Am. Nat.* 112, 609–623. doi: 10.1086/283301

Sukhorukov, A. P., Liu, P. L., and Kushunina, M. (2019). Taxonomic revision of Chenopodiaceae in Himalaya and Tibet. *PhytoKeys* 116, 1–141. doi: 10.3897/phytokeys.116.27301

Suttie, J. M., and Reynolds, S. G. (2003). Transhumant Grazing Systems in Temperate Asia. *FAO Plant Prod. Prot. Ser.* 31.

Takhtajian, A. (1992). *Floristic Regions of the World* (Berkeley: University of California Press).

Tavakkoli, S., Kazempour Osaloo, S., and Maassoumi, A. (2010). The Phylogeny of *Calligonum* and *Pteropyrum* (Polygonaceae) Based on Nuclear Ribosomal DNA ITS and Chloroplast trnL-F Sequences. *Iranian J. Biotechnol.* 8, 7–15.

Teeri, J., and Stowe, L. (1976). Climatic patterns and the distribution of C_4 grasses in North America. *Oecologia* 23, 1–12. doi: 10.1007/BF00351210

Tieszen, L., Senyimba, M., Imbamba, S., and Troughton, J. (1979). The distribution of C_3 and C_4 grasses and carbon isotope discrimination along an altitudinal and moisture gradient in Kenya. *Oecologia* 37, 337–350. doi: 10.1007/BF00347910

Toderich, K., Black, C. C., Juylova, E., Kozan, O., Mukimov, T., and Matsuo, N. (2007). *C3/C4 plants in the vegetation of Central Asia, geographical distribution and environmental adaptation in relation to climate. Climate Change and Terrestrial Carbon Sequestration in Central Asia*. (London: CRC Press), 33–63. doi: 10.1201/9780203932698.ch3

Tubeileh, A., Rennie, T., and Goss, M. (2016). A review on biomass production from C_4 grasses: yield and quality for end-use. *Curr. Opin. Plant Biol.* 31, 172–180. doi: 10.1016/j.pbi.2016.05.001

Turland, N. J., Wiersema, J. H., Barrie, F. R., Greuter, W., Hawksworth, D. L., Herendeen, P. S., et al. (2018). *(Eds) International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017* (Glashütten: Koeltz Botanical Books (Regnum Vegetabile 159).

Uddin, M., Juraimi, A., Hossain, M., Nahar, M., Ali, M., and Rahman, M. (2014). Purslane Weed (Portulaca oleracea): A Prospective Plant Source of Nutrition, Omega-3 Fatty Acid, and Antioxidant Attributes. *Sci. World J.* doi: 10.1155/2014/951019

Vaezi, A., Ghazban, F., Tavakoli, V., Routh, J., Naderi Beni, A., Bianchi, T. S., et al. (2019). A late Pleistocene-Holocene multi-proxy record of climate variability in the Jazmurian playa, southeastern Iran. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 514, 754–767. doi: 10.1016/j.palaeo.2018.09.026

Varghese, M., Yadav, S., and Thomas, J. (2006). Taxonomic Status of Some of the *Tribulus* Species in the Indian Subcontinent. *Saudi J. Biol. Sci.* 3, 7–12.

Von Caemmerer, S., Ghannoum, O., Pengelly, J. J. L., and Cousins, A. B. (2014). Carbon isotope discrimination as a tool to explore C_4 photosynthesis. *J. Exp. Bot.* 65, 3459–3470. doi: 10.1093/jxb/eru127

Voznesenskaya, E., Franceschi, V., Kiirats, O., Freitag, H., and Edwards, G. (2001). Kranz anatomy is not essential for terrestrial C_4 plant photosynthesis. *Nature* 414, 543–546. doi: 10.1038/35107073

Voznesenskaya, E., Franceschi, V., Kiirats, O., Artyusheva, E., Freitag, H., and Edwards, G. (2002). Proof of C_4 photosynthesis without Kranz anatomy in *Bienertia cycloptera* (Chenopodiaceae). *Plant J.: For Cell Mol. Biol.* 31, 649–662. doi: 10.1046/j.1365-313X.2002.01385.x

Voznesenskaya, E., Koteyeva, N., Chuong, S., Akhani, H., Edwards, G., and Franceschi, V. (2005). Differentiation of cellular and biochemical features of the single-cell C_4 syndrome during leaf development in *Bienertia cycloptera* (Chenopodiaceae). *Am. J. Bot.* 92, 1784–1795. doi: 10.3732/ajb.92.11.1784

Voznesenskaya, E. V., Akhani, H., Koteyeva, N. K., Chuong, S. D., Roalson, E. H., Kiirats, O., et al. (2008). Structural, biochemical, and physiological characterization of photosynthesis in two C_4 subspecies of *Tecticornia indica* and the C_3 species *Tecticornia pergranulata* (Chenopodiaceae). *J. Exp. Bot.* 59, 1715–1734. doi: 10.1093/jxb/ern028

Voznesenskaya, E. V., Koteyeva, N. K., Akhani, H., Roalson, E. H., and Edwards, G. E. (2013). Structural and physiological analyses in Salsoleae (Chenopodiaceae) indicate multiple transitions among C_3, intermediate, and C_4 photosynthesis. *J. Exp. Bot.* 64, 3583–3604. doi: 10.1093/jxb/ert191

Waha, K., Krummenauer, L., Adams, S., Aich, V., Baarsch, F., Coumou, D., et al. (2017). Climate change impacts in the Middle East and Northern Africa (MENA) region and their implications for vulnerable population groups. *Reg. Environ. Change* 17, 1623–1638. doi: 10.1007/s10113-017-1144-2

Walker, D., Lutts, S., Sánchez-García, M., and Correal, E. (2014). *Atriplex halimus* L.: Its biology and uses. *J. Arid Environ.* 100–101, 111–121. doi: 10.1016/j.jaridenv.2013.09.004

Walker, J., Yang, Y., Feng, T., et al. (2018). From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. *Am. J. Bot.* 105, 446–462. doi: 10.1002/ajb2.1069

Waller, S., and Lewis, J. (1979). Occurrence of C_3 and C_4 Photosynthetic Pathways in North American Grasses. *J. Range Manage.* 32, 12–28. doi: 10.2307/3897378

Wang, R., and Ma, L. (2016). Climate-driven C_4 plant distributions in China: divergence in C_4 taxa. *Sci. Rep.* 6, 27977. doi: 10.1038/srep27977

Wang, Y., Bräutigam, A., Weber, A. P. M., and Zhu, X. (2014). Three distinct biochemical subtypes of C4 photosynthesis? A modelling analysis. *J. Exp. Bot.* 65, 3567–3578. doi: 10.1093/jxb/eru058

Welk, E. (2015). *Phytogeography of the Nepalese flora and its floristic links to neighbouring regions. Nepal: An introduction to the natural history, ecology and human environment of the Himalayas. A companion volume to the Flora of Nepal* (Edinburgh: Royal Botanic Garden Edinburgh), 140–144.

Wendrich, W. (1989). "Preliminary report on the Amarna basketry and cordage," in *Amarna Reports*. (London: Egypt Exploration Society), 169–201.

Whigham, D. F., Dykyjova, D., and Hejn, S. (1993). Wetlands of the world I lnventory, ecology, and management. *Springer Sci.* doi: 10.1007/978-94-015-8212-4

White, F., and Leonard, J. (1991). Phytogeographical links between Africa and Southwest Asia. *Flora Vegetatio Mundi* 9, 229–246.

Winter, K. (1981). C_4 plants of high biomass in arid regions of asia-occurrence of C_4 photosynthesis in Chenopodiaceae and Polygonaceae from the Middle East and USSR. *Oecologia* 48, 100–106. doi: 10.1007/BF00346994

Wooler, M., Swain, D., Street-Perrot, E., Mathai, S., and Agnew, A. (2001). An Altitudinal And Stable Carbon Isotope Survey of C_3 and C_4 Graminoids On Mount Kenya. *J. East Afr. Natural Hist.* 90, 69–85. doi: 10.2982/0012-8317(2001)90[69:AAASCI]2.0.CO;2

Zarrin, A., Ghaemi, H., Azadi, M., and Farajzadeh, M. (2009). The spatial pattern of summertime subtropical anticyclones over Asia and Africa: a climatological review. *Int. J. Climatol.* 30, 159–173. doi: 10.1002/joc.1879

Zarrin, A., Ghaemi, H., Azadi, M., Mofidi, A., and Mirzaei, E. (2011). The effect of the Zagros Mountains on the formation and maintenance of the Iran Anticyclone using RegCM4. *Meteorol. Atmos. Phys.* 112, 91–100. doi: 10.1007/s00703-011-0134-z

Zohary, M. (1973). *Geobotanical foundations of the Middle East* (Stuttgart: Fischer).

Zohary, M., and Hopf, M. (2000). *Domestication of plants in the Old World. The Origin and Spread of Cultivated Plants in West Asia, Europe, and the Nile Valley*. (Oxford: Oxford University Press), 316.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

NOMENCLATURAL APPENDIX

CHENOPODIACEAE

Disintegration of *Salsola* s.l. into several lineages following molecular evidences lead to ample nomenclatural changes (Akhani et al., 2007). A major change was distinguishing *Kali* as a separate genus encompassing species of *Salsola* sect. *Salsola* with the type of *Kali turgida* (Dumort.) Guterm., (Gutermann, 2011). This system was accepted by many botanists and most nomenclatural changes were fixed according to the rules of International Code of Nomenclature (Hakobyan, 2011; Brullo et al., 2015; see additional notes and references in Sukhorukov et al., 2019). Two additional independent studies using one nuclear (ITS) and four plastid markers (*atp*B-*rbcL* spacer, *ndh*F-*rpl*32 spacer, *trn*Q-*rps*16 spacer, *rpl*16 intron) strongly confirmed the monophyly of *Salsola* s. str. sensu Akhani et al. (2007) (Voznesenskaya et al., 2013; Schüssler et al., 2017). A proposal to retypify the genus *Salsola* with *Salsola kali* by Mosyakin et al. (2014) did not get a consensus in Nomenclatural Committee (Applequist, 2016). However, in the last edition of the ICN, *Salsola kali* L. is accepted as the type of the genus *Salsola* (Turland et al., 2018). This unexpected decision necessitates several nomenclatural changes which we apply for those affected species occurring in the area of this paper.

Caroxylon omanense* (Boulos) Akhani and Rudov *comb. nov. Basionym: *Salsola omanensis* Boulos, Kew Bull. 46(2): 297 (1991).

Soda austro-iranica* (Akhani) Akhani *comb. nov. Basionym: *Salsola austro-iranica* Akhani in Pl. Veg. N. W. Persian Gulf 286 (2015).

Soda cyrenaica* (Maire and Weiller) Akhani *comb. nov. Basionym: *Darniella cyrenaica* Maire and Weiller, Bull. Soc. Hist. Nat. Afrique N. 1939, xxx. 301.

Soda drummondii* (Ulbr.) Akhani *comb. nov. Basionym: Nat. Pflanzenfam., ed. 2 [Engler and Prantl] 16c: 565 (1934).

Soda florida* (M. Bieb.) Akhani *comb. nov. Basionym: *Anabasis florida* M.Bieb., Fl. Taur.-Caucas. 1: 190 (1808).

Soda foliosa* (L.) Akhani *comb. nov. Basionym: *Anabasis foliosa* L., Sp. Pl. 1: 223 (1753).

Soda grandis* (Freitag, Vural and N. Adiguzel) Akhani *comb. nov. Basionym: *Salsola grandis* Freitag, Vural and Adıgüzel, Willdenowia 29(1–2): 131 (1999).

Soda cinerea* (Moq.) Akhani *comb. nov. Basionym: *Halogeton*? *cinerea* Moq., Chenopod. Monogr. 159 (1840).

Soda kerneri* (Wol.) Akhani *comb. nov. Basionym: *Hypocylix kerneri* Woł., Denkschr. Kaiserl. Akad. Wiss., Wien. Math.-Naturwiss. Kl. 51(2): 276 (1886).

Soda longifolia* (Forssk.) Akhani *comb. nov. Basionym: *Salsola longifolia* Forssk., Fl. Aegypt.-Arab. 55. (1775).

Soda makranica* (Freitag) Akhani *comb. nov. Basionym: *Salsola makranica* Freitag, Fl. Iranica [Rechinger] 172: 166 (1997).

Soda oppositifolia* (Desf.) Akhani *comb. nov. Basionym: *Salsola oppositifolia* Desf., Fl. Atlant. 1: 219 (1798).

Soda rosmarinus* (Ehrenb. ex Boiss.) Akhani *comb. nov. Basionym: *Seidlitzia rosmarinus* Ehrenb. ex Boiss., Fl. Or. 4: 951, 1879.

Soda schweinfurthii* (Solm.) Akhani *comb. nov. Basionym: *Salsola schweinfurthii* Solms, Bot. Zeitung, 2. Abt. 59: 173, in obs. (1901).

***Soda stocksii* (Boiss.) Akhani comb. nov.** Basionym: *Salsola stocksii* Boiss., Diagn. Pl. Orient. ser. 2, 4: 75 (1859).

Poaceae

The phylogenetic relationships within Poaceae has been investigated in several recent revisions (Hodkinson et al., 2002; Kellogg, 2015; Soreng et al., 2015; 2017). These revisions regarded also changes within the PACMAD clade, which includes all C_4 Poaceae lineages. According to these revisions *Pennisetum* und *Snowdenia* has been synonymized with *Cenchrus* and *Brachiaria* (s.s.) with *Urochloa*. Several species have been however neglected by the taxonomic changes. We propose here new combinations for the following four species, that have been affected by the last revision by Soreng et al. (2017):

Cenchrus aethiopicus* (Fresen.) Rudov *comb. nov. Basionym: *Beckera polystachya* Fresen., Mus. Senckenberg. ii. 132 (1837).

Cenchrus glaucifolius* (Hochst. ex A.Rich.) Rudov and Akhani *comb. nov. Basionym: *Pennisetum glaucifolium* Hochst. ex A. Rich. Tent. Fl. Abyss. 2: 382 (1841).

Cenchrus nubicus* (Hochst.) Rudov and Akhani *comb. nov. Basionym: *Gymnotrix nubica* Hochst., Flora 27(1): 251 (1844).

***Cenchrus yemensis* (Deflers) Rudov and Akhani comb. nov.** Basionym: *Pennisetum yemense* Deflers, Voyage Yemen 217 (1889).

Urochloa arida (*Mez) Rudov *comb. nov. Basionym: *Panicum aridum* Mez, Bot. Jahrb. Syst. 34(1): 139 (1905.

Urochloa chusqueoides* (Hack.) Rudov *comb. nov. Basionym: *Panicum chusqueoides* Hack., Bull. Herb. Boissier iii. 377 (1895).

Urochloa ovalis* (Stapf) Rudov *comb. nov. Basionym: *Brachiaria ovalis* Stapf, Fl. Trop. Afr. [Oliver et al.] 9(3): 546 (1919).

frontiers in Plant Science

ORIGINAL RESEARCH
published: 08 April 2021
doi: 10.3389/fpls.2021.559967

ZmOrphan94 Transcription Factor Downregulates *ZmPEPC1* Gene Expression in Maize Bundle Sheath Cells

Alicja M. Górska[1,2†‡], Paulo Gouveia[1,2†], Ana Rita Borba[1,2‡], Anna Zimmermann[1,2,3], Tânia S. Serra[1,2], Pedro Carvalho[1], Tiago F. Lourenço[1,2], M. Margarida Oliveira[1,2], Christoph Peterhänsel[3] and Nelson J. M. Saibo[1,2]*

[1]*Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal,* [2]*Instituto de Biologia Experimental e Tecnológica, Oeiras, Portugal,* [3]*Institut für Botanik, Leibniz Universität Hannover, Hannover, Germany*

Spatial separation of the photosynthetic reactions is a key feature of C_4 metabolism. In most C_4 plants, this separation requires compartmentation of photosynthetic enzymes between mesophyll (M) and bundle sheath (BS) cells. The upstream region of the gene encoding the maize PHOSPHOENOLPYRUVATE CARBOXYLASE 1 (ZmPEPC1) has been shown sufficient to drive M-specific *ZmPEPC1* gene expression. Although this region has been well characterized, to date, only few *trans*-factors involved in the *ZmPEPC1* gene regulation were identified. Here, using a yeast one-hybrid approach, we have identified three novel maize transcription factors ZmHB87, ZmCPP8, and ZmOrphan94 as binding to the *ZmPEPC1* upstream region. Bimolecular fluorescence complementation assays in maize M protoplasts unveiled that ZmOrphan94 forms homodimers and interacts with ZmCPP8 and with two other *ZmPEPC1* regulators previously reported, ZmbHLH80 and ZmbHLH90. Trans-activation assays in maize M protoplasts unveiled that ZmHB87 does not have a clear transcriptional activity, whereas ZmCPP8 and ZmOrphan94 act as activator and repressor, respectively. Moreover, we observed that ZmOrphan94 reduces the trans-activation activity of both activators ZmCPP8 and ZmbHLH90. Using the electromobility shift assay, we showed that ZmOrphan94 binds to several *cis*-elements present in the *ZmPEPC1* upstream region and one of these *cis*-elements overlaps with the ZmbHLH90 binding site. Gene expression analysis revealed that *ZmOrphan94* is preferentially expressed in the BS cells, suggesting that ZmOrphan94 is part of a transcriptional regulatory network downregulating *ZmPEPC1* transcript level in the BS cells. Based on both this and our previous work, we propose a model underpinning the importance of a regulatory mechanism within BS cells that contributes to the M-specific *ZmPEPC1* gene expression.

Keywords: C4 metabolism, photosynthesis, transcriptional regulation, *cis*-elements, phosphoenolpyruvate carboxylase 1, cell-specific gene expression

OPEN ACCESS

Edited by:
Sarah Covshoff,
Independent researcher, Las Vegas, United States

Reviewed by:
Tsuyoshi Furumoto,
Ryukoku University, Japan
Pinghua Li,
Shandong Agricultural University, China

***Correspondence:**
Nelson J. M. Saibo
saibo@itqb.unl.pt

[†]These authors have contributed equally to this work

***Present address:**
Alicja M. Górska, Institute of Plant Sciences, University of Graz, Graz, Austria
Ana Rita Borba, Department of Plant Sciences, Cambridge University, Cambridge, United Kingdom

Specialty section:
This article was submitted to Plant Physiology, a section of the journal Frontiers in Plant Science

Received: *12 May 2020*
Accepted: *01 February 2021*
Published: *08 April 2021*

Citation:
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. doi: 10.3389/fpls.2021.559967

INTRODUCTION

Most plants use ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) as the primary carbon dioxide (CO_2) fixing enzyme in a process called Calvin-Benson cycle. However, due to its dual activity (carboxylase and oxygenase) and the high atmospheric O_2 concentration, RuBisCO shows a high oxygenase activity in C_3 plants (Portis and Parry, 2007). This activity leads to the production of 2-phosphoglycolate, which is toxic for the plant and needs to be recycled through a process called photorespiration. This is a wasteful process, which leads to the loss of C, N, and ATP, thus decreasing photosynthetic efficiency (Bauwe et al., 2010). C_4 plants, which have evolved independently from C_3 species over 60 times (Sage, 2004), have a carbon concentrating mechanism that significantly minimizes photorespiration. To concentrate CO_2 around RuBisCO, C_4 plants developed a spatial separation of photosynthetic reactions between mesophyll (M) and bundle sheath (BS) cells. In M cells, carbonic anhydrase (CA) converts atmospheric CO_2 to bicarbonate (HCO_3^-), which in presence of phosphoenolpyruvate (PEP) is fixed by phosphoenolpyruvate carboxylase (PEPC) to form oxaloacetate (OAA). Subsequently, OAA is rapidly converted to malate or aspartate, which diffuses into BS cells, where RuBisCO is present. Decarboxylation of these organic acids in the BS cells leads to a high CO_2 concentration around RuBisCO, which then incorporates CO_2 into the Calvin-Benson cycle in a highly efficient way (Furbank and Taylor, 1995).

All enzymes required for C_4 photosynthesis are present in C_3 species but they are low abundant and/or present in both cell types (Aubry et al., 2011). Thus, for C_4 genes to be highly expressed and restricted to either M or BS cells, changes at *trans*- and *cis*-regulatory level had to occur (Hibberd and Covshoff, 2010; Reeves et al., 2017). A number of *cis*-elements involved in cell-specific expression of C_4 genes have already been identified. For example, BS-specific expression of C_4 genes has been associated with sequences within their promoter (Wiludda et al., 2012), untranslated region (Patel et al., 2006), and coding sequences (Brown et al., 2011; Reyna-Llorens et al., 2018), whereas sequences within untranslated regions (Kajala et al., 2012; Williams et al., 2016) and promoters (Stockhaus et al., 1997; Nomura et al., 2000; Gowik et al., 2016; Gupta et al., 2020) have been associated with M-specific expression.

M-specific *PEPC* gene expression has been mainly associated with regulation at promoter level. For example, in *Flaveria trinervia*, a specific promoter domain called M expression module 1 (MEM1) was reported to drive M-specific *PEPC* expression (Gowik et al., 2004). MEM1 is a 41 bp element located in a distal *ppcA* promoter region. It functions as an enhancer element, conferring M-specific gene expression, and acts as a repressor of *ppcA* expression in BS and in vascular bundle (Gowik et al., 2004; Akyildiz et al., 2007). In maize, a 0.6 kb *ZmPEPC1* upstream region was shown to be sufficient to drive M-specific gene expression (Taniguchi et al., 2000; Kausch et al., 2001). M-specific gene expression was also observed when a 1.2 kb *ZmPEPC1* upstream region driving the *β-glucuronidase* (*GUS*) reporter gene was transformed in rice, a C_3 plant (Matsuoka et al., 1994). Although, the upstream regions involved in M-specific *ZmPEPC1* expression have been determined, the knowledge about trans-factors involved in this regulation remains scarce. The transcription factors (TFs) identified as binding to the *ZmPEPC1* upstream region include maize nuclear factors (MNFs; Yanagisawa and Izui, 1990, 1992), PEP-I (Kano-Murakami et al., 1991), DOF1 and DOF2 (Yanagisawa and Sheen, 1998), and ZmbHLH80 and ZmbHLH90 (Górska et al., 2019). Interestingly, in the *zmdof1* maize knockdown mutant, no alterations in *ZmPEPC1* gene expression were observed, suggesting a possible redundant function of TFs binding to the *ZmPEPC1* promoter (Cavalar et al., 2007). Thus, to understand the mechanism regulating the *ZmPEPC1* expression, a different approach is needed. Instead of focusing on individual TFs, it is important to investigate the transcriptional regulatory machinery involved in C_4 *ZmPEPC1* expression. It is, therefore, important to identify novel TFs binding to the *ZmPEPC1* promoter and to determine their interactions with other TFs as well as with other proteins. The biological meaning of all these interactions must also be studied.

In this study, we aimed to identify novel maize TFs involved in the M-specific *ZmPEPC1* gene expression and to determine their function. Using a yeast one-hybrid (Y1H) system, three TFs belonging to different TF families were identified as binding to *ZmPEPC1* upstream region and functionally characterized. We show that one of the novel TFs is a BS-preferentially expressed repressor that interacts with other *ZmPEPC1* regulators and reduces their trans-activation activity. Based on our findings, we propose a model in which a new repressor together with previously identified TFs jointly contribute to the *ZmPEPC1* cell-specific gene expression.

MATERIALS AND METHODS

Yeast One-Hybrid Screening of the Maize Leaf cDNA Expression Library

The 781 bp *ZmPEPC1* (GRMZM2G083841) upstream region (starting from ATG) was divided into three overlapping fragments (5 U, F1, and F2) and these were amplified by PCR using primers listed in **Supplementary Table S3**. The isolated fragments were cloned into the pINT/HIS vector system (Ouwerkerk and Meijer, 2001) and integrated into the yeast strain Y187 (Clontech, CA, United States) to originate the yeast bait strains. The yeast baits were then transformed with 1 μg of maize cDNA expression library as described by Ouwerkerk and Meijer (2001). The maize cDNA expression library used in this study was described by Borba et al. (2018). For each bait, at least 1×10^6 yeast colonies were screened. The screenings were performed in complete minimal (CM)/-His/-Leu medium supplemented with: 25 mM (5 U), 10 mM (F1), and 5 mM (F2) of 3-amino-1,2,4-triazole (3-AT). Plasmids from the yeast colonies grown on the selection media were extracted, sequenced, and the obtained cDNA insert sequences were analyzed using BLAST program. The plasmids containing ZmHB87 (GRMZM26163641), ZmOrphan94 (GRMZM2G127426), and ZmCPP8 (GRMZM2G096600) and the empty vector (EV) were re-transformed into the *ZmPEPC1*

yeast baits (5 U, F1, and F2). The growth of the transformed yeast was analyzed on CM/-His/-Leu medium supplemented with increasing concentration of 3-AT.

Plant Materials and Growth Conditions

Maize plants (B73) used for M protoplast isolation, M and BS cell isolation, as well as for diurnal gene expression studies were grown as described by Górska et al. (2019).

Isolation and Transformation of Maize and Rice Protoplasts

Maize M protoplasts were isolated from second leaves of 10-day-old maize etiolated seedlings using a modified protocol from Sheen (1991). In brief, the mid veins from the second leaves were removed and leaves were cut into approximately 0.5–1 cm strips. The strips were transferred to an enzyme solution (0.4 M mannitol, 10 mM MES pH 5.7, 1 mM $CaCl_2$, 0.1% BSA, 50 mg L-1 ampicillin, 5 mM β-mercaptoethanol), 1.5% Cellulase R10 (Duchefa, Haarlem, The Netherlands), and 0.3% Macerozyme R10 (Duchefa, Haarlem, The Netherlands) and subjected to vacuum infiltration for 30 min. Afterward, digestion was continued for 5 h at 25–27°C with gentle agitation (40 rpm) in dark. After 5 h, the protoplasts were released by increasing the agitation to 85 rpm for 15 min. The enzyme solution containing protoplasts was filtered twice through a 100 µm mesh, washed with 1x volume of wash solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, and 2 mM MES pH 5.7) and filtered again through a 50 µm filter. Protoplasts were harvested by centrifugation in a "swing-out" bucket (150 × *g*, 5 min) and resuspended in 200 µl MMg solution (0.4 M mannitol, 4 mM MES pH 5.7, 15 mM MgCl2). Afterward, protoplasts were diluted to a 2×10^6 ml^{-1} concentration and permeabilized/transformed with polyethylene glycol (PEG) by gentle mixing 200 µl of protoplast solution with 20 µl of plasmid DNA mix and 220 µl of PEG solution (PEG 4000 40%, 0.3 M mannitol, 0.1 M $CaCl_2$). After being incubated at room temperature (RT) in dark for 20 min., protoplasts were diluted with 3x volume of wash solution, harvested by centrifugation in a "swing-out" bucket (150 × *g*, 5 min) and resuspended in 750 µl of incubation solution (0.6 M mannitol, 4 mM MES pH 5.7, 4 mM KCl). After this, protoplasts were transferred to a 24-well-plate containing additional 750 µl of incubation solution with 50 mg L-1 ampicillin and incubated for 15–16 h at RT in dark.

Rice protoplasts were obtained using a protocol described by Cordeiro et al. (2016), with modifications. 2-day-old rice (cv. Nipponbare) cell suspension cultures were collected by centrifugation in a "swing-out" bucket (150 × *g*, 5 min). Collected cells were resuspended in an enzyme solution and subjected to vacuum infiltration. Following enzymatic digestion, the enzyme solution containing protoplasts was filtered through a 100 µm mesh, washed with 1x volume of wash solution and filtered through a 50 µm mesh. Protoplasts were harvested by centrifugation in a "swing-out" bucket (150 × *g*, 5 min) and resuspended in 200 µl MMg solution. These protoplasts were then diluted to a 1×10^6 ml^{-1} concentration and transformed with PEG by gentle mixing 200 µl of protoplast solution with 10 µl of plasmid DNA mix and 220 µl of PEG solution. Afterwards, transformed protoplasts were incubated at RT in dark for 20 min., diluted with 3x volume of wash solution, harvested by centrifugation in a "swing-out" bucket (150 × *g*, 5 min) and resuspended in 750 µl of incubation solution (0.4 M mannitol, 4 mM MES pH 5.7, 20 mM KCl). The transformed protoplasts were incubated for 15–16 h at RT in dark.

Bimolecular Fluorescence Complementation Assay

ZmHB87, ZmOrphan94, and ZmCPP8 full-length CDS were amplified from maize (*Zea mays* B73) cDNA by PCR, using primers listed in **Supplementary Table S3**. The amplified products were recombined into pDONR221 (Invitrogen, CA, United States) according to manufacturer's instructions, confirmed by sequencing, and cloned into pYFPN43 and pYFPC43 plasmids to be in fusion with the N- and C-terminal halves of the yellow fluorescent protein (YFP), respectively. The final vectors prepared were then analyzed by digestion with restriction enzymes. Cloning of ZmbHLH80 and ZmbHLH90 TFs into pYFPC43 and pYFPN43 vectors was described by Górska et al. (2019). The resulting pYFPC43 and pYFPN43 constructs (6 µg of each plasmid) were transformed into maize M protoplasts. Each transformation was performed in triplicate. pYFPN43::ZmOrphan94 co-transformed with pYFPC::Akin3 (*Arabidopsis* SNF1 Kinase Homologue 3) and pYFPC43::ZmOrphan94 co-transformed with pYFPN43::Akin10 (*Arabidopsis* SNF1 Kinase Homologue 10) served as negative controls. The transformed protoplasts were incubated for 15–16 h at RT in dark, and reconstitution of the fluorescence signal was observed using confocal laser scanning microscopy (Leica SP5).

Trans-Activation Assays in Maize and Rice Protoplasts

The construction of 5 U-*ZmPEPC1* and unrelated DNA sequence (US) reporter vectors, *firefly* luciferase (LUC) transformation control plasmid, as well as ZmbHLH80, ZmbHLH90, and EV effector plasmids was described by Górska et al. (2019). For F2-*ZmPEPC1* reporter plasmid, a sequence of the F2-*ZmPEPC1* fragment used in the Y1H screening was cloned (through BP-Gateway reaction) into the p2GW7m35S::GUS plasmid. For the effector constructs, the ZmHB87, ZmOrphan94, and ZmCPP8 entry clones were recombined *via* LR-Gateway reaction into p2GW7 plasmid. The trans-activations assays were carried out by transforming maize M protoplasts with 6, 6, and 16 µg of reporter, transformation control, and effector plasmids, respectively. Each transformation was performed in triplicate. The trans-activation activity of the TFs was calculated as GUS/LUC ratio. The transformed protoplasts were incubated for 15–16 h at RT in dark. Cell lysis and determination of GUS and LUC activity levels were performed as described by Figueiredo et al. (2012).

To analyze the trans-activation activity of ZmOrphan94 on 5 U-*ZmPEPC1* and mut5U-*ZmPEPC1*, we used a dual luciferase system. Thus, to construct the 5 U-*ZmPEPC1* and mut5U_*ZmPEPC1* reporter plasmids, a sequence of the 5 U-*ZmPEPC1* fragment, used in the Y1H screening, and mut5U-ZmPEPC1 (5 U-*ZmPEPC1* fragment with the ZmOrphan94 binding sites mutated from CACA to TATA), were cloned into the pGreenII 0800-LUC plasmid (Hellens et al., 2005), upstream of a minimal CaMV35S promoter driving LUC, using the restriction enzymes *Nco*I and *Spe*I. The transformation control is calculated using the activity of *renilla* luciferase (REN), which is expressed by the same vector under the control of the CaMV35S promoter. Regarding the effector, the ZmOrphan94 entry clone was recombined *via* LR-Gateway reaction into the p2GW7 plasmid. The trans-activations assays were carried out by transforming rice protoplasts with 5 µg of reporter plus 5 µg of effector plasmids. Each transformation was performed in triplicate and results shown correspond to two independent experiments. The trans-activation activity of the TFs was calculated as LUC/REN ratio. The transformed protoplasts were incubated for 15–16 h at RT in dark. Cell lysis and determination of LUC and REN activity were performed with the Dual-Luciferase® reporter assay (Promega, United States), using a modified protocol. Briefly, protoplasts were collected by centrifugation and cell lysis performed using 100 µl of 1x Passive Lysis Buffer. LUC and REN activity reactions were performed in 96-well plates using 50 µl of cell lysate, to which 30 µl of LARII reagent was added for LUC activity and 30 µl of Stop & Glo® reagent for REN activity. Luminescent was detected using FLUOStar Optima (BMG LabTech, Germany) microplate reader. Each sample was analyzed for each luciferase (LUC and REN) with measurements every 0.5 s during 12 s of luminescence acquisition.

Production of Recombinant Trx::ZmOrphan94 Protein

The full-length CDS of ZmOrphan94 was amplified from maize (*Zea mays* B73) cDNA by PCR using primers listed in **Supplementary Table S3**. The amplified sequence was cloned as an *Eco*RI-*Xho*I fragment into pET32a vector (Novagen) to raise an N-terminal translational fusion with Thioredoxin (Trx) and transformed into *Escherichia coli* Rosetta (DE3) pLysS competent cells for the expression of the recombinant protein. The bacterial cells transformed with Trx::ZmOrphan94 plasmid were grown in Luria-Bertani (LB) medium at 37°C to an OD_{600} of 0.6. Subsequently, the expression of Trx::ZmOrphan94 recombinant protein was induced with 4 mM isopropyl-D-1-thiogalactopyranoside (IPTG) and continued for 16 h at 18°C. The Trx::ZmOrphan94 protein purification was performed as described by Cordeiro et al. (2016).

Electrophoretic Mobility Shift Assay

For electrophoretic mobility shift assay (EMSA), DNA probes were generated by annealing oligonucleotide pairs and radiolabelling as described by Serra et al. (2013). Oligonucleotide sequences and respective annealing temperatures are listed in **Supplementary Table S2**. The binding reactions were performed in a 10 µl volume containing 1 µg of Trx::ZmOrphan94, 50 fmol of radiolabelled probe, 10 mM HEPES (pH 7.9), 40 mM KCl, 1 mM EDTA (pH 8), 1 mM DTT, 50 ng herring sperm DNA, 15 µg BSA and 10% (v/v) glycerol for 1 h on ice. The resulting complexes were resolved on a native 5% polyacrylamide gel (37.5:1). Competition assays were performed by adding 200- to 400-fold molar excess of the unlabelled probe. Trx protein was used as negative control. Gel electrophoresis and detection of radioactive signal were performed as described by Serra et al. (2013).

Isolation of Maize Mesophyll and Bundle Sheath Cells

Mesophyll cells were isolated from the third leaves of 10-day-old maize seedlings according to Covshoff et al. (2013) with the following modification: the isolated M cells were collected to a tube containing 450 µl RLT buffer from the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). The same leaves that were used to extract M cells, were further used in BS isolation. From this point on, the isolation of BS was performed as described by Markelz et al. (2003) with the following modification: high speed shredding was carried out three times for 1 min each. Samples for cell isolation were harvested at 6 and 2 h before lights turned on (−6 h, −2 h), when the lights turned on (0 h), and 2 h after illumination (+2 h). The samples collection at time point −6 and −2 h was performed under green light. For each time point, three biological replicates of M and BS were prepared using five leaves per replicate.

RNA Isolation, cDNA Synthesis, and Quantitative PCR

Total RNA was extracted from purified M and BS samples, and from whole maize leaves using the RNeasy Plant kit (Qiagen, Hilden, Germany). After isolation, total RNA was treated with Turbo DNase (Ambion, CA, United States) according to the manufacturer's instructions. The quality of the RNA was assessed by NanoDrop 2000c Spectrophotometer (Thermo Fisher Scientific, MA, United States) and by gel electrophoresis. First strand cDNA was synthesized using SuperScript® III First-Strand Synthesis System (Invitrogen, CA, United States) following the manufacturer's instructions. Two-hundred and eighty and five-hundred nanogram of total RNA from the purified cell samples and whole maize leaves, respectively, were used to synthesize cDNA using oligo (dT) primers. The quantitative PCR (qPCR) was performed using SYBR Green I Master mix (Roche, Basel, Switzerland) on a LightCycler 480 system (Roche, Basel, Switzerland). Threshold cycles (Ct) values were calculated from means of three biological replicates, three technical replicates each. The Ct values were normalized against *ZmActin*1 (GRMZM2G126010) for diurnal analysis, and GRMZM2G144843 and GRMZM2G044552 for cell-specific analysis. GRMZM2G144843 and GRMZM2G044552 were selected based on their stable

expression between M and BS cells and along the time (data not shown). Gene specific primers used in qPCR are listed in **Supplementary Table S3**.

Direct Yeast One-Hybrid

The full-length coding sequence of *OsOrphan65* was amplified from rice (*Oryza sativa* L, cv. Nipponbare) cDNA using the primers listed in **Supplementary Table S3**. The amplified product was recombined in pDONR221 (Invitrogen, CA, United States), confirmed by sequencing and cloned into the pDEST22 (Invitrogen, CA, United States) plasmid, according to manufacturer's instructions. Integrity of the expression clone was analyzed by digestion with restriction enzymes. For the direct yeast Y1H, OsOrphan65::pDEST22, ZmbHLH90::pDEST22 (positive control), and pDEST22 (negative control) were individually transformed into the yeast bait strains containing 5 U, F1, and F2 *ZmPEPC1* upstream fragments. The growth of the transformed yeast baits was analyzed on CM/-His/-Trp medium and with increasing concentrations of 3-AT.

RESULTS

ZmHB87, ZmCPP8, and ZmOrphan94 Bind to the *ZmPEPC1* Upstream Region

To identify additional components of the *ZmPEPC1* regulatory network and isolate TFs involved in the M-specific *ZmPEPC1* gene expression, we used a Y1H approach to screen a maize cDNA expression library. The Y1H screening was carried out using overlapping fragments of the *ZmPEPC1* upstream region (5 U, F1, and F2; **Figure 1A**) as baits. These fragments were cloned upstream of the *HIS3* reporter gene and the resulting constructs were individually integrated into yeast genome to generate 5 U-*ZmPEPC1*, F1-*ZmPEPC1*, and F2-*ZmPEPC1* yeast bait strains, respectively. These baits were transformed with the maize cDNA expression library and the yeast growth on CM/-Leu/-His selection medium supplemented with 3-amino-1,2,4-triazole (3-AT), a competitive inhibitor of the *HIS3* gene product, was analyzed. Among the colonies that grew on the selection media, we identified three different TFs, ZmCPP8, ZmHB87, and ZmOrphan94. ZmHB87 and ZmOrphan94 were identified as binding to 5 U fragment, whereas ZmCPP8 was found to interact with F2 (**Figure 1A**). No TFs were identified as binding to the F1 fragment. To validate the TF-DNA interactions and determine their specificity, we isolated the plasmids and re-transformed each of the *ZmPEPC1* baits (5 U-*ZmPEPC1*, F1-*ZmPEPC1*, and F2-*ZmPEPC1*) with the plasmids expressing the identified TFs and with the EV as negative control. The growth of the re-transformed bait strains was analyzed on CM/-Leu/-His media supplemented with increasing concentrations of 3-AT. According to our results, ZmHB87 and ZmOrphan94 bind specifically to the 5 U fragment (**Figure 1B**). Though the 5 U-*ZmPEPC1* bait strain transformed with the EV grew on the CM/-Leu/-His selection medium without 3-AT, the presence of 5 mM 3-AT was enough to abolish yeast growth. The same bait strain transformed with the plasmids expressing ZmHB87 or ZmOrphan94 grew on the CM/-Leu/-His selection medium supplemented with up to 10 mM 3-AT (**Figure 1B**), showing the authenticity of their protein-DNA interactions. Regarding ZmCPP8, our results indicate that it binds specifically to the F2 fragment. When the F2-*ZmPEPC1* bait was transformed with the EV, it did not grow on CM/-Leu/-His, but it grew when transformed to express ZmCPP8 (**Figure 1B**). However, our results suggest that the interaction between ZmCPP8 and the F2 fragment is not very strong as the yeast growth is eliminated with 5 mM 3-AT.

In silico analysis of ZmCPP8, ZmHB87, and ZmOrphan94 protein sequences revealed that all three TFs contain DNA-binding domains and nuclear localization signals, supporting their role as transcriptional regulators (**Figure 1C, Supplementary Figure S1A**).

ZmOrphan94 Forms Homodimers and Interacts With Other TFs Binding to the *ZmPEPC1* Upstream Region

Given that protein-protein interactions are an important feature influencing TF activity, we decided to investigate whether ZmCPP8, ZmHB87, and ZmOrphan94 form homodimers and/or heterodimers. To test this, all TFs were cloned into pYFN43 and pYFC43 vectors to raise N-terminal translational fusions with N- and C-halves of the YFP and then used to perform bimolecular fluorescent complementation (BiFC) assays. These assays were carried out in maize M protoplasts and all possible interactions were tested. According to our results, among the novel TFs, only ZmOrphan94 forms homodimers. A strong fluorescent signal was observed when YFP^N::*ZmOrphan94* was co-transformed with YFP^C::*ZmOrphan94* (**Figure 2**), but no fluorescence was detected in protoplasts co-transformed with YFP^N::*ZmCPP8* and YFP^C::*ZmCPP8*, nor with YFP^N::*ZmHB87* and YFP^C::*ZmHB87* (data not shown). Regarding the interactions between the novel TFs, our results show that ZmOrphan94 interacts with ZmCPP8 (**Figure 2**). An YFP fluorescent signal was detected in the nuclei of protoplasts co-transformed with YFP^N::*ZmOrphan94* and YFP^C::*ZmCPP8*, as well as with YFP^C::*ZmOrphan94* and YFP^N::*ZmCPP8*. No interactions were observed between ZmOrphan94 and ZmHB87, neither between ZmHB87 and ZmCPP8 (data not shown).

Our previous studies have identified two ZmbHLH TFs, ZmbHLH80 and ZmbHLH90, as binding to the *ZmPEPC1* upstream region (Górska et al., 2019). Thus, to gain deeper insights into the *ZmPEPC1* regulatory network, we decided to analyze whether the novel TFs could interact with these ZmbHLHs. Interestingly, we found that ZmOrphan94 interacts with either ZmbHLH80 or ZmbHLH90. Reconstitution of the YFP signal was observed in nuclei of protoplasts co-transformed with YFP^N::*ZmOrphan94* and YFP^C::*ZmbHLH90* or YFP^C::*ZmbHLH80*, and YFP^C::*ZmOrphan94* with YFP^N::*ZmbHLH80* or YFP^N::*ZmbHLH90*. No signal was detected when ZmOrphan94 was co-transformed with the unrelated proteins Akin10 or Akin3, thus validating the observed interactions. We also observed that neither ZmCPP8 nor

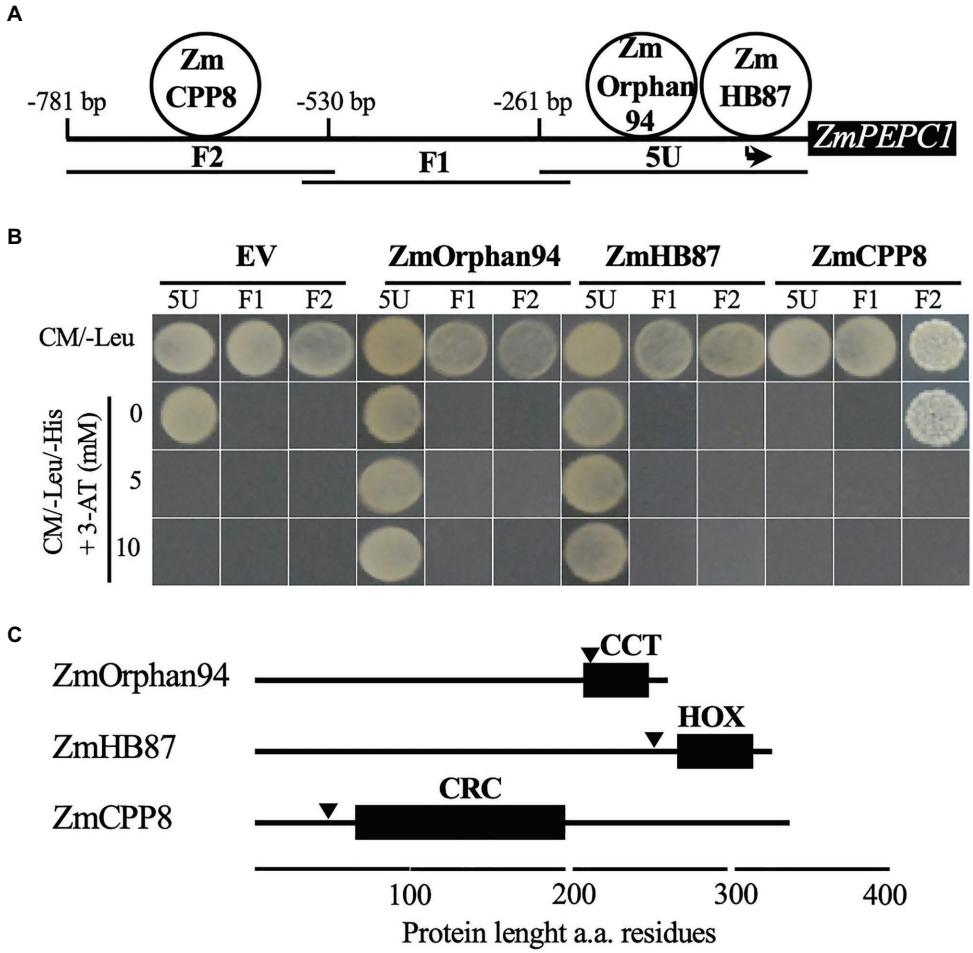

FIGURE 1 | Analysis of the transcription factors (TFs) binding to *ZmPEPC1* upstream region. **(A)** Schematic representation of the three *ZmPEPC1* upstream region fragments used in the yeast one-hybrid screenings. ZmHB87 and ZmOrphan94 were identified as binding to the 5 U fragment located between –261 and 0 bp upstream of *ZmPEPC1* ATG. The order of the TFs binding to 5 U fragment is representative but was not determined experimentally. ZmCPP8 was identified as binding to the F2 fragment located between –530 and –781 bp upstream of *ZmPEPC1* ATG. **(B)** Validation of the interactions between identified TFs and *ZmPEPC1* upstream region. Yeast bait strains carrying the different fragments of the *ZmPEPC1* upstream region were transformed with plasmids to express ZmHB87, ZmOrphan94, and ZmCPP8, as well as with empty vector (EV). Growth of the transformed bait strains was analyzed on complete minimal (CM)/-Leu/-His medium (CM medium lacking Leucine and Histidine) supplemented with increasing concentrations of 3-amino-1,2,4-triazole (3-AT). **(C)** Schematic representation of ZmHB87, ZmOrphan94, and ZmCPP8 protein structures with homeodomain (HOX), CONSTANS, CO-like, and TOC1 (CCT), and C1-RNPXAFXPK-C2 (CRC) DNA-binding domains, respectively, determined by ScanProsite. Nuclear localization signals (arrowheads) were predicted by cNLS mapper. a.a. refers to amino acid.

ZmHB87 interacts with either ZmbHLH80 or ZmbHLH90 (data not shown).

ZmOrphan94 Acts as a Repressor and Impairs ZmbHLH90-Mediated *ZmPEPC1* Trans-Activation

To determine trans-activation activity of the novel TFs on the *ZmPEPC1* promoter, we conducted a trans-activation assay in maize M protoplasts (prepared from etiolated seedlings). We transiently transformed the protoplasts with reporter constructs (e.g., 5 U-*ZmPEPC1*::m35::GUS) and effector plasmids expressing TFs under the control of the CaMV35S promoter (**Figure 3A**). Co-transformation of maize protoplasts with the reporter US (containing an unrelated DNA sequence before the minimal 35S) together with either EV or ZmHB87 did not change the GUS/LUC ratio (**Figure 3B**). The same result was observed when the reporter 5 U-*ZmPEPC1* was co-transformed with either EV or ZmHB87, indicating that ZmHB87 has no trans-activation activity. Regarding ZmCPP8, it activated the US reporter vector, but its activation activity on the F2-*ZmPEPC1*, which contains the fragment bound by ZmCPP8, was not statistically significant (**Figure 3C**). These results suggest that ZmCPP8 may act as an activator and that the US reporter contains *cis*-element(s) recognized by ZmCPP8. According to our data, ZmOrphan94 acts as a transcriptional repressor. When ZmOrphan94 was co-transformed with either US, 5 U-*ZmPEPC1*, or F2-*ZmPEPC1* reporters, it always reduced the GUS/LUC ratio, as compared with the EV. In addition,

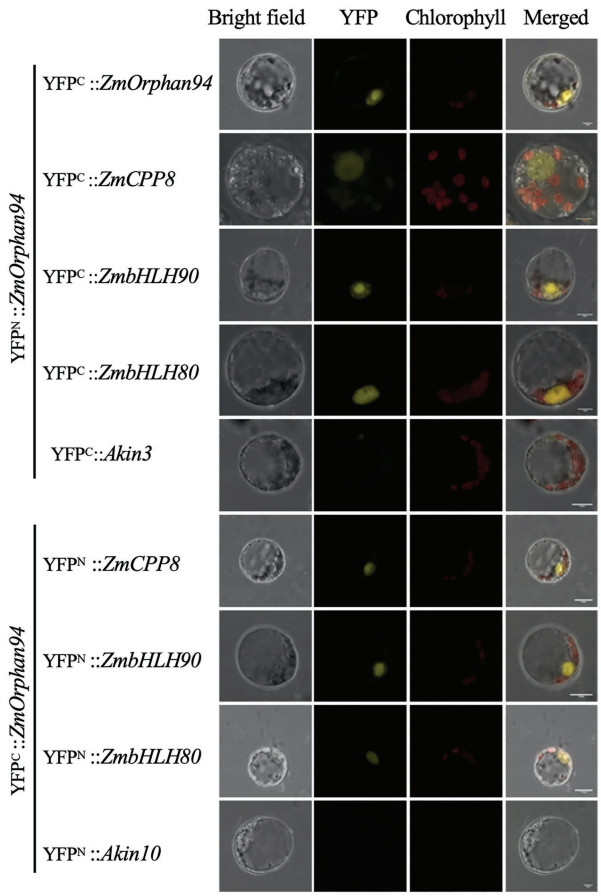

FIGURE 2 | Bimolecular fluorescence complementation (BiFC) analysis of protein-protein interactions between OsOrphan94 and other TF binding to *ZmPEPC1* upstream region. Pairs of proteins fused to complementary yellow fluorescent protein (YFP) halves were transiently expressed in etiolated maize mesophyll protoplasts. BiFC fluorescence is indicated as the YFP signal. Maize mesophyll protoplasts co-transformed with YFP^C::*Akin3* and YFP^N::*ZmOrphan94*, and YFP^N::*Akin10* with YFP^C::*ZmOrphan94* were used as negative controls. Scale bars = 10 μm.

the strongest repression was observed when ZmOrphan94 was co-transformed with 5 U-*ZmPEPC1* reporter (**Figure 3C**), which contains the *ZmPEPC1* upstream fragment where ZmOrphan94 binds (**Figure 1B**).

Our BIFC assays showed that ZmOrphan94 forms multiple heterodimers (**Figure 2**). It interacts with ZmCPP8, identified as binding to the F2-*ZmPEPC1* fragment, as well as ZmbHLH80 and ZmbHLH90, which were previously shown as binding to 5 U-*ZmPEPC1* (to which ZmOrphan94 also binds). To understand the role of these interactions regulating *ZmPEPC1* promoter activity, we also analyzed the trans-activation activity of ZmOrphan94 when acting together with the interacting TFs. Given that ZmCPP8 and ZmOrphan94 interact with each other but bind to different *ZmPEPC1* promoter fragments, we analyzed their trans-activation activity on both fragments. Consistent with its specific binding to the F2-*ZmPEPC1* fragment, ZmCPP8 did not show trans-activation activity on the 5 U-*ZmPEPC1* reporter vector, as compared with the EV (**Figure 3C**). However, when ZmCPP8 was co-transformed with ZmOrphan94, we observed a reduction of the GUS/LUC ratio for all the analyzed reporter vectors, as compared with that observed for ZmCPP8 alone (**Figure 3C**). When the ZmOrphan94 was co-transformed with the ZmbHLH80, we did not observe changes in the GUS/LUC ratio for US and 5 U-*ZmPEPC1*, as compared with those of ZmOrphan94 or ZmbHLH80 alone (**Figure 3D**). However, when ZmOrphan94 and ZmbHLH90 were co-transformed, ZmOrphan94 reduced the ZmbHLH90-mediated activation observed with US and 5 U-*ZmPEPC1* reporter vectors (**Figure 3D**), clearly indicating that ZmOrphan94 impairs the activation of *ZmPEPC1* caused by ZmbHLH90.

ZmOrphan94 and *ZmbHLH90* Show Similar Expression Profile

To understand the role of the interaction between ZmOrphan94 and ZmbHLH90 regulating *ZmPEPC1* gene expression, we analyzed the gene expression pattern of *ZmOrphan94*, *ZmbHLH90*, and *ZmPEPC1* over a period of 24 h. As shown in **Figure 4**, *ZmOrphan94* and *ZmbHLH90* have a similar diurnal transcript profile. Moreover, despite the difference in the transcript levels observed between these TFs and their target gene *ZmPEPC1*, all three genes show a very similar diurnal expression pattern. The three genes have a peak of expression at the end of the night or at the beginning of the day and all three are downregulated during the 1st hour (0.5–4 h after dawn) of the day. After this, with the exception of *ZmOrphan94*, which shows an unexpected increase of expression in the middle of the photoperiod (8 h after dawn), the transcript level of these genes is downregulated till the end of the photoperiod (16 h). The expression of the three genes is then induced all night long to reach the peak at the end of the night (0.5 h pre-dawn)/beginning of the day (0.5 h after dawn; **Figure 4**). The high correlation between *ZmOrphan94*, *ZmbHLH90*, and *ZmPEPC1* gene expression patterns indicates that ZmOrphan94 and ZmbHLH90 may act together to regulate *ZmPEPC1* gene expression.

ZmOrphan94 Binds to Different CACA Motifs Within the *ZmPEPC1* Upstream Region and One of Its Binding Sites Overlaps With the Binding Site of ZmbHLH90

Based on the trans-activation data, ZmOrphan94 impairs ZmbHLH90-mediated *ZmPEPC1* activation (**Figure 3D**). Given that ZmOrphan94 and ZmbHLH90 bind to the same 5 U-*ZmPEPC1* upstream fragment, it is possible that the effect observed on the *ZmPEPC1* expression is due to their binding to the same *cis*-element or to different *cis*-elements in close proximity. To understand where ZmOrphan94 binds within 5 U-*ZmPEPC1*, we first reviewed the literature to search for *cis*-elements described as binding sites for CCT domain-containing proteins, such as ZmOrphan94. It is reported that *Arabidopsis* TIMING OF CAB EXPRESSION 1 (TOC1), a CCT domain-containing TF, can bind to a 5′-CACA-3′ sequence

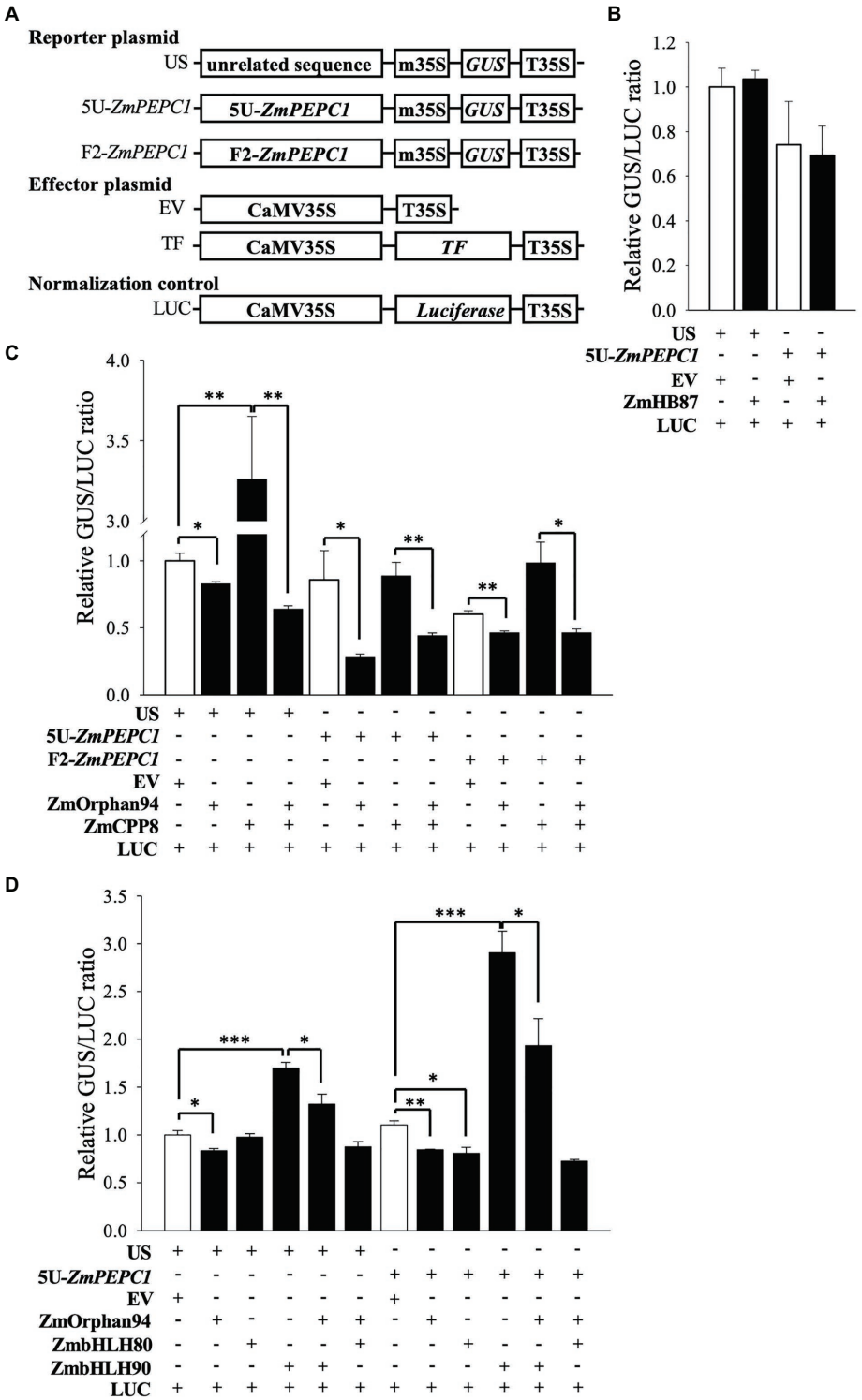

FIGURE 3 | Trans-activation assays to test the activity of the novel TFs. **(A)** Schematic representation of the constructs used in trans-activation assays in maize mesophyll protoplasts. T35S, cauliflower mosaic virus 35S terminator; EV, empty vector; m35S, minimal cauliflower mosaic virus 35S promoter; CaMV35S, full cauliflower mosaic virus 35S promoter; GUS, *β-glucuronidase*; LUC, *luciferase*; US, reporter vector harboring an Unrelated Sequence; 5 U-*ZmPEPC1* and F2-*ZmPEPC1*, reporter vectors harboring the *ZmPEPC1* upstream region fragments (5 U and F2, respectively) used in Y1H screening. Trans-activation activity of ZmHB87 **(B)**, ZmCPP8 and ZmOrphan94 individually and co-transformed **(C)**, and ZmOrphan94 individually and co-transformed with ZmbHLH80 or ZmbHLH90 **(D)** represented as a GUS/LUC ratio. Data represent means ± SEM ($n = 3$). Differences are statistically significant (*t*-test, $^{*}p < 0.05$; $^{**}p < 0.01$; and $^{***}p < 0.001$).

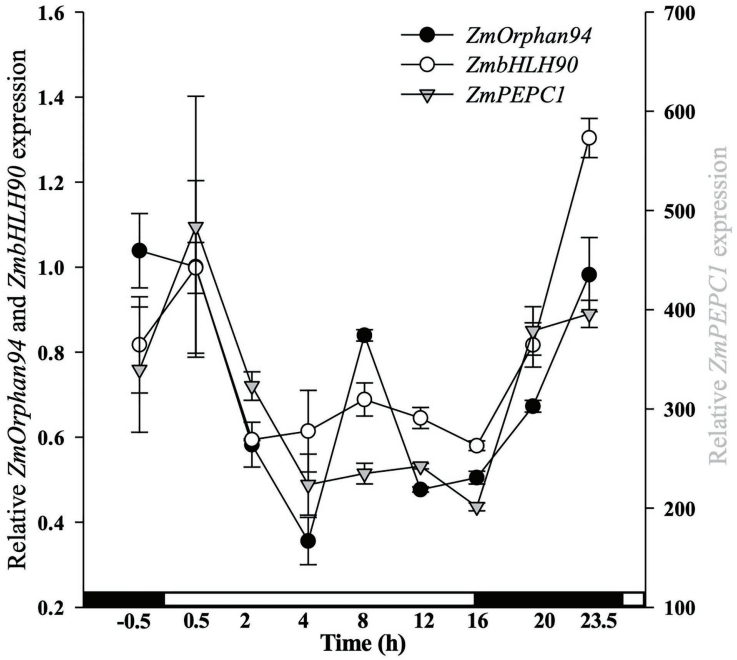

FIGURE 4 | Analysis of diurnal gene expression for *ZmOrphan94*, *ZmbHLH90*, and *ZmPEPC1*. The transcript level was assessed by RT-qPCR over a 24 h period. On the *x*-axis, white and black boxes indicate light and dark periods, respectively. Transcript level values were normalized to the expression of *ZmActin1* (GRMZM2G126010). Data represent means ± SEM (n = 3).

(Gendron et al., 2012). *In silico* analysis of the 5 U-*ZmPEPC1* fragment, revealed five predicted CACA motifs within the 5 U-*ZmPEPC1* sequence, with one of them overlapping with the ZmbHLH90 binding site (E-box; **Figures 5A**,**B**). To determine whether ZmOrphan94 binds to the 5'-CACA-3' sequences, we produced a full-length ZmOrphan94 recombinant protein and performed EMSA. As shown in **Figure 5C**, Trx::ZmOrphan94 bound to all 5 U-*ZmPEPC1* fragments containing CACA motifs (5 U-*ZmPEPC1*-1, 5 U-*ZmPEPC1*-2, and 5 U-*ZmPEPC1*-3), causing an uplift of the radiolabeled probes. The strongest band intensity was observed for 5 U-*ZmPEPC1*-3 probe, which contains multiple 5'-CACA-3' elements (**Figure 5C**). Binding of Trx::ZmOrphan94 to the labeled wild-type (WT) probes could be efficiently out-competed by unlabeled WT probes, thus validating the TF-DNA binding (**Figure 5C**). As a negative control, the probe containing CACA sequence was incubated with Trx alone and no gel mobility shift was observed (**Supplementary Figure S2**). In addition, Trx::ZmOrphan94 did not bind to 5 U-*ZmPEPC1*-0 probe lacking 5'-CACA-3' sequence (**Supplementary Figure S2**). To test whether ZmOrphan94 binds specifically to the CACA sequence(s) within the 5 U-*ZmPEPC1* fragment, we also generated mutated probes. As observed in **Figure 5C**, the mutations within the 5'-CACA-3' sequence led to a strong decrease in the Trx::ZmOrphan94-DNA complex band intensities.

In order to test the biological function of the ZmOrphan94 binding sites, we have performed an additional trans-activation assay, using as reporter the firefly LUC gene driven by either 5 U-*ZmPEPC1* fragment or mut5U-*ZmPEPC1* (5 U-*ZmPEPC1* sequence in which all ZmOrphan94 binding sites were mutated; **Figure 5A**). We observed that, when the reporter gene is driven by 5 U-*ZmPEPC1* fragment, ZmOrphan94 represses its activity (**Figure 5D**). However, when all ZmOrphan94 binding sites are mutated in the 5 U-*ZmPEPC1* sequence, the ZmOrphan94 repression activity is impaired (**Figure 5D**). This shows that indeed ZmOrphan94 binds *in vivo* to the CACA elements present in the 5 U-*ZmPEPC1* sequence and that this binding is essential for its function as transcriptional repressor.

Altogether, our results showed that ZmOrphan94 binds specifically to multiple 5'-CACA-3 sequences within the 5 U-*ZmPEPC1* upstream region, being this binding crucial for its function as repressor. Furthermore, we showed that one of the ZmOrphan94 binding sites within the 5 U-*ZmPEPC1* overlaps with the DNA-binding site of ZmbHLH90. This suggests a possible competition of ZmbHLH90 and ZmOrphan94 to the same binding site within the *ZmPEPC1* upstream region. However, the fact that ZmOrphan94 also binds to motifs in close proximity and that ZmOrphan94 and ZmbHLH90 proteins can interact may also underlie the observed impairment of the ZmbHLH90-mediated *ZmPEPC1* activation by ZmOrphan94.

ZmOrphan94 Rice Homologue Does Not Bind to the *ZmPEPC1* Upstream Region

In our previous studies, we have shown that the rice TF OsbHLH112 and its two maize homologues, ZmbHLH80 and ZmbHLH90, bind to the same *ZmPEPC1* upstream region (Górska et al., 2019). Given that C_4 plants evolved from the C_3, we proposed that these ZmbHLHs were co-opted during

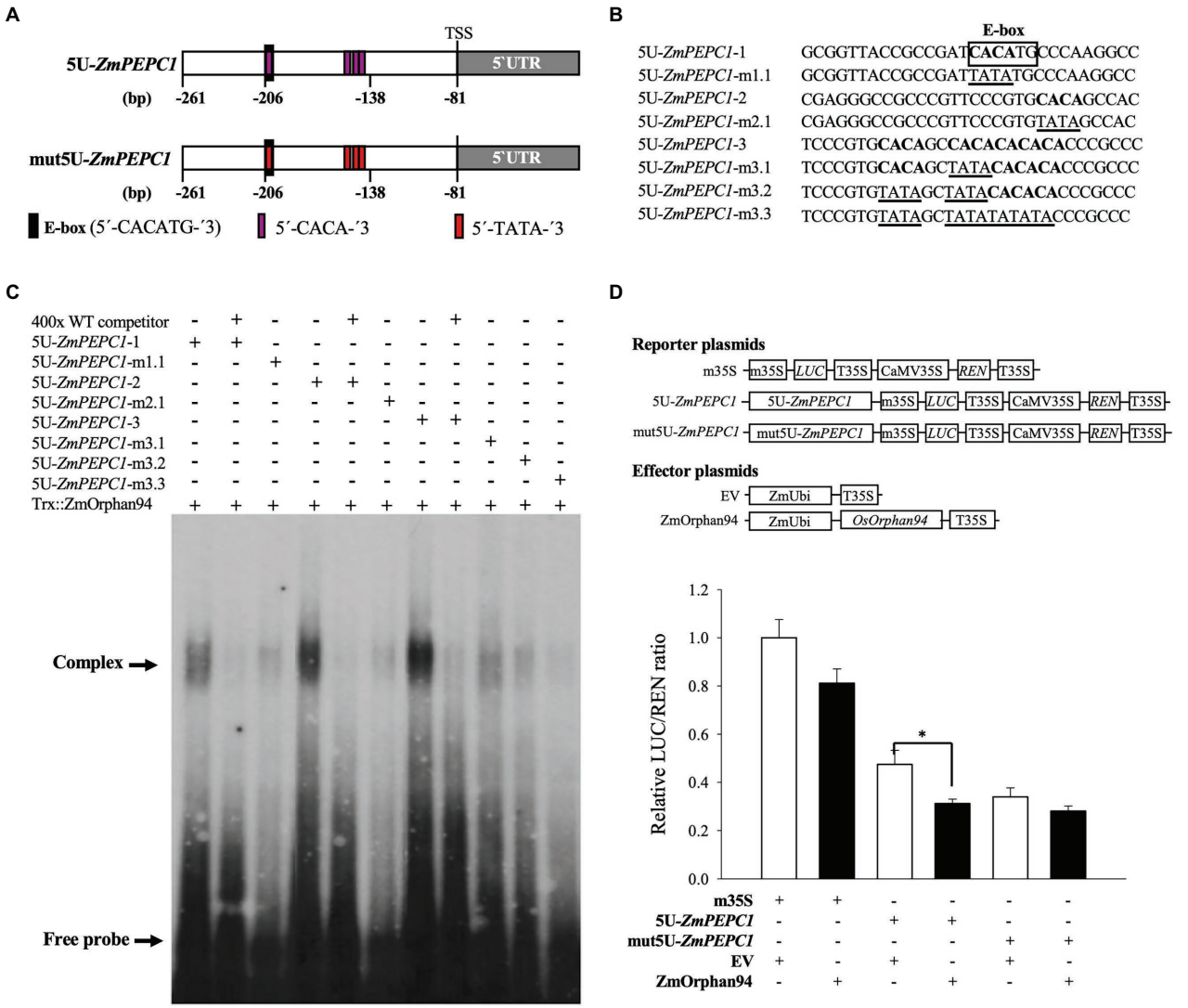

FIGURE 5 | Analysis of the interactions between ZmOrphan94 and the *cis*-elements present in the 5 U *ZmPEPC1* fragment. **(A)** Schematic representation of the 5 U fragment (within *ZmPEPC1* upstream region) indicating the relative positions of the E-box (ZmbHLH90 binding site; black rectangle) and 5'-CACA-3' (ZmOrphan94 binding site; magenta rectangle) sequences, as well as its mutated form (mut5U-*ZmPEPC1*). mut5U-*ZmPEPC1* fragment has the ZmOrphan94 binding sites mutated from 5'-CACA-3' to 5'-TATA-3'. Nucleotide numbers refer to the *ZmPEPC1* translational start codon ATG (+1). **(B)** Nucleotide sequences of DNA probes used in electromobility shift assays (EMSA). Putative ZmOrphan94 binding sites (5'-CACA-3') are represented in bold. The CACA sequences mutated to TATA are underlined. The binding site for ZmbHLH90 (E-box) is indicated. **(C)** EMSA of the Trx::ZmOrphan94 with probes derived from the 5 U *ZmPEPC1* upstream region. The "+" and "−" indicate presence and absence of respective protein or probe. The "400x" indicates a 400-time excess of unlabeled wild-type (WT) probe. **(D)** Trans-activation assay to analyze the function of the ZmOrphan94 binding sites in the 5 U-*ZmPEPC1* fragment. Data represent means ± SEM (n = 10–12). Differences are statistically significant (t-test, $^*p < 0.05$).

evolution of C_4 photosynthesis. To investigate whether ZmOrphan94 was also recruited from C_3 plants, we first searched for a ZmOrphan94 rice homologue and checked whether this homologue binds to the *ZmPEPC1* upstream region. To identify a rice homologue of ZmOrphan94, BLASTp search of the *O. sativa* genome with the ZmOrphan94 protein sequence was performed. This search identified rice OsOrphan65 (LOC_Os05g51690.1) as the best hit, showing 76.4% amino acid identity to the ZmOrphan94 protein sequence. OsOrphan65 full-length CDS was cloned into pDEST22 to be in fusion with the GAL4 activation domain (AD) and a direct Y1H assay was performed. The 5 U-*ZmPEPC1* bait was transformed with OsOrphan65, EV, and ZmbHLH90 (positive control) and the growth of the transformed bait strain was analyzed on a CM/-Trp/-His selection medium supplemented with increasing concentrations of 3-AT. Our results showed that OsOrphan65 does not interact with the 5 U-*ZmPEPC1* fragment. The growth of the 5 U-*ZmPEPC1* bait strain transformed with OsOrphan65 or EV was repressed on a CM/-Trp/-His selection media supplemented with 5 mM 3-AT (**Supplementary Figure S3**).

On the other hand, the 5 U-*ZmPEPC1* bait strain transformed with ZmbHLH90 (positive control) grew on CM/-Trp/-His + 20 mM 3-AT (**Supplementary Figure S3**).

ZmOrphan94 Is Preferentially Expressed in Maize BS Cells

M-specific expression of *ZmPEPC1* is known to be regulated at transcriptional level, activated in M, and repressed in BS cells (Kausch et al., 2001). Our results indicate that ZmOrphan94 is a *ZmPEPC1* transcriptional repressor that binds to the *ZmPEPC1* upstream region involved in the cell-specific *ZmPEPC1* gene expression. To better understand the role of ZmOrphan94 in the regulation of *ZmPEPC1* gene expression, we analyzed *ZmOrphan94* transcript levels in M and BS cells. For this, we purified M and BS cells from fully expanded maize third leaves at several time points, in the dark (−6 and −2 h), at the transition between light and dark (0 h) and during the light (+2 h, **Figure 6A**). Purity of isolated M and BS cells was assessed using the cell-specific gene markers, *ZmPEPC1* for M and *ZmNADP-ME* for BS cells (**Supplementary Figure S4**). As shown in **Figure 6B**, *ZmOrphan94* shows a much higher transcript level in BS than in M cells for all the time points analyzed. The higher expression of *ZmOrphan94* in BS cells, as compared with M cells, is particularly striking 2 h before dawn, and at dawn, when *ZmOrphan94* transcript level is, respectively, 4.75 and 3.9 times higher in BS than in M cells. The smallest difference in *ZmOrphan94* transcript accumulation between the two cell types is observed at 6 h before dawn. Nevertheless, at this time point, *ZmOrphan94* is approximately two times more expressed in BS than in M cells (**Figure 6B**).

DISCUSSION

The Role of ZmHB87 and ZmCPP8 in the Regulation of *ZmPEPC1* Gene Expression

In this study, we have identified three novel TFs binding to the *ZmPEPC1* upstream region (ZmOrphan94, ZmHB87, and ZmCPP8). However, despite our efforts, the role of ZmHB87 and ZmCPP8 on the regulation of *ZmPEPC1* gene expression is still not clear and requires further studies. ZmHB87 is a member of the homeobox TF family reported to form homo- and heter-odimers (Meijer et al., 2000; Mukherjee and Brocchieri, 2010). Nevertheless, under our experimental conditions, ZmHB87 did not form homodimers, did not interact with the other TFs analyzed, and did not show any trans-activation activity on the *ZmPEPC1* upstream region. In addition, *in silico* analysis of ZmHB87 amino acid sequence using TargetP (Emanuelsson et al., 2000) revealed that besides the NLS, ZmHB87 also carries a chloroplast transit peptide (cTP) localized at N-terminus of the protein (**Supplementary Figure S1B**). Overall, our results suggest that ZmHB87 may not regulate *ZmPEPC1* gene expression.

ZmCPP8 is a member of the cysteine-rich polycomb-like protein (CPP) TF family. Members of this family contain a highly conserved cysteine rich domain (CXC) within their DNA binding motif (Hauser et al., 2000; Schmit et al., 2009). CPPs are known to be involved in different processes, such as female and male sterility in *Arabidopsis* (Andersen et al., 2007), salt tolerance in rice (Almeida et al., 2016), and root nodule formation in soybean (Cvitanich et al., 2000). According to our results, ZmCPP8 may also be involved in the *ZmPEPC1* regulation in maize. ZmCPP8 was identified by Y1H as binding to the *ZmPEPC1* upstream region. Nevertheless, the addition of 5 mM 3-AT was enough to impair the ZmCPP8 binding to the F2-*ZmPEPC1* fragment, suggesting a weak interaction. In addition, the trans-activation assays to test ZmCPP8 activity showed a stronger activation of the US reporter (control) as compared to the F2-*ZmPEPC1* reporter, to which ZmCPP8 binds specifically. Thus, to determine whether ZmCPP8 is indeed involved in the *ZmPEPC1* regulation, further work is needed. First, it is essential to identify the *cis*-regulatory elements within the *ZmPEPC1* upstream region where ZmCPP8 can bind. LIN54 is perhaps the best characterized member of the CPP family and it was shown as binding to the *cis*-elements CDE (5'-TAGCGCGGT-3') and CHR (5'-TTYRAA-3', where Y is a pyrimidine and R is a purine; Schmit et al., 2009; Marceau et al., 2016). *In silico* analysis of the F2-*ZmPEPC1* sequence revealed the presence of CDE and CHR resembling motifs (**Supplementary Table S1**). Therefore, the next step would be to determine whether ZmCPP8 binds to these elements *in vitro* and *in vivo*. Recently, members of the maize CPP family have been characterized in terms of their response to abiotic stresses (Song et al., 2016). *ZmCPP8* (named *ZmCPP11* in that study) was upregulated in response to cold and heat treatment. The expression of *ZmCPP8* was strongly induced after 12 h of cold (4°C) and heat (42°C) treatments (Song et al., 2016). These findings suggest a putative role of ZmCPP8 in response to cold and heat stress in maize. Given that *PEPC1* transcript level is upregulated under cold stress (Li et al., 2019), it would be interesting to determine whether ZmCPP8 mediates the effect of cold on *ZmPEPC1* gene expression.

ZmOrphan94 May Play an Important Role in *ZmPEPC1* Regulatory Network

Our data indicate that ZmOrphan94 forms multiple heterodimers with other TFs binding to the *ZmPEPC1* upstream region. Interactions of ZmOrphan94 with other proteins occur likely through the CCT domain shown to mediate protein-protein interactions (Kurup et al., 2000; Wenkel et al., 2006). ZmOrphan94 forms homodimers, suggesting intra-family interactions, but also interacts with TFs from other TF families (e.g., CPP and bHLH). Even though intra-family interactions are most common and have been extensively studied (Amoutzias et al., 2008), the importance of cross-family interactions in a global regulatory network is well known (Bemer et al., 2017). The ability to interact with various TFs belonging to different TF families suggests that ZmOrphan94 plays an important role in the transcriptional network regulating *ZmPEPC1* gene expression, thus increasing its complexity and flexibility. There are several possible ways on how TF dimerization may affect TF activity. First, a TF complex may increase or reduce binding affinity of individual TFs to DNA. For example, in *Arabidopsis thaliana* a complex of auxin-response factor (ARF6) with PIF4/BZR1 increased DNA-binding affinity to ARF6/PIF4/BZR1 common targets, whereas decreased affinity for ARF6 specific

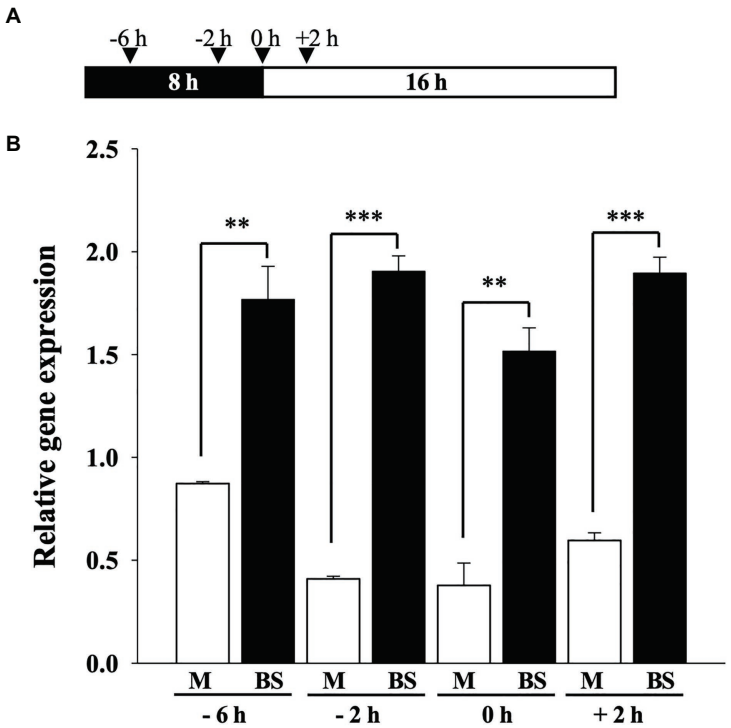

FIGURE 6 | Analysis of *ZmOrphan94* transcript level in mesophyll (M) and bundle sheath (BS) cells. **(A)** Schematic representation of the sampling time points. Black and white box indicate dark and light period, respectively. Arrowheads indicate sampling points. **(B)** *ZmOrphan94* expression in BS and M cells in the analyzed time points. *ZmOrphan94* transcript levels were analyzed by RT-qPCR and normalized against the expression of two housekeeping genes (GRMZM2G144843 and GRMZM2G044552). Data represent means ± SEM ($n = 3$). Statistical significance (*t*-test, $^{**}p < 0.01$ and $^{***}p < 0.001$).

targets (Oh et al., 2014). The effect of heterodimer formation on DNA binding specificity was also reported for MADS-domain containing proteins. The distinct binding preferences of MADS-box SEPALLATA3 (SEP3) and AGAMOUS (AG) TF heterodimer are involved in specification of reproductive organs in *Arabidopsis* (Smaczniak et al., 2017). TF-TF interactions may also affect the trans-activation activity of the individual TFs. For example, in sweet potato, bHLH heterodimerization is involved in plant defense against herbivory (Chen et al., 2016). In response to wounding, the TF IbbHLH3 activates the defense network through binding to the promoter and activation of IbNAC1 gene expression (Chen et al., 2016). To terminate the response, the IbbHLH3-IbbHLH4 heterodimer, which downregulates IbNAC1 expression, competes with IbbHLH3 homodimer for binding to the IbNAC1 promoter (Chen et al., 2016). Different transcriptional activity of TF-TF heterodimer, comparing to individual TFs, was also reported for the TFs involved in plant response to water deficit and osmotic stress conditions. ANAC096, a NAC (for NAM, ATAF1/2, and CUC2) TF, and ABF2, a bZIP-type TF, interact and are activators of *RD29A* gene expression. When co-transformed, ANAC096 and ABF2 act synergistically to activate *RD29A* expression (Xu et al., 2013). According to our results, ZmOrphan94 acts as a transcriptional repressor and co-expression of ZmOrphan94 with the identified activators, ZmCPP8 and ZmbHLH90, reduces the transactivation activity of the individual TFs on the *ZmPEPC1* promoter. Nevertheless, our transactivation experimental setup did not allow to determine whether this effect is due to competition for the same *cis*-element(s) or to heterodimerization, and its effect on DNA-binding or on TF trans-activation activity. Therefore, to understand in more detail the effect of ZmOrphan94 on the trans-activation activity of ZmCPP8 and ZmbHLH90, further work is needed.

ZmOrphan94 Is Part of a Regulatory Mechanism That Downregulates *ZmPEPC1* Gene Expression in Bundle Sheath Cells

Our results clearly show that ZmOrphan94 is part of the transcriptional network regulating *ZmPEPC1* gene expression. Based on our findings, we propose that together with the previously identified TFs, ZmbHLH80 and ZmbHLH90, ZmOrphan94 contributes to the M-specific *ZmPEPC1* gene expression (**Figure 7**). According to our results, ZmOrphan94 regulates *ZmPEPC1* transcript level in a similar manner as ZmbHLH80 (Górska et al., 2019). ZmOrphan94 acts as a repressor and impairs ZmbHLH90-mediated *ZmPEPC1* activation. This impairment is clear but may occur due to different reasons: (a) ZmOrphan94 and ZmbHLH90 form heterodimers, thus impairing ZmbHLH90 binding and/or ZmbHLH90 trans-activation activity; (b) ZmOrphan94 and

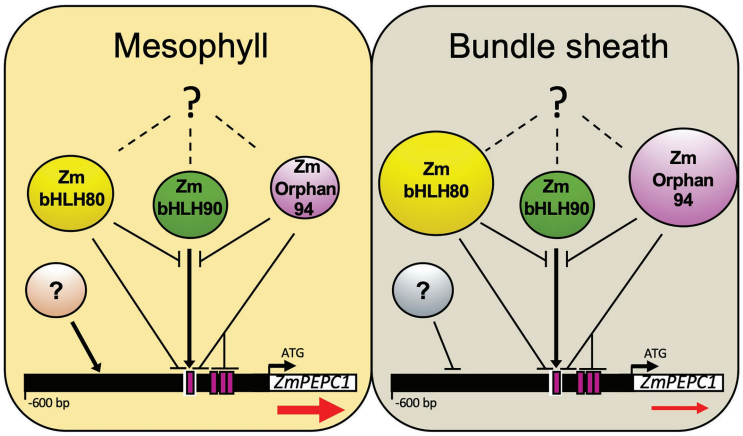

FIGURE 7 | Model integrating the contribution of ZmOrphan94, ZmbHLH80, and ZmbHLH90 to regulate M-specific *ZmPEPC1* expression. ZmOrphan94 and ZmbHLH80 are repressors, whereas ZmbHLH90 is an activator of *ZmPEPC1* expression. ZmOrphan94 and ZmbHLH80 impair ZmbHLH90-mediated *ZmPEPC1* activation through competitive binding to the same *cis*-element (E-Box; white rectangle) and/or heterodimerization. ZmOrphan94 can also impair ZmbHLH90-mediated *ZmPEPC1* activation through its binding to the CACA motifs (magenta rectangles) present within the *ZmPEPC1* upstream region, being that one of them is present within the E-Box (white rectangle). *ZmOrphan94* and *ZmbHLH80* are preferentially expressed in BS cells, thus having a predominant role repressing *ZmPEPC1* expression in this cell type. Dashed lines with question marks represent mechanisms that may regulate ZmOrphan94, ZmbHLH80, and ZmbHLH90 post-transcriptionally. Question marks in circles represent yet unidentified TFs that may be involved in M-specific *ZmPEPC1* expression. Arrows and horizontal lines indicate activation and repression, respectively. Sizes of ZmOrphan94, ZmbHLH80, and ZmbHLH90 circles represent differences in transcript abundance. Red arrows below *ZmPEPC1* gene indicate the difference in *ZmPEPC1* transcript abundance between M and BS cells.

ZmbHLH90 compete to the same *cis*-element; (c) ZmOrphan94 and ZmbHLH90 bind to *cis*-elements in close proximity and their interaction weakens ZmbHLH90 trans-activation activity; and (d) ZmOrphan94 has a repressor activity that overcomes ZmbHLH90 activator activity. Given that *ZmOrphan94* shows higher transcript levels in BS cells, as compared to M cells, we propose that together with BS-preferentially expressed ZmbHLH80, ZmOrphan94 plays a role in maintaining the *ZmPEPC*1 transcript levels low in BS cells (**Figure 7**). In maize, cell-specific downregulation has been reported for the *RbcS* gene family encoding RuBisCO small subunit. *RbcS* genes become BS-specific upon illumination (Sheen and Bogorad, 1986, 1987) and sequences within the upstream region and 3'UTR fragment of *RbcS-m3* have been involved in this expression pattern (Viret et al., 1994). Moreover, it was proposed that a TF belonging to the Krüppel-type zinc finger family, TRANSCRIPTION REPRESSOR MAIZE 1 (ZmTTM1), may be involved in this regulation. Though *ZmTTM1* expression, unlike *ZmOrphan94*, is not more abundant in a given cell type, mutations of any of its binding sites within *RbcS-m3* gene eliminate the repression of the *RBCS-m3* reporter gene in M cells (Xu et al., 2001). Interestingly, ZmOrphan94, ZmbHLH80, and ZmbHLH90 bind to the *ZmPEPC1* upstream region, within the conserved nucleotide sequences (CNSs) motifs recently identified by Gupta et al. (2020). ZmOrphan94 binds within CNS-1 and CNS-3B motifs, whereas ZmbHLH80 and ZmbHLH90 bind to the CNS-1. The CNS motifs are conserved among the C_4 *PEPC* genes from the Panicoid clade and are essential for driving M-cell specific gene expression in rice (Gupta et al., 2020), highlighting the importance of these CNS motifs as well as their binding TFs for the C_4 *PEPC* cell-specific gene expression. The current attempts to engineer the C_4 metabolism into rice (C_3 plant) require cell-specific accumulation of C_4 enzymes. To successfully accomplish this ambitious goal, we need to identify and characterize the function of the different *cis*-regulatory elements as well as the binding TFs and the molecular mechanisms underlying this feature.

Taken together, our data reveal the importance of the regulatory mechanisms within BS cells that contribute to the M-specific *ZmPEPC1* gene expression. We show that at least two TFs, ZmOrphan94 and ZmbHLH80, act to suppress *ZmPEPC1* gene expression in BS cells. It is likely that ZmbHLH80 was co-opted from the ancestral C_3 pathway, whereas ZmOrphan94 was recruited during evolution of C_4 photosynthesis.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

AG and PG conceived the project, performed the experiments, analyzed the data, and wrote the manuscript. AB, AZ, TS, PC, and TL performed the experiments and analyzed the data. MO conceived the project and revised the manuscript. CP conceived the project and raised funds. NS conceived the project, raised funds, analyzed the data, and revised the

manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the European Union project 3to4 (grant agreement no: 289582) and Research Unit GREEN-IT "Bioresources for Sustainability" (UID/Multi/04551/2013). FCT funded AG (SFRH/BD/89743/2012), AB (SFRH/BD/105739/2014), PC (PD/BD/128403/2017), TL (SFRH/BPD/102872/2014 and CEECIND/03641/2017), and NS (IF/01126/2012 – POPH-QREN).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.559967/full#supplementary-material

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. doi: 10.1105/tpc.107.053322

Almeida, D. M., Gregorio, G. B., Oliveira, M. M., and Saibo, N. J. M. (2016). Five novel transcription factors as potential regulators of *OsNHX1* gene expression in a salt tolerant rice genotype. *Plant Mol. Biol.* 93, 61–77. doi: 10.1007/s11103-016-0547-7

Amoutzias, G. D., Robertson, D. L., Van de Peer, Y., and Oliver, S. G. (2008). Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci.* 33, 220–229. doi: 10.1016/j.tibs.2008.02.002

Andersen, S. U., Algreen-Petersen, R. G., Hoedl, M., Jurkiewicz, A., Cvitanich, C., Braunschweig, U., et al. (2007). The conserved cysteine-rich domain of a tesmin/TSO1-like protein binds zinc in vitro and TSO1 is required for both male and female fertility in *Arabidopsis thaliana*. *J. Exp. Bot.* 58, 3657–3670. doi: 10.1093/jxb/erm215

Aubry, S., Brown, N. J., and Hibberd, J. M. (2011). The role of proteins in C3 plants prior to their recruitment into the C4 pathway. *J. Exp. Bot.* 62, 3049–3059. doi: 10.1093/jxb/err012

Bauwe, H., Hagemann, M., and Fernie, A. R. (2010). Photorespiration: players, partners and origin. *Trends Plant Sci.* 15, 330–336.

Bemer, M., van Dijk, A. D. J., Immink, R. G. H., and Angenent, G. C. (2017). Cross-family transcription factor interactions: an additional layer of gene regulation. *Trends Plant Sci.* 22, 66–80. doi: 10.1016/j.tplants.2016.10.007

Borba, A. R., Serra, T. S., Górska, A., Gouveia, P., Cordeiro, A. M., Reyna-Llorens, I., et al. (2018). Synergistic binding of bHLH transcription factors to the promoter of the maize NADP-ME gene used in C4 photosynthesis is based on an ancient code found in the ancestral C3 state. *Mol. Biol. Evol.* 35, 1690–1705. doi: 10.1093/molbev/msy060

Brown, N. J., Newell, C. A., Stanley, S., Chen, J. E., Perrin, A. J., Kajala, K., et al. (2011). Independent and parallel recruitment of preexisting mechanisms underlying C_4 photosynthesis. *Science* 331, 1436–1439. doi: 10.1126/science.1201248

Cavalar, M., Phlippen, Y., Kreuzaler, F., and Peterhänsel, C. (2007). A drastic reduction in DOF1 transcript levels does not affect C4-specific gene expression in maize. *J. Plant Physiol.* 164, 1665–1674. doi: 10.1016/j.jplph.2006.09.008

Chen, S. P., Kuo, C. H., Lu, H. H., Lo, H. S., and Yeh, K. W. (2016). The sweet potato NAC-domain transcription factor IbNAC1 is dynamically coordinated by the activator IbbHLH3 and the repressor IbbHLH4 to reprogram the defense mechanism against wounding. *PLoS Genet.* 12:e1006397. doi: 10.1371/journal.pgen.1006397

Cordeiro, A. M., Figueiredo, D. D., Tepperman, J., Borba, A. R., Lourenço, T., Abreu, I. A., et al. (2016). Rice phytochrome-interacting factor protein OsPIF14 represses *OsDREB1B* gene expression through an extended N-box and interacts preferentially with the active form of phytochrome B. *Biochim. Biophys. Acta, Gene Regul. Mech.* 1859, 393–404. doi: 10.1016/j.bbagrm.2015.12.008

Covshoff, S., Furbank, R. T., Leegood, R. C., and Hibberd, J. M. (2013). Leaf rolling allows quantification of mRNA abundance in mesophyll cells of sorghum. *J. Exp. Bot.* 64, 807–813. doi: 10.1093/jxb/ers286

Cvitanich, C., Pallisgaard, N., Nielsen, K. A., Hansen, A. C., Larsen, K., Pihakaski-Maunsbach, K., et al. (2000). CPP1, a DNA-binding protein involved in the expression of a soybean leghemoglobin C3 gene. *Proc. Natl. Acad. Sci. U. S. A.* 97, 8163–8168. doi: 10.1073/pnas.090468497

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016. doi: 10.1006/jmbi.2000.3903

Figueiredo, D. D., Barros, P., Cordeiro, A. M., Serra, T., Lourenço, T., Subhash, C., et al. (2012). Seven zinc-finger transcription factors are novel regulators of the stress responsive gene *OsDREB1B*. *J. Exp. Bot.* 63, 695–709. doi: 10.1093/jxb/ers035

Furbank, R. T., and Taylor, W. C. (1995). Regulation of photosynthesis in C3 and C4 plants: a molecular approach. *Plant Cell* 7, 797–807. doi: 10.2307/3870037

Gendron, J. M., Pruneda-Paz, J. L., Doherty, C. J., Gross, A. M., Kang, S. E., and Kay, S. A. (2012). *Arabidopsis* circadian clock protein, TOC1, is a DNA-binding transcription factor. *Proc. Natl. Acad. Sci.* 109, 3167–3172. doi: 10.1073/pnas.1200355109

Górska, A. M., Gouveia, P., Borba, A. R., Zimmermann, A., Serra, T. S., Lourenco, T. F., et al. (2019). ZmbHLH80 and ZmbHLH90 transcription factors act antagonistically and contribute to regulate *PEPC1* cell-specific gene expression in maize. *Plant J.* 99, 270–285. doi: 10.1111/tpj.14323

Gowik, U., Burscheidt, J., Akyildiz, M., Schlue, U., Koczor, M., Streubel, M., et al. (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. doi: 10.1105/tpc.019729

Gowik, U., Schulze, S., Saladié, M., Rolland, V., Tanz, S. K., Westhoff, P., et al. (2016). A MEM1-like motif directs mesophyll cell-specific expression of the gene encoding the C4 carbonic anhydrase in Flaveria. *J. Exp. Bot.* 68, 311–320. doi: 10.1093/jxb/erw475

Gupta, S. D., Levey, M., Schulze, S., Karki, S., Emmerling, J., Streubel, M., et al. (2020). The C4Ppc promoters of many C4 grass species share a common regulatory mechanism for gene expression in the mesophyll cell. *Plant J.* 101, 204–216. doi: 10.1111/tpj.14532

Hauser, B. A., He, J. Q., Park, S. O., and Gasser, C. S. (2000). TSO1 is a novel protein that modulates cytokinesis and cell expansion in *Arabidopsis*. *Development* 127, 2219–2226.

Hellens, R. P., Allan, A. C., Friel, E. N., Bolitho, K., Grafton, K., Templeton, M. D., et al. (2005). Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. *Plant Methods* 1:13. doi: 10.1186/1746-4811-1-13

Hibberd, J. M., and Covshoff, S. (2010). The regulation of gene expression required for C4 photosynthesis. *Annu. Rev. Plant Biol.* 61, 181–207. doi: 10.1146/annurev-arplant-042809-112238

Kajala, K., Brown, N. J., Williams, B. P., Borrill, P., Taylor, L. E., and Hibberd, J. M. (2012). Multiple *Arabidopsis* genes primed for recruitment into C4 photosynthesis. *Plant J.* 69, 47–56. doi: 10.1111/j.1365-313X.2011.04769.x

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. doi: 10.1007/BF00269849

Kausch, A. P., Owen, T. P., 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. doi: 10.1023/A:1006487326533

Kurup, S., Jones, H. D., and Holdsworth, M. J. (2000). Interactions of the developmental regulator ABI3 with proteins identified from developing *Arabidopsis* seeds. *Plant J.* 21, 143–155. doi: 10.1046/j.1365-313x.2000.00663.x

Li, M., Sui, N., Lin, L., Yang, Z., and Zhang, Y. (2019). Transcript profiling revealed genes involved in response to cold stress in maize. *Funct. Plant Biol.* 46, 830–844. doi: 10.1071/FP19065

Marceau, A. H., Felthousen, J. G., Goetsch, P. D., Iness, A. N., Lee, H. W., Tripathi, S. M., et al. (2016). Structural basis for LIN54 recognition of CHR elements in cell cycle-regulated promoters. *Nat. Commun.* 7, 1–11. doi: 10.1038/ncomms12301

Markelz, N. H., Costich, D. E., and Brutnell, T. P. (2003). Photomorphogenic responses in maize seedling development. *Plant Physiol.* 133, 1578–1591. doi: 10.1104/pp.103.029694

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. doi: 10.1046/j.1365-313X.1994.06030311.x

Meijer, A. H., de Kam, R. J., D'Erfurth, I., Shen, W., and Hoge, J. H. C. (2000). HD-zip proteins of families I and II from rice: interactions and functional properties. *Mol. Gen. Genet.* 263, 12–21. doi: 10.1007/PL00008671

Mukherjee, K., and Brocchieri, L. (2010). Evolution of plant homeobox genes. *eLS* 2, 31–45. doi: 10.1002/9780470015902.a0022865

Nomura, M., Sentoku, N., Nishimura, A., Lin, J. H., Honda, C., Taniguchi, M., et al. (2000). The evolution of C4 plants: acquisition of cis-regulatory sequences in the promoter of C4-type pyruvate, orthophosphate dikinase gene. *Plant J.* 22, 211–221. doi: 10.1046/j.1365-313x.2000.00726.x

Oh, E., Zhu, J. Y., Bai, M. Y., Arenhart, R. A., Sun, Y., and Wang, Z. Y. (2014). Cell elongation is regulated through a central circuit of interacting transcription factors in the *Arabidopsis* hypocotyl. *eLife* 3, 1–19. doi: 10.7554/eLife.03031

Ouwerkerk, P. B., and Meijer, A. H. (2001). Yeast one-hybrid screening for DNA-protein interactions. *Curr. Protoc. Mol. Biol.* 55, 12.12.1–12.12.12. doi: 10.1002/0471142727.mb1212s55

Patel, M., Siegel, A. J., and Berry, J. O. (2006). Untranslated regions of FbRbcS1 mRNA mediate bundle sheath cell-specific gene expression in leaves of a C4 plant. *J. Biol. Chem.* 281, 25485–25491. doi: 10.1074/jbc.M604162200

Portis, A. R., and Parry, M. A. J. (2007). Discoveries in Rubisco (Ribulose 1,5-bisphosphate carboxylase/oxygenase): a historical perspective. *Photosynth. Res.* 94, 121–143. doi: 10.1007/s11120-007-9225-6

Reeves, G., Grangé-Guermente, M. J., and Hibberd, J. M. (2017). Regulatory gateways for cell-specific gene expression in C4 leaves with Kranz anatomy. *J. Exp. Bot.* 68, 107–116. doi: 10.1093/jxb/erw438

Reyna-Llorens, I., Burgess, S. J., Reeves, G., Singh, P., Stevenson, S. R., Williams, B. P., et al. (2018). Ancient duons may underpin spatial patterning of gene expression in C4 leaves. *Proc. Natl. Acad. Sci.* 115, 1931–1936. doi: 10.1073/pnas.1720576115

Sage, R. F. (2004). The evolution of C4 photosynthesis. *New Phytol.* 161, 341–370. doi: 10.1111/j.1469-8137.2004.00974.x

Schmit, F., Cremer, S., and Gaubatz, S. (2009). LIN54 is an essential core subunit of the DREAM/LINC complex that binds to the *cdc2* promoter in a sequence-specific manner. *FEBS J.* 276, 5703–5716. doi: 10.1111/j.1742-4658.2009.07261.x

Serra, T. S., Figueiredo, D. D., Cordeiro, A. M., Almeida, D. M., Lourenço, T., Abreu, I. A., et al. (2013). OsRMC, a negative regulator of salt stress response in rice, is regulated by two AP2/ERF transcription factors. *Plant Mol. Biol.* 82, 439–455. doi: 10.1007/s11103-013-0073-9

Sheen, J. (1991). Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes. *Plant Cell* 3, 225–245. doi: 10.1105/tpc.3.3.225

Sheen, J. -Y., and Bogorad, L. (1986). Expression of the ribulose-1, 5-bisphosphate carboxylase large subunit gene and three small subunit genes in two cell types of maize leaves. *EMBO J.* 5:3417. doi: 10.1002/j.1460-2075.1986.tb04663.x

Sheen, J. Y., and Bogorad, L. (1987). Differential expression of C4 pathway genes in mesophyll and bundle sheath cells of greening maize leaves. *J. Biol. Chem.* 262, 11726–11730. doi: 10.1016/S0021-9258(18)60871-3

Smaczniak, C., Muiño, J. M., Chen, D., Angenent, G. C., and Kaufmann, K. (2017). Differences in DNA-binding specificity of floral homeotic protein complexes predict organ-specific target genes. *Plant Cell* 29, 1822–1835. doi: 10.1105/tpc.17.00145

Song, X. Y., Zhang, Y. Y., Wu, F. C., and Zhang, L. (2016). Genome-wide analysis of the maize (*Zea may* L.) CPP-like gene family and expression profiling under abiotic stress. *Genet. Mol. Res.* 15, 1–11. doi: 10.4238/gmr.15038023

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. doi: 10.1105/tpc.9.4.479

Taniguchi, M., Izawa, K., Ku, M. S. B., Lin, J. -H., Saito, H., Ishida, Y., et al. (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. doi: 10.1023/A:1026565027772

Viret, J. F., Mabrouk, Y., and Bogorad, L. (1994). Transcriptional photoregulation of cell-type-preferred expression of maize Rbcs-m3-3' and 5' sequences are involved. *Proc. Natl. Acad. Sci. U. S. A.* 91, 8577–8581. doi: 10.1073/pnas.91.18.8577

Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A., et al. (2006). CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. *Plant Cell* 18, 2971–2984. doi: 10.1105/tpc.106.043299

Williams, B. P., Burgess, S. J., Reyna-Llorens, I., Knerova, J., Aubry, S., Stanley, S., et al. (2016). An untranslated cis-element regulates the accumulation of multiple C4 enzymes in *Gynandropsis gynandra* mesophyll cells. *Plant Cell* 28, 454–465. doi: 10.1105/tpc.15.00570

Wiludda, C., Schulze, S., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., et al. (2012). Regulation of the photorespiratory GLDPA gene in C4 Flaveria: an intricate interplay of transcriptional and posttranscriptional processes. *Plant Cell* 24, 137–151. doi: 10.1105/tpc.111.093872

Xu, Z. -Y., Kim, S. Y., Hyeon, D. Y., Kim, D. H., Dong, T., Park, Y., et al. (2013). The *Arabidopsis* NAC transcription factor ANAC096 cooperates with bZIP-type transcription factors in dehydration and osmotic stress responses. *Plant Cell* 25, 4708–4724. doi: 10.1105/tpc.113.119099

Xu, T., Purcell, M., Zucchi, P., Helentjaris, T., and Bogorad, L. (2001). TRM1, a YY1-like suppressor of rbcS-m3 expression in maize mesophyll cells. *Proc. Natl. Acad. Sci.* 98, 2295–2300. doi: 10.1073/pnas.041610098

Yanagisawa, S., and Izui, K. (1990). Multiple interactions between tissue-specific nuclear proteins and the promoter of the phosphoenolpyruvate carboxylase gene for C4 photosynthesis in *Zea mays*. *Mol. Gen. Genet.* 224, 325–332. doi: 10.1007/BF00262425

Yanagisawa, S., and Izui, K. (1992). MNF1, a leaf tissue-specific DNA-binding protein of maize, interacts with the cauliflower mosaic virus 35S promoter as well as the C4 photosynthetic phosphoenolpyruvate carboxylase gene promoter. *Plant Mol. Biol.* 19, 545–553. doi: 10.1007/BF00026781

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. doi: 10.1105/tpc.10.1.75

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 30 August 2021
doi: 10.3389/fpls.2021.665279

Genome-Wide Identification and Analysis of the Phosphoenolpyruvate Carboxylase Gene Family in *Suaeda aralocaspica*, an Annual Halophyte With Single-Cellular C_4 Anatomy

*Jing Cao†, Gang Cheng†, Lu Wang, Tayier Maimaitijiang and Haiyan Lan**

Xinjiang Key Laboratory of Biological Resources and Genetic Engineering, College of Life Science and Technology, Xinjiang University, Urumqi, China

OPEN ACCESS

Edited by:
Sarah Covshoff, Independent Researcher, Las Vegas, NV, United States

Reviewed by:
Steven James Burgess, University of Illinois Urbana-Champaign, United States
Jacob Washburn, Plant Genetics Research Unit, USDA-ARS, United States

***Correspondence:**
Haiyan Lan
lanhaiyan@xju.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Physiology, a section of the journal Frontiers in Plant Science

Received: *04 March 2021*
Accepted: *09 July 2021*
Published: *30 August 2021*

Citation:
Cao J, Cheng G, Wang L, Maimaitijiang T and Lan H (2021) Genome-Wide Identification and Analysis of the Phosphoenolpyruvate Carboxylase Gene Family in Suaeda aralocaspica, an Annual Halophyte With Single-Cellular C_4 Anatomy. Front. Plant Sci. 12:665279. doi: 10.3389/fpls.2021.665279

Phosphoenolpyruvate carboxylase (PEPC) plays pivotal roles in the carbon fixation of photosynthesis and a variety of metabolic and stress pathways. *Suaeda aralocaspica* belongs to a single-cellular C_4 species and carries out a photosynthetic pathway in an unusually elongated chlorenchyma cell, which is expected to have PEPCs with different characteristics. To identify the different isoforms of *PEPC* genes in *S. aralocaspica* and comparatively analyze their expression and regulation patterns as well as the biochemical and enzymatic properties in this study, we characterized a bacterial-type PEPC (BTPC; SaPEPC-4) in addition to the two plant-type PEPCs (PTPCs; SaPEPC-1 and SaPEPC-2) using a genome-wide identification. *SaPEPC-4* presented a lower expression level in all test combinations with an unknown function; two SaPTPCs showed distinct subcellular localizations and different spatiotemporal expression patterns but positively responded to abiotic stresses. Compared to *SaPEPC-2,* the expression of *SaPEPC-1* specifically in chlorenchyma cell tissues was much more active with the progression of development and under various stresses, particularly sensitive to light, implying the involvement of *SaPEPC-1* in a C_4 photosynthetic pathway. In contrast, *SaPEPC-2* was more like a non-photosynthetic PEPC. The expression trends of two SaPTPCs in response to light, development, and abiotic stresses were also matched with the changes in PEPC activity *in vivo* (native) or *in vitro* (recombinant), and the biochemical properties of the two recombinant SaPTPCs were similar in response to various effectors while the catalytic efficiency, substrate affinity, and enzyme activity of SaPEPC-2 were higher than that of SaPEPC-1 *in vitro.* All the different properties between these two SaPTPCs might be involved in transcriptional (e.g., specific *cis*-elements), posttranscriptional [e.g., 5′-untranslated region (5′-UTR) secondary structure], or translational (e.g., PEPC phosphorylation/dephosphorylation) regulatory events. The comparative studies on the different isoforms of the PEPC gene family in *S. aralocaspica* may help to decipher their exact role in C_4 photosynthesis, plant growth/development, and stress resistance.

Keywords: enzyme kinetics, genome-wide identification, PEPC, single-cellular C_4 anatomy, *Suaeda aralocaspica*, transcriptional expression

INTRODUCTION

Phosphoenolpyruvate carboxylase (PEPC, EC 4.1.1.31) is widely distributed in photosynthetic organisms such as vascular plants, algae, and photosynthetic bacteria. Plant PEPCs consist of a small gene family, which encodes several plant-type PEPCs (PTPCs) and at least one "distant relative" – bacterial-type PEPC (BTPC; Sánchez and Cejudo, 2003). All PTPCs encode approximately 100–110 kDa polypeptides with a conserved N-terminal seryl-phosphorylation domain and a distinguishing C-terminal tetrapeptide QNTG (Q-glutamine, N-asparagine, T-threonine, and G-glycine) signature (Izui et al., 2004; Xu et al., 2006). The 116–118 kDa-long polypeptides encoded by BTPCs show only about 40% identity with PTPC sequences, which harbor a prokaryotic-like (R/K) NTG (R-arginine, K-lysine) C-terminal tetrapeptide, but is lacking the N-terminal phosphorylatable Ser residue (Gennidakis et al., 2007; O'Leary et al., 2011a,b). With the completion of plant genome project and gene sequencing, numerous *PEPC* genes of different plant species have been identified, e.g., four *PEPCs* in *Arabidopsis* (Sánchez and Cejudo, 2003), five in tomato (Waseem and Ahmad, 2019), 10 in soybean (Wang et al., 2016), 5–9 in different peanut species (Yu et al., 2010; Pan et al., 2017; Tu et al., 2018), 6–11 among different cotton cultivars (Zhao et al., 2019), etc., however, limited information is available concerning *PEPC* from C_4 species without Kranz anatomy. So far, except the cases of our previously uploaded two *PEPC* complementary DNA (cDNA) sequences of *Suaeda aralocaspica* in the GenBank database (KP985714.1 and KX009562.1 for *SaPEPC-1* and *SaPEPC-2*, respectively), only partial cDNA sequences of some *PEPC* genes were isolated from three single-cell (SC) C_4 species (*S. aralocaspica*, *Bienertia cycloptera*, and *Bienertia sinuspersici*) (Lara et al., 2006; Rosnow et al., 2014). Currently, the introduction of the draft genome assembly of *S. aralocaspica* makes it possible to identify genome-wide *PEPC* genes in *S. aralocaspica* (Wang et al., 2019).

Suaeda aralocaspica (Bunge) Freitag and Schütze (Chenopodiaceae) is an annual halophyte, which is distributed in the southern margin of Junggar Basin in China, and is restricted to the saline–alkaline sandy soils of Gobi desert in central Asia (Commissione Redactorum Florae Xinjiangensis, 1994). It is the first terrestrial plant species discovered possessing the SC C_4 photosynthetic pathway, since then, three more SC C_4 species have been found in the genus *Bienertia* (Voznesenskaya et al., 2001; Sharpe and Offermann, 2014). In *S. aralocaspica*, the unusually long chlorenchyma cells are arranged in a single layer in the leaf, and the dimorphic chloroplasts have a spatially polar distribution between distal and proximal ends of the chlorenchyma cells, which is analogous to the Kranz anatomy but lacks the intervening cell wall (Edwards and Voznesenskaya, 2011). The key photosynthetic enzymes are biochemically compartmentalized in different regions of the cytoplasm (Voznesenskaya et al., 2004; for a review see Sharpe and Offermann, 2014), and it is speculated that the PEPCs might be different in types, enzymatic properties, and/or functions to those in Kranz C_4 species. PEPC carboxylation rate is apparently higher in *S. aralocaspica* than in Kranz C_4 species (Edwards et al., 2004; Smith et al., 2009; Liu et al., 2020); whereas some biochemical characteristics of PEPC in *S. aralocaspica* leaves are similar to Kranz C_4 plants (i.e., protein phosphorylation in response to light/dark and some aspects of enzymatic kinetics) (Lara et al., 2006). In our previous study, a full-length (FL) cDNA sequence of *PEPC* gene in *S. aralocaspica* was isolated and termed as *SaPEPC-1* (GenBank: KP985714.1) according to the classification of Rosnow et al. (2014), and suggested its roles in development and stress tolerance in *S. aralocaspica* (Cheng et al., 2016). Recently, we cloned another *PEPC* gene cDNA sequence from *S. aralocaspica* termed as *SaPEPC-2* (GenBank: KX009562.1), which share 76.2% similarity with *SaPEPC-1*; we also achieved the complete genomic DNA sequences of *SaPEPC-1* and *SaPEPC-2* (GenBank: KU870624 and KU870625). So far, these two PEPCs remained uncertain in terms of category (C_3 or C_4 type), regulatory characteristics, biochemical properties, and the enzymatic kinetics. To dissect these questions, the major aims of the present study are: (1) To characterize all members of *PEPC* genes in *S. aralocaspica* at the genome-wide level. (2) To comparatively analyze the functional differences among *SaPEPC* isoforms in phylogenetic relations, gene structure, protein motifs, subcellular localization, transcriptional regulation, and enzyme biochemistry. (3) To investigate the contribution of different types of SaPEPCs in response to the development, light/dark, and abiotic stresses, as well as their role in a C_4 photosynthetic pathway. The abovementioned goals needed to be achieved might help in further understanding of the roles of PEPC isoforms in C_4 species.

MATERIALS AND METHODS

Plant Materials and Cultivation

The mature seeds of *S. aralocaspica* were harvested from dry inflorescence of natural plants growing in the Gurbantunggut desert at Wujiaqu 103 regiment (44°37′N, 87°26′E; 423 mH) in October 2014, in the Xinjiang Uygur Autonomous Region, China. Seeds were air-dried indoor and cleaned and then stored at 4°C in a sealed brown paper bag. The brown seeds can germinate within 3 h upon contact with water, whereas the black seeds germinate much slower (Wang et al., 2008; He et al., 2013). However, the descendants from dimorphic seeds present no significant difference in morphological and physiological characteristics as well as gene expression patterns (Cao et al., 2015). Therefore, in this study, brown seeds were used in all the experiments.

Seed Germination and Treatments

To collect samples for total RNA extraction in seed germination, approximately 150 brown seeds were sown on the two layers of a filter paper in a 15 cm Petri dish, to which 20 ml of distilled water or other aqueous solutions were added. For different germination times, germinated seeds (seedlings) were harvested at 8 h, 12 h, 24 h, 2 days, 5 days, 10 days, and 15 days, respectively, and dry seeds at 0 h were used as control; for different tissues, cotyledons, hypocotyls, and radicles were sampled from the seedlings germinated for 7 days; and for different stress treatments, a filter paper was saturated with 20 ml of different concentrations

(conc.) of aqueous solutions: NaCl (100, 300, and 500 mmol·L^{-1}), isoosmotic mannitol (200, 600, and 1,000 mmol·L^{-1}), and ABA (1, 5, and 10 μmol·L^{-1}), respectively, only distilled water was used as control, and the seedlings were harvested after 7-day germination. All the above experiments were treated with normal light intensity (500 μmol·m^{-2}·s^{-1}) and under darkness (the Petri dish was properly wrapped with a foil to avoid light penetration). For different light treatments, all Petri dishes were subjected to a photoperiod of 16 h light/8 h dark with the light intensities of 30, 300, and 900 μmol·m^{-2}·s^{-1}, respectively, and the seedlings were harvested after 7-day germination.

Seedling Growth and Treatments

The brown seeds were sown in pots containing perlite: vermiculite (1:3, v/v) in a growth chamber, under the conditions of a 16 h light/8 h dark photoperiod with the light intensity of 500–700 μmol·m^{-2}·s^{-1}, a temperature regime of 24–30°C, and a relative humidity of 10–20%. The pot soil was carefully sprayed with distilled water or other aqueous solutions by a mini-sprinkler till the distilled water (solutions) was drained out from the bottom of the pot when the draining solution volume and the pot volume were approximately the same, a fresh distilled water (corresponding solutions) was transferred into the pot in a tray for 2 h to keep the soil saturated with distilled water (solutions) during initiation, consequently supplemented with the distilled water (corresponding solutions) at an interval of 1 week. Seedlings cultivated with distilled water were used as control [in addition to the half-strength Hoagland solution (Arnon and Hoagland, 1940) at an interval of 2–3 weeks]. For different tissues, the leaves, stems, and roots on day 15 (seedling) and day 90 (adult plant) after emergence were collected; for different developmental stages, whole seedlings (for gene expression analysis) or cotyledons/leaves (for PEPC activity measurement) were harvested on day 3, 15, 30, and 60, respectively, after emergence; for salt stress, seedlings were treated with the half-strength Hoagland solution containing 100, 300, or 500 mmol·L^{-1} NaCl and the cotyledons/leaves were harvested on day 15 after emergence; for drought stress, seedlings at 30 days of emergence were subjected to natural drought for 7, 14, and 28 days, respectively; and for different photoperiods, 60-day-old plants after emergence were cultivated in a greenhouse during the continual sunny days, on the 2nd or 3rd day, the leaves on the top of the plants were harvested at an interval of 2 h from the morning at 8:00 to the evening at 22:00 within the same day. **Supplementary Table 1** provides the detailed descriptions for different experimental designs and sampling times.

All samples were immediately frozen in liquid nitrogen on harvesting and then stored at −80°C until use. Four biological replicates were applied to each treatment.

Identification of *PEPC* Genes in *S. aralocaspica*

To identify the potential members of *PEPC* gene family in *S. aralocaspica* genome, firstly, the amino acid sequences of the four PEPCs in *Arabidopsis thaliana* were used as a query to conduct a local BLASTP search by a cut-off *E*-value of 1×10^{-5} (**Supplementary Table 2**). Subsequently, the Hidden Markov Model- (HMM-) based profile of the PEPCase domain (PF00311) obtained from the Pfam database[1] was used to verify the candidates of *PEPC* gene homologs by HMMER[2] and SMART[3] searches. Finally, the candidates of *PEPC* homologs were further validated in the presence of a PEPC family domain (IPR021135), a lysine active site (IPR018129), and a histidine active site (IPR033129) on the InterProScan website[4].

Multi-Sequence Alignment and Phylogenetic Analysis

Multiple alignments were performed by FL amino acid sequences using the ClustalW program of MEGA X with the default settings (Kumar et al., 2018). The phylogenetic tree of PEPC proteins from 27 plant species was constructed using the unrooted neighbor-joining method of MEGA X with the following parameters: Poisson correction, pairwise deletion, and a bootstrap analysis with 1,000 replicates. The amino acid sequences of the other 26 representative species (Monocots: *Brachypodium distachyon*, *Oryza sativa*, *Panicum virgatum*, *Setaria italica*, *Sorghum bicolor*, and *Zea mays*; Dicots: *A. thaliana*, *Arachis hypogaea*, *Brassica rapa*, *Chenopodium quinoa*, *Glycine max*, *Gossypium raimondii*, *Linum usitatissimum*, *Manihot esculenta*, *Medicago truncatula*, *Phaseolus vulgaris*, *Populus trichocarpa*, *Ricinus communis*, and *Solanum lycopersicum*; Pteridophytes: *Selaginella moellendorffii*; Bryophyte: *Physcomitrella patens*; Photosynthetic algae: *Chlamydomonas reinhardtii*, *Coccomyxa subellipsoidea*, *Micromonas pusilla*, *Ostreococcus lucimarinus*, and *Volvox carteri*) were acquired from the Phytozome database[5]. **Supplementary Table 2** provides a detailed description of the abovementioned proteins and their corresponding accession numbers.

Analyses of Gene Structures, Conserved Motifs, and *Cis*-Regulatory Elements

For gene structure analysis, the exons and introns of *PEPC* genes were identified due to the alignment of cDNA sequences with the corresponding genomic DNA sequences and were illustrated using the GSDS 2.0 server[6]. The MEME program[7] was employed to identify and analyze the conserved motifs of PEPC proteins with default parameters, and the maximum number of motifs to be detected was set as 10. The *cis*-regulatory elements in the promoter sequences (2,500 bp upstream of the start codon) of *PEPC* genes were identified using the PlantCARE database[8]. The Mfold RNA/DNA folding program[9] was used to predict the secondary structure of a 5′-untranslated region (5′-UTR) of *PEPC* genes. The MEME and

[1]http://pfam.xfam.org/
[2]http://hmmer.janelia.org/
[3]http://smart.embl-heidelberg.de/
[4]http://www.ebi.ac.uk/interpro/
[5]https://phytozome.jgi.doe.gov
[6]http://gsds.cbi.pku.edu.cn/
[7]http://meme-suite.org/tools/meme
[8]http://bioinformatics.psb.ugent.be/webtools/plantcare/html/
[9]http://www.unafold.org/

PlantCARE results were visualized using the TBtools software (Chen et al., 2020). Meanwhile, the theoretical molecular weight (*MW*), isoelectric point (*pI*), and grand average of hydropathicity (GRAVY) of PEPC candidates were predicted using the ExPASy website[10].

Determination of Subcellular Localization

Plant-mPLoc[11] and YLoc+[12] websites were used to predict the subcellular localization of candidate PEPCs in *S. aralocaspica*, which were further verified by a transient expression system in tobacco epidermal cells. The open reading frame (ORF) sequence of *SaPEPC-1* or *SaPEPC-2* (with the stop codon deletion) was fused to an enhanced green fluorescent protein (eGFP) ORF, and then inserted into the plant binary expression vector pCAMBIA1300, which resulted in the construct *35S::SaPEPCs-eGFP*. The primers used for vector construction are presented in **Supplementary Table 3**.

The abovementioned recombinant vectors were transformed into *Agrobacterium tumefaciens* strain GV3101 through a $CaCl_2$ method. The correct single colony was inoculated in a YEB medium (50 mg·L^{-1} kanamycin, 50 mg·L^{-1} gentamicin, and 50 mg·L^{-1} rifampicin) and cultivated with a shaking speed of 220 rpm at 28°C till the OD_{595} value reached the range of 0.8–1.0. Then, 2 ml of cultures were removed for centrifugation at 12,000 rpm for 2 min to collect cells, which were then resuspended in an infiltration buffer (10 mmol·L^{-1} MES, 10 mmol·L^{-1} $MgCl_2$, and 150 μmol·L^{-1} acetosyringone) at a final concentration of OD_{595} = 0.8. *A. tumefaciens* suspension (A) of the abovementioned constructs was evenly mixed with *35S::CBL-RFP*/GV3101 (B) [Calcineurin B-like protein 1 (CBL1), located on the plasma membrane, used as control] (Batistic et al., 2010) and *35S::P19*/GV3101 (C) (P19 protein: promoted protein expression) suspensions with a volume ratio of 450 μl (A): 300 μl (B): 300 μl (C); for *SaPEPC-2*, *35S::ABI5-BFP*/GV3101 (D) [abscisic acid insensitive 5 (ABI5), located in the nucleus, used as control] (Bensmihen et al., 2005) was also included in a volume ratio of 450 μl (A): 300 μl (B): 300 μl (C): 300 μl (D) of *A. tumefaciens* suspension. The mixture was held at room temperature for 2–3 h in the dark before use. About 5- to 6-week-old *Nicotiana benthamiana* plants were prepared for infiltration. The tip end of a syringe (without a needle) is placed against the underside of the leaf (in avoidance of the veins) with one finger supporting on the upper side, then gently pressing the syringe to infiltrate *A. tumefaciens* mixture into the fresh leaf and labeled the infiltration area for further recognition. The treated plants were held in the dark overnight, and were then transferred to the normal growth conditions for another 48 h. The fluorescent signals in the leaf of *N. benthamiana* were examined and photographed using the Zeiss LSM 800 confocal microscope (Carl Zeiss, Jena, Germany).

[10]http://web.expasy.org/compute_pi/

[11]http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/

[12]https://abi-services.informatik.uni-tuebingen.de/yloc/webloc.cgi

Assay of the Promoter Activity

A series of 5′-deletions of *SaPEPC-1* and *SaPEPC-2* promoters were generated according to the predicted sites of light-response elements (**Supplementary Figure 1**). FL of each promoter was truncated into six fragments and labeled as fragments 1–6 in an order from the smaller to the larger upstream of ATG. Seven specific upstream primers (Ppc-FLF and Ppc-F1 to F6) and a single downstream primer (Ppc-FLR) were designed for each fragment in a *SaPEPC* promoter. In addition, another downstream primer Ppc-R_{TSS} was designed to combine with the upstream primer Ppc-F6 to amplify the 5′-flank regions from the transcription start site (TSS), which was labeled as TSS (**Supplementary Table 3**). Following the digestion with endonucleases *Hin*dIII and *Bam*HI, the CaMV35S promoter sequence was replaced by the abovementioned fragments in the plant expression vector pBI121 to drive β-glucuronidase (*GUS*) gene. The recombinant constructs were transformed into *A. tumefaciens* strain EHA105 for the transient expression test in *N. benthamiana*, the manipulation was similar to that in the determination of subcellular localization except that the suspension of each construct was not necessary to mix with *35S::CBL1*/GV3101, *35S::ABI5*/GV3101, and *35S::P19*/GV3101 strains. Treated tobacco plants were exposed to normal illumination (500 μmol·m^{-2}·s^{-1}) for 3 days in a growth chamber, and the darkness group was placed in the dark cabinet to avoid light until sampling. A GUS fluorometric assay was performed according to Jefferson et al. (1987). All leaves were ground in liquid nitrogen and homogenized in 1.0 ml of the freshly prepared GUS extraction buffer [200 mmol·L^{-1} NaH_2PO_4, 200 mmol·L^{-1} Na_2HPO_4, 500 mmol·L^{-1} ethylenediaminetetraacetic acid (EDTA), 0.1% (v/v) Triton X-100, 0.1% (v/v) β-mercaptoethanol, and 10% (w/v) sodium dodecyl sulfate (SDS)]. After centrifugation at 12,000 rpm, 4°C for 15 min, the supernatant was employed to determine the GUS activity using 4-methylumbelliferyl glucuronide (4-MUG) as a substrate. The fluorescence of 4-methylumbelliferone (4-MU) produced by GUS-catalyzed hydrolysis was measured by the FLx800^{TM} Fluorescence Reader (BioTek, Winooski, VT, United States). The protein concentration of the supernatant was assessed by the method of Bradford (1976) using bovine serum albumin (BSA) as the standard. GUS activity was normalized to the protein concentration of each supernatant extract and calculated as pmol of 4-MU per microgram of soluble protein per minute.

Analysis of Gene Expression Profile

Gene expression profile was analyzed based on public released data. Published gene expression data sets in different tissues (matured leaves, stems, roots, and fruits) of *S. aralocaspica* were downloaded from the NCBI (SRA: SRP128359; BioProject: JNA428881) (Wang et al., 2019). The RNA sequencing (RNA-Seq) data sets of dimorphic seeds in the germination of *S. aralocaspica* were obtained from the BioProject of PRJNA325861 (Wang et al., 2017). Gene expression levels were estimated by the fragments per kilobase of exon per million mapped reads (FPKM) values using the Cufflinks software

(Trapnell et al., 2012). The heatmap was generated using the TBtools software (Chen et al., 2020), the color scale represents FPKM counts, and the ratios were expressed as log2 transformed.

Quantitative Real-Time PCR

To validate the transcriptomic data, we performed a quantitative real-time PCR (qRT-PCR) to analyze the expression of *SaPEPC-1* and *SaPEPC-2* genes. Total RNA was extracted from the collected plant samples using the E.Z.N.A.® Plant RNA Kit (Cat. R6827, OMEGA, Norcross, GA, United States) according to the instructions of the manufacturer. RNA conc. and absorbance ratios (A_{260}/A_{280} and A_{260}/A_{230}) were measured using the NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, United States). Each reverse transcription reaction was performed with 1 μg of total RNA in a final volume of 20 μl using the M-MLV RTase cDNA Synthesis Kit (D6130, TaKaRa, Shiga, Japan) with a 2.5 $\mu mol \cdot L^{-1}$ oligo (dT) primer following the instructions of the manufacturer. cDNA was stored at −20°C until use. qRT-PCR was carried out using GoTaqR® qPCR Master Mix (Promega, Madison, WI, United States) in the GeneAmp® 7500 Real-Time PCR System (ABI, Vernon, CA, United States). Gene-specific primers of *SaPEPC-1* and *SaPEPC-2* were designed using the Primer-Blast tools[13] (**Supplementary Table 3**). To ensure the amplification of the desired product, a melt-curve analysis was performed to determine that only a single peak was present to represent a unique PCR product as per the MIQE guidelines (Bustin et al., 2009). Standard curves were generated for each primer to assess efficiency, and all primers had a value of efficiency between 1.9 and 2.1 (**Supplementary Data**). β-*tubulin* gene of *S. aralocaspica* was used as an internal reference (Cao et al., 2016). The reaction mixture consisted of 1 μl cDNA samples, 0.5 μl each of the forward and reverse primers (10 $\mu mol \cdot L^{-1}$), 10 μl GoTaqR® qPCR master mix, and 8 μl nuclease-free H_2O in a final volume of 20 μl. qRT-PCR was performed as follows: 2 min initial denaturation at 95°C, followed by 40 cycles at 95°C for 15 s, and 60°C for 1 min. Four biological replicates with two technical replicates for each treatment were applied, and the data were analyzed using the $2^{-\Delta\Delta Cq}$ method (Taylor et al., 2019). The final value of relative quantification was described as a normalized fold change in the gene expression of each target gene compared to the control. Data were expressed as geometric mean ± 95% CI of four biological replicates for each treatment.

Expression and Detection of Recombinant Protein

The ORF of *SaPEPC-1* or *SaPEPC-2* was inserted into the prokaryotic expression vector pET28a. The primers used for vector construction are shown in **Supplementary Table 3**. The recombinant plasmids pET28a-*SaPEPCs* were transformed into *Escherichia coli* Transetta (DE3) strain. The positive clones were sequenced and cultivated in a liquid LB medium supplemented with 100 $mg \cdot L^{-1}$ kanamycin and 0.8 $mmol \cdot L^{-1}$ isopropyl β-D-1-thiogalactopyranoside (IPTG) at 37°C for 4 h to induce the expression of *SaPEPCs*. The cell pellets of the recombinant strains were ultrasonically treated, and the total amount of proteins was harvested by centrifuging at 12,000 *g*, 4°C for 10 min. The precipitation was resuspended and resolved by –SDS-polyacrylamide gel electrophoresis (SDS-PAGE), and the recombinant protein was detected by an immunoblot according to the following steps: upon separation on 10% (w/v) PAGE, the proteins were electroblotted onto a polyvinylidene fluoride (PVDF) membrane, which was then blocked overnight at 4°C in a Tris-buffered saline (TBS) buffer (20 $mmol \cdot L^{-1}$ Tris–HCl, pH 7.5; 150 $mmol \cdot L^{-1}$ NaCl) containing 5% (w/v) powdered milk. After the incubation with mouse anti-His monoclonal antibody (1:1000 diluted) for 2 h at 37°C, the membrane was washed four times in a TBS buffer, and then incubated with a 1:1000 diluted goat anti-mouse IgG secondary antibody. The 3,3-diaminobenzidine (DAB) was added as a chromogen for staining.

Assay of Stress Tolerance of Recombinant Protein

The recombinant (Transetta: pET-28a-*SaPEPCs*) and control (Transetta: pET-28a) strains were inoculated in a fresh LB medium containing 100 $mg \cdot L^{-1}$ kanamycin and cultured overnight at 37°C, which (1% of the culture) was then reinoculated to a fresh LB medium (in addition of 100 $mg \cdot L^{-1}$ kanamycin) and cultivated for about 4 h till the OD_{600} value reached 0.5. After the addition of 0.8 $mmol \cdot L^{-1}$ IPTG, the cultures were incubated for another 4 h at 37°C. About 1% of the diluted culture (0.8 OD_{600}) was inoculated into a 50 ml fresh LB medium (in addition of 100 $mg \cdot L^{-1}$ kanamycin) with the supplement of 400 $mmol \cdot L^{-1}$ NaCl, 10% (w/v) PEG 6000, or 25 $\mu mol \cdot L^{-1}$ methyl viologen (MV; to mimic the oxidative stress). For the test of acid or base response, the pH value of a LB medium was adjusted to 5.0. Except for the case of temperature assay (at 30°C), all other cultures were incubated at 37°C with a shaking speed of 220 rpm overnight. For the measurement of the time course of growth under different abiotic stresses, cultures (10 ml) were harvested at an interval of 3 h to a total of 12 h to measure the enzyme activity.

Assay of the Kinetic Property and Stability of Enzymes

The recombinant proteins SaPEPC-1 and SaPEPC-2 were purified under native conditions with an Ni-NTA agarose resin (Qiagen, Hilden, Germany) according to the protocol of the manufacturer. After the determination of the concentration of the purified proteins by using PEP as a substrate, the enzyme activity was determined by the addition of the different conc. of PEP (0.5, 1, 3, 4, and 5 $mmol \cdot L^{-1}$), $NaHCO_3$ (0.5, 3, 5, 10, and 20 $mmol \cdot L^{-1}$), and $MgCl_2$ (0.5, 3, 5, 10, and 20 $mmol \cdot L^{-1}$) at pH 8.0 and 25°C. The Michaelis constant K_m and the maximum reaction rate V_{max} were calculated according to the Lineweaver–Burk plot method, and the catalytic constant K_{cat} ($K_{cat} = V_{max}$/enzyme concentration) was calculated to measure the speed of an enzymatic reaction. The heat stability of the purified SaPEPCs was determined by measuring the enzyme activity at various temperatures (15–55°C). The pH stability was determined by an incubation with a 50 $mmol \cdot L^{-1}$ Tris–HCl

[13]https://www.ncbi.nlm.nih.gov/tools/primer-blast/

reaction buffer under the pH values from 7.0 to 10.0 at 25°C. The metal ion stability was determined using a reaction buffer containing 10 mmol·L^{-1} EDTA and 10 mmol·L^{-1} metal ions (Cu^{2+}, Al^{3+}, and Mn^{2+}), respectively. The deionized water was used as the control, and the enzyme activity was measured at pH 8.0 and 25°C. The effect of metabolic effectors on enzyme activity was estimated in the presence of varying amounts of allosteric activators (0, 5, 10, 20, and 40 mmol·L^{-1} glucose-6-phosphate or glycine) and inhibitors (0, 2, 5, 10, 15, 20, and 40 mmol·L^{-1} L-malate) at pH 8.0 and 25°C.

Measurement of PEPC Enzyme Activity

For PEPC activity in *S. aralocaspica*, leaves (approximately 0.1 g) were homogenized on ice with 1.0 ml of an extraction buffer containing 100 mmol·L^{-1} Tris-H_2SO_4 (pH 8.2), 7 mmol·L^{-1} β-mercaptoethanol, 1 mmol·L^{-1} EDTA, and 5% (v/v) glycerol. The homogenate was then centrifuged at 2,000 rpm, 4°C for 20 min. The supernatant was immediately used for the assay of PEPC activity according to the protocols described by Cao et al. (2015). For the enzymatic activity of recombinant proteins, the 10 ml samples from the different bacterial cultures were centrifuged at 12,000 *g* for 10 min, cell pellets were washed with a phosphate buffer, then sonicated for 10 times of 3 s of each with an interval of 10 s and centrifuged at 12,000 *g*, 4°C for 10 min. The supernatant was employed as a crude enzyme and kept on ice for immediate use. Enzyme activity was measured as described by Cheng et al. (2016). The absorbance of reaction mixtures was recorded by monitoring NADH oxidation at 340 nm on an UV-3010 spectrophotometer (Shimadzu, Kyoto, Japan). The total amount of proteins was determined at 595 nm (Bradford, 1976). One unit of PEPC enzyme activity was defined as an optical density value decrease of 0.01 per minute (Nomenclature Committee of the International Union Of Biochemistry (NC-IUB)., 1979).

Immunoblot Analysis of Photosynthetic Enzymes

Leaves (∼0.2 g) of *S. aralocaspica* were used for the extraction of soluble proteins according to the method described by Koteyeva et al. (2011). The supernatant was mixed with a loading buffer [250 mmol·L^{-1} Tris–HCl, pH 6.8, 10% (w/v) SDS, 50% (v/v) glycerol, 5% (v/v) β-mercaptoethanol, and 0.5% (w/v) bromphenol blue] as 4:1 in volume and boiled for 10 min after centrifugation at 10,000 *g*, 4°C for 10 min, the supernatant was subjected to the SDS-PAGE analysis. Protein concentration was determined using the Bradford Protein Assay Kit (Cat. PC0010, Solarbio, Beijing, China). The resolved protein samples (10 mg of each) were transferred to a PVDF membrane for an immunoblot analysis of the photosynthetic enzymes. All the primary antibodies used in this study were raised against the predicted optimal epitopic antigens of the conserved amino acid sequences of PEPC, pyruvate orthophosphate dikinase (PPDK), and ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) from *S. aralocaspica*, the amino acid residues of the epitopic antigens, and the working dilution of these antibodies were as follows: anti-SaPEPC-C (EKLSSIDAQLR, common to PEPC) IgG (1:500), anti-SaPEPC-M {[EKLS(pS)IDAQLR], for the detection of the phosphorylation of the serine residue in this sequence of PEPC} IgG (1:500), anti-SaPPDK (KLATEKGRAAKPSL) IgG (1:200), and anti-SaRubisco large subunit (RBCL) (QARNEGRDLAREGN) IgG (1:500). The secondary antibody goat anti-rabbit IgG (conjugated horseradish peroxidase) (1:2000) was used for detection. Bound antibodies were visualized by enhanced chemiluminescence (Biosharp, Beijing, China), and the images were acquired by an luminescent image analyzer (FUJIFILM LAS-4000, Tokyo, Japan).

Statistical Analysis

All data were plotted using GraphPad Prism Version 7.0 (GraphPad Software, San Diego, CA, United States) and analyzed using SPSS version 26.0 (SPSS Inc., Chicago, IL, United States). Univariate scatterplots displaying parametric data were present as mean and SD (Weissgerber et al., 2015). One-way ANOVA was used to test the significance of different treatments, and Tukey's HSD test was performed for multiple comparisons to determine significant differences between the samples at 0.05, 0.01, 0.001, and 0.0001 significance levels. When the homogeneity of variance assumption was not met, differences were analyzed using Welch's ANOVA and Games Howell *post hoc* test. Statistically significant differences between the groups at 0.05 significance level were determined by an unpaired Student's *t*-test (homoscedastic) or unpaired Student's *t*-test using Welch's correction (heteroscedastic) of a two-tailed distribution (McDonald, 2014).

RESULTS

Identification of *PEPC* Gene Family in *S. aralocaspica*

A total of three putative *PEPC* genes were identified by a local BLASTP search of *S. aralocaspica* genome, the deduced proteins were subjected to Pfam, SMART, and InterProScan databases to analyze the domains and active sites. Three non-redundant genes (GOSA_00009595-RA, GOSA_00006741-RA, and GOSA_00018957-RA) were confirmed as *SaPEPCs*, the first two were recognized as *SaPEPC-1* (GenBank: KP985714.1) and *SaPEPC-2* (GenBank: KX009562.1), respectively, which were identified in our previous work; the third one was similar to *AtPPC4* in *Arabidopsis* and denominated as *SaPEPC-4*. Detailed information of *SaPEPC* family members is shown in **Table 1**.

Phylogenetic Analysis and Sequence Alignment of PEPCs in *S. aralocaspica*

To investigate the evolutionary relationship between SaPEPCs and PEPCs of other 26 representative eukaryotic species, including dicots, monocots, ferns, mosses, and algae, we constructed an unrooted neighbor-joining phylogenetic tree using 135 PEPC proteins (**Figure 1**). The results indicate that all PEPC family members can be categorized into two distinct clades: PTPCs and BTPCs, about 70% of the PEPCs were PTPCs. Similar distribution patterns of PTPCs and BTPCs were also

TABLE 1 | Characteristics of *PEPC* gene family in *S. aralocaspica*.

Gene name	Gene model name	Gene length (bp)	ORF (bp)	Exons Introns	Protein					Subcellular location	
					Size (aa)	MW (kDa)	*pI*	GRAVY	Aliphatic index	Plant-mPLoc	YLoc+
SaPEPC-1	GOSA_00009595-RA	5651	2901	10/9	966	110.2	6.10	−0.396	90.06	Cytoplasm	Cytoplasm (60%) or nucleus (33.7%)
SaPEPC-2	GOSA_00006741-RA	6701	2901	10/9	966	110.0	5.61	−0.390	88.35	Cytoplasm	Cytoplasm (74.1%) or nucleus (15.5%)
SaPEPC-4	GOSA_00018957-RA	10637	3099	20/19	1032	116.9	6.27	−0.463	88.25	Cytoplasm	Cytoplasm (74.6%) or nucleus (23.2%)

found in different species (**Supplementary Table 4**). In the PTPC subfamily, compared with a dicot branch, monocots, mosses, and ferns were gathered together and formed another independent branch, which could further be divided into seven groups (PTPC I–PTPC VII). BTPC subfamily was distinctly classified into four groups (BTPC I–BTPC IV). Phylogenetic analysis also identified some closely related orthologous PEPCs among *S. aralocaspica*, *A. thaliana*, and *C. quinoa* (a C_3 plant, Geissler et al., 2015): SaPEPC-1, AtPPC2, and CqPPC1 were located on the same branch of PTPC IV; SaPEPC-2, AtPPC1/3, and CqPPC2 were assigned to PTPC II; whereas SaPEPC-4 was grouped into AtPPC4 and CqPPC4 cluster in BTPC III, suggesting that an ancestral set of PEPCs may exist prior to the divergence of *S. aralocaspica*, *A. thaliana*, and *C. quinoa*.

The amino acid sequence alignment among SaPEPCs, AtPPCs, CqPPCs, and ZmPEPCs showed that they shared typical conserved domains and functional sites in PTPC and BTPC genes. However, no N-terminal phosphorylation domain (SIDAQLR) was found in the polypeptide deduced from *SaPEPC-4*, *CqPPC4*, *AtPPC4*, and *ZmPEPC3* genes, instead of harboring an RNTG tetrapeptide at the C-terminus, which was commonly found in BTPC (**Figure 2A**).

Gene Structure Analysis of *PEPCs*

Different *PEPC* genes in the same species displayed great discrepancies in size. *PEPC* genes in PTPC II/IV groups contained 10 exons, whereas it was 20 in BTPC III (**Figure 2B**). By predicting the exon–intron structure in 27 plant species (including *S. aralocaspica*), we found that the length of PTPC genes was from about 4 to 13 kb, and that of BTPC genes ranged from 7.5 to 14 kb; whereas the length of *PEPC* exons was similar in the same branch, and the number of exons/introns was conserved (**Supplementary Figure 2**). The PTPC genes of dicots, monocots, and ferns contained 10–12 exons and 9–11 introns, whereas the BTPC genes contained 18–21 exons and 17–20 introns. The moss PTPC genes generally consisted of 11–14 exons and 10–13 introns, except for one member with only five exons and four introns; while the independent branches of moss BTPCs contained 32 exons and 31 introns. All algal *PEPCs* belonged to BTPC, generally containing 20–30 exons and 19–29 introns, but exceptionally, only one exon was found in that of *Ostreococcus lucimarinus* and *Microcystis aeruginosa*.

Although the exon number was largely different, the exon length of PTPC genes was similar to that of BTPC, especially exons 8, 9, and 10; whereas the intron length differed greatly, indicating that the size of *PEPC* genes largely depends on the introns. Similarly, in *S. aralocaspica*, although the length of coding region of *SaPEPC-1* and *SaPEPC-2* (both were 2,901 bp) and the number of exons/introns (10/9) were conserved, the size of *SaPEPC-2* was 1,000 bp longer than that of *SaPEPC-1*, for the 3rd intron of the former was about 10 times longer than that of the latter (**Figure 2B**).

Conserved Motif Analysis of PEPC Family

Multi-sequence alignment showed that PEPCs were highly conserved among *S. aralocaspica*, *A. thaliana*, and *C. quinoa*. The analysis of the top 10 conserved motifs of 135 PEPCs from 27 plant species revealed that all these motifs bore the prints (IPR021135) of PEPCase family, besides, motif-4 and motif-8 also contained the histidine (IPR033129) and lysine (IPR018129) active sites, respectively (**Supplementary Figure 3** and **Supplementary Table 5**). SaPEPCs and other 121 PEPCs contained all these 10 motifs and were arranged in the same order, whereas the other five PTPCs and six BTPCs were lacking some of these motifs (**Supplementary Table 6**). It suggests that PEPCs in different species are generally conserved in gene structures, protein domains, and functional motifs but can also be genetically diverse.

Subcellular Localization of SaPEPCs

The prediction of *in silico* subcellular localization showed that all three SaPEPC proteins were most probably localized in the cytoplasm, with an average possibility of 24.1% in the nucleus (**Table 1**). Our transient transformation assay in tobacco epidermal cells showed that SaPEPC-1 had a strong fluorescence signal in the cytoplasm, plasma membrane, and nucleus (**Figure 3B**) while the fluorescence signal of SaPEPC-2 was mainly observed in the nucleus (**Figure 3C** and **Supplementary Figure 4**), which is consistent with the prediction by the Plant-mPLoc and $YLoc^{+}$ software.

Analysis of the *Cis*-Regulatory Elements and Activity of *SaPEPC* Promoters

The retrieved 2,500 bp sequences upstream of the start codon of *SaPEPC* genes were queried to the PlantCARE database

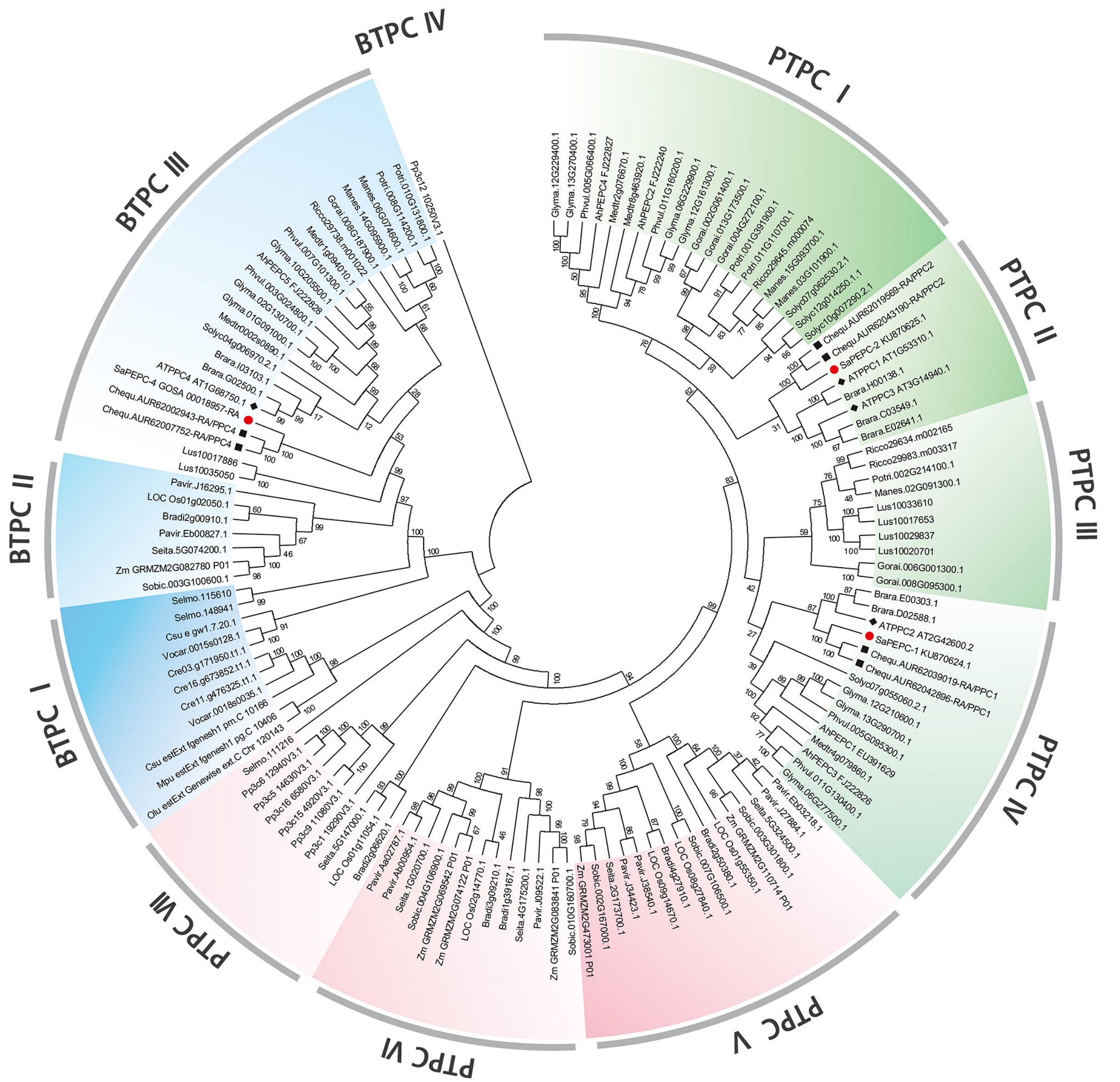

FIGURE 1 | Phylogenetic tree of phosphoenolpyruvate carboxylase (*PEPC*) gene family in different plant species. Green (dicots) and red (monocots, mosses, and ferns) portions are the plant types of PEPC (PTPC); the blue part is a bacterial type of PEPC (BTPC). At, *Arabidopsis thaliana*; Ah, *Arachis hypogaea*; Bradi, *Brachypodium distachyon*; Brara, *Brassica rapa*; Chequ, *Chenopodium quinoa*; Cre, *Chlamydomonas reinhardtii*; Csu, *Coccomyxa subellipsoidea*; Glyma, *Glycine max*; Gorai, *Gossypium raimondii*; Lus, *Linum usitatissimum*; Manes, *Manihot esculenta*; Mt, *Medicago truncatula*; Mpu, *Micromonas pusilla*; LOC Os, *Oryza sativa*; Olu, *Ostreococcus lucimarinus*; Pavir, *Panicum virgatum*; Phvul, *Phaseolus vulgaris*; Pp3c, *Physcomitrella patens*; Potri, *Populus trichocarpa*; Rc, *Ricinus communis*; Selmo, *Selaginella moellendorffii*; Seita, *Setaria italica*; Solyc, *Solanum lycopersicum*; Sobic, *Sorghum bicolor*; Sa, *Suaeda aralocaspica*; Vocar, *Volvox carteri*; and Zm, *Zea mays*. The symbols of red circle, black box, and black diamond box indicate the *PEPC* gene family members in *S. aralocaspica*, *C. quinoa*, and *Arabidopsis*, respectively.

for a *cis*-regulatory element prediction. A total of 98 *cis*-elements were detected from the three *SaPEPC* genes, in addition to the phytohormone and specific expression-related *cis*-elements, the other two *cis*-elements were involved in palisade mesophyll cell differentiation and four (ARE, WUN-motif, LTR, and MBS) in response to abiotic stresses. In particular, light-responsive *cis*-elements were up to 12 varieties (i.e., AE-box, AT1-motif, ATCT-motif, Box-4, chs-CMA1a, G-box, GA-motif, GT1-motif, I-box, MRE, TCCC-motif, and TCT-motif) (**Figure 4A** and **Supplementary Table 7**), suggesting that *SaPEPCs* may largely be involved in light regulation. Further analysis of these results might help in understanding the role of *SaPEPC* genes in development, photosynthesis, and response to stresses.

Two *SaPEPC* promoters driving *GUS* expression *in vitro* revealed that the GUS activity was gradually decreased with

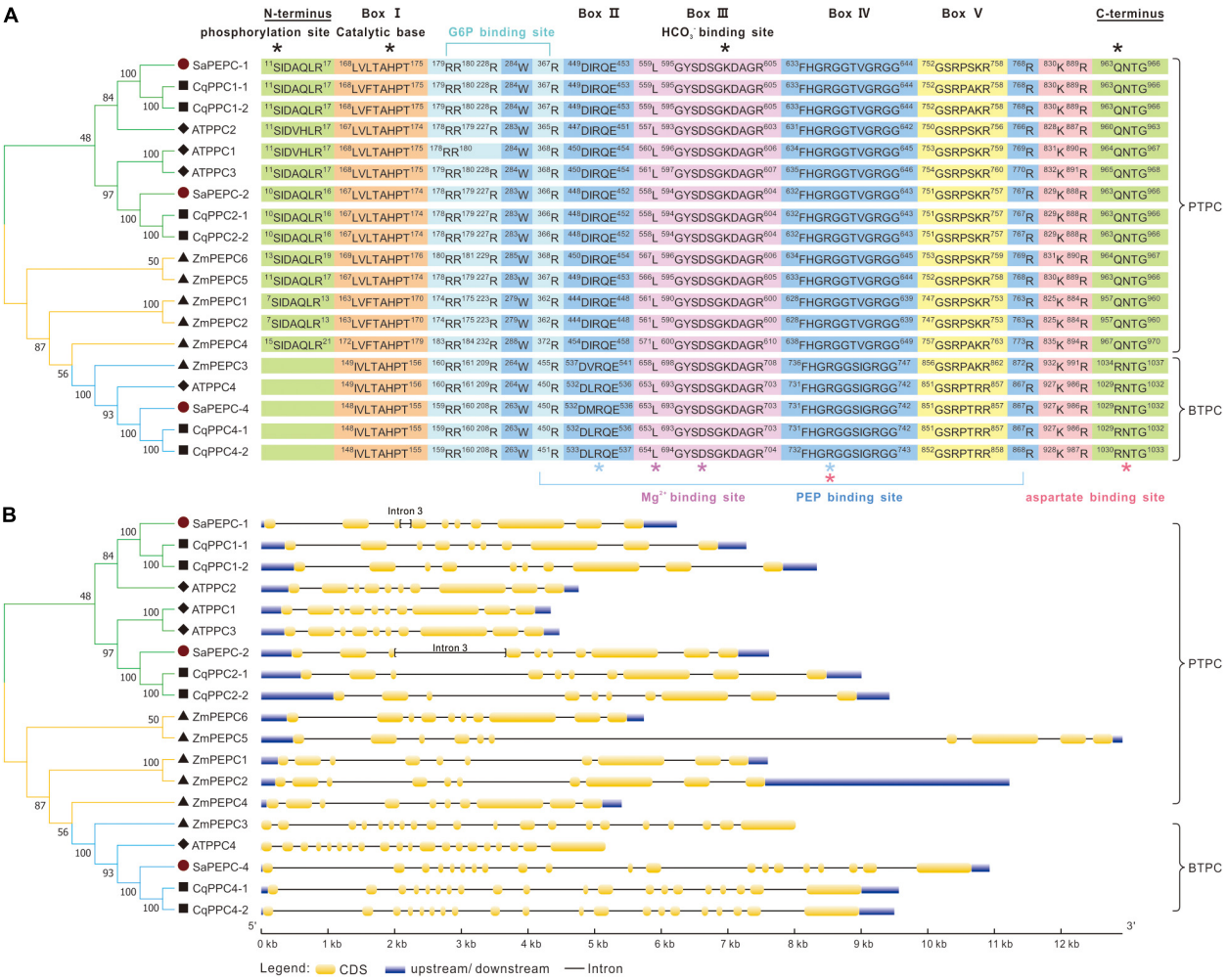

FIGURE 2 | Sequence alignment **(A)** and gene structure analysis **(B)** of PEPCs among *S. aralocaspica*, *A. thaliana*, *C. quinoa*, and *Z. mays*. Box I–V are conserved sequences responsible for the activity of PEPC enzyme. *: Indicates the key domains of the catalytic reaction and binding sites for the substrate and inhibitor.

the progressive 5′-flanking deletion of the upstream sequence ($F_{7,24}$ = 46.26 and 42.69, $p < 0.0001$ for *SaPEPC-1* and *SaPEPC-2* under light, respectively; Welch's $F_{7,10.03}$ = 263.99 and Welch's $F_{7,10.14}$ = 32.85, $p < 0.0001$ for *SaPEPC-1* and *SaPEPC-2* under darkness, respectively) and significantly responded to light (e.g., t_6 = 6.748, p = 0.0005 for TSS1 and t_6 = 9.255, $p < 0.0001$ for P2-5, respectively) (**Figures 4B,C**). Interestingly, the upstream sequence starting from the TSS (named these segments as TSS1 and TSS2 for *SaPEPC-1* and *SaPEPC-2*, respectively, which were lacking the 5′-UTR sequence compared with P1-6 and P2-6 fragments) significantly promoted GUS activity compared to that of P1-6 fragment of *SaPEPC-1* (t_6 = 5.848, p = 0.0011 under light; t_6 = 3.343, p = 0.0155 under darkness) or P2-6 fragment of *SaPEPC-2* (t_6 = 7.426, p = 0.0003 under light; t_6 = 5.310, p = 0.0018 under darkness), respectively, suggesting that a 5′-UTR region may apply a repressive effect on the activity of the *SaPEPC* promoter (**Supplementary Figure 1**). Further analysis revealed that multiple stem-loop structures might be formed at the RNA level in the 5′-UTR sequence of *SaPEPCs*, that of *SaPEPC-2* presented a more complicated secondary structure with a folding free energy (ΔG) of −51.62 kcal mol^{-1} (**Supplementary Figure 5**). These facts imply that the 5′-UTR sequence of *SaPEPCs* may play an important regulatory role in gene expression.

Spatial and Temporal Expression Patterns of *SaPEPC* Genes

Based on the available RNA-Seq data (Wang et al., 2017, 2019), the temporal expression patterns of three *SaPEPC* genes in dry, imbibed, and germinated seeds were analyzed by using the dimorphic seeds. *SaPEPC-1* accumulated more transcripts with the germination progression, and the brown seedlings responded quicker than the black ones; the transcripts of *SaPEPC-2* were accumulated to remain relatively constant and abundant in dimorphic seeds and seedlings while *SaPEPC-4* was expressed at lower levels in germination and both types of seeds (**Figure 5**). In our previous study, the transcriptional expression patterns

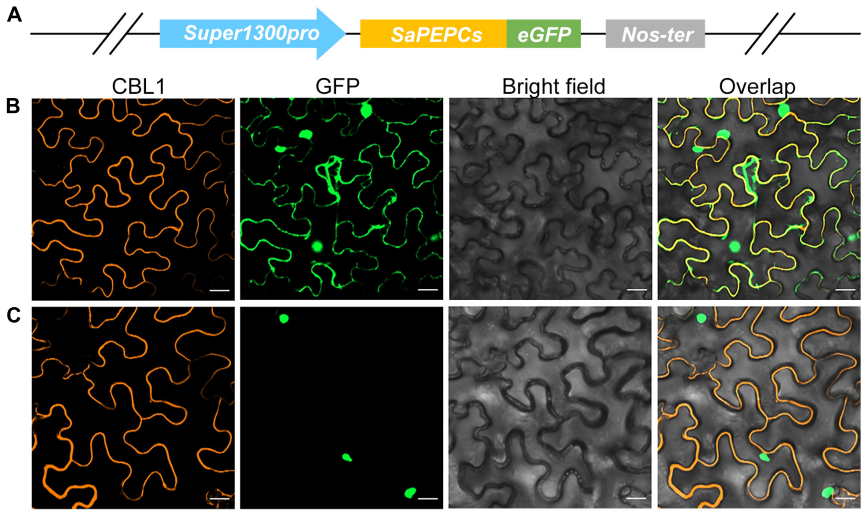

FIGURE 3 | Subcellular localization of two types of SaPEPC in tobacco epidermal cells. **(A)** Schematic diagram of vector construction; **(B)** SaPEPC-1; **(C)** SaPEPC-2. CBL1, calcineurin B-like protein 1, membrane marker control; GFP, green fluorescent protein. Bar = 20 μm.

of *SaPEPC-1* and *SaPEPC-2* were compared in terms of seed germination (0, 5, 10, and 15 days), different tissues (radicle, hypocotyl, and cotyledon), and salt stress (0, 100, 300, and 500 mmol·L^{-1}) in dimorphic seeds under normal conditions (Cheng et al., 2016). In this study, we strengthened these data by emphasizing the effect of light/dark on the expression of two *SaPEPC* genes without distinguishing dimorphic seeds, and a much longer developmental period (i.e., 0 h, 8 h, 12 h, 24 h, 2 days, 3 days, 5 days, 10 days, 15 days, 30 days, and 60 days), more different tissue types (radicles, hypocotyls, cotyledons, and roots, stems, leaves of 15- and 90-day plants), and more different types of abiotic stresses (NaCl, mannitol, ABA, and light intensity) were applied. In developing seedlings and adult plants from brown seeds, the transcript accumulation of two genes was increased gradually with seedling growth, and reached the highest value at 10 days after germination and 30 days after emergence, respectively (**Figure 6**). Light or darkness treatment in the germination stage revealed that *SaPEPC-1* expression was more sensitive to light, and the transcript copies were approximately 20 times more than that in the dark on the 10th day of germination (Welch's $t_{3.028}$ = 14.01, p = 0.0008) (**Figures 6A,B**). The results are consistent with our previous study (Cheng et al., 2016).

The spatial expression patterns of *SaPEPC* genes according to the available data showed that *SaPEPC-2* was widely expressed in different tissues (root, stem, leaf, and fruit), *SaPEPC-1* was preferentially expressed in leaves and fruits, with lower expression in roots while the expression of *SaPEPC-4* was lower in all the tested tissues (**Figure 5**). With the help of qRT-PCR analysis, the expression patterns of *SaPEPC-1* and *SaPEPC-2* in different tissues were validated, partial results were consistent with the RNA-Seq data and our previous result (Cheng et al., 2016). In general, the accumulation of *SaPEPC-1* transcripts was significantly higher than that of *SaPEPC-2* in different tissues, especially in cotyledons (Welch's $t_{3.089}$ = 15.03, p = 0.0005 under light and t_6 = 2.709, p = 0.0351 under darkness) and leaves (Welch's $t_{3.005}$ = 14.40, p = 0.0007 for 15-day seedling and Welch's $t_{3.000}$ = 13.74, p = 0.0008 for 90-day-adult plant). In comparison with the high level in leaves, *SaPEPC-1* transcripts were hardly detected in developing roots and stems, whereas the *SaPEPC-2* expressed relatively higher in roots compared to other tissue types (**Figure 7**). These results suggest that *SaPEPC-1* and *SaPEPC-2* may play diverse biological functions in plant developmental processes.

Expression Profiles of *SaPEPC* Genes in Response to Light and Abiotic Stresses

To investigate the responses of *SaPEPCs* to NaCl, mannitol, ABA, and different light intensities in developing seedlings, the transcriptional expression patterns of *SaPEPC-1* and *SaPEPC-2* were analyzed. The results showed that two *SaPEPCs* were significantly upregulated by increasing the light intensity ($F_{2,9}$ = 39.64, p < 0.0001 for *SaPEPC-1* and $F_{2,9}$ = 28.99, p = 0.0001 for *SaPEPC-2*), especially under 300 (t_6 = 3.500, p = 0.0128) and 900 μmol·m^{-2}·s^{-1} (t_6 = 4.586, p = 0.0037), the expression level of *SaPEPC-1* was significantly higher than that of *SaPEPC-2* (**Figure 8D**). Under various abiotic stresses, both *SaPEPCs* exhibited similarly positive responses to light or darkness while the transcript copies of *SaPEPC-1* were significantly higher than that of *SaPEPC-2* when exposed to 300 mmol·L^{-1} NaCl (t_6 = 6.082, p = 0.0009), 600 mmol·L^{-1} mannitol (t_6 = 9.432, p < 0.0001), and 10 μmol·L^{-1} ABA (t_6 = 2.805, p = 0.0309) treatments under normal light conditions (**Figures 8A–C**).

PEPC Enzyme Activity in Response to Development, Light, and Abiotic Stresses

Enzyme activity of SaPEPCs was measured in the leaves of different developmental stages and under different light intensities as well as abiotic stresses. As shown in **Figures 9A–C**,

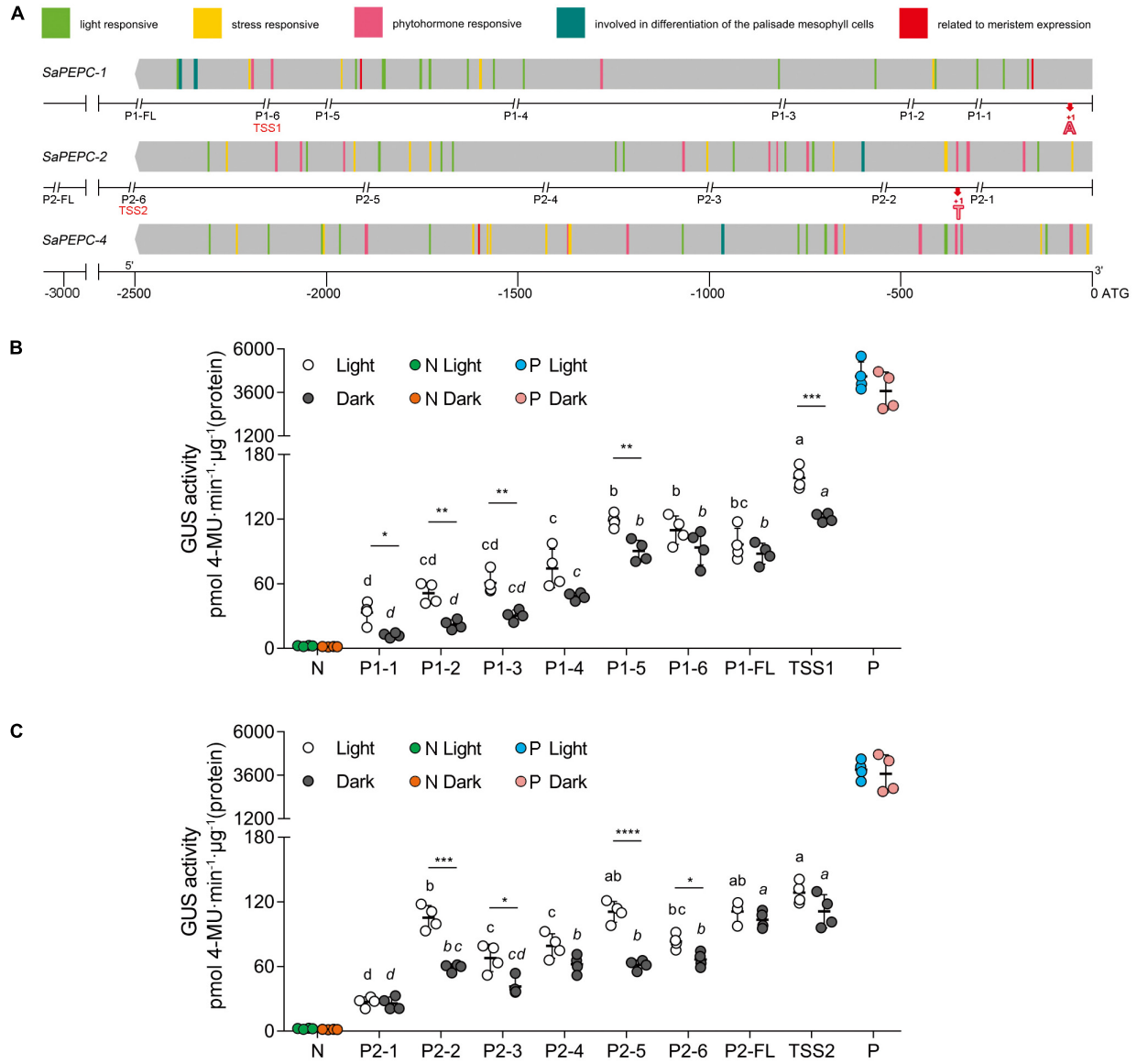

FIGURE 4 | Sequence and activity analysis of *SaPEPC* promoters. **(A)** Predicted *cis*-elements on promoters of three *SaPEPC* genes and a schematic diagram of truncated sites of two *SaPEPC* promoters; **(B)** β-glucuronidase (GUS) enzyme activity driven by *SaPEPC-1* promoter; **(C)** GUS enzyme activity driven by *SaPEPC-2* promoter. The promoter regions (2,500 bp) of three *SaPEPC* genes were analyzed, and different *cis*-elements were present in color boxes. The numbers at the bottom of figure **(A)** indicate the distance to the start codon ATG. P1 and P2 represent *SaPEPC-1* and *SaPEPC-2* promoter, respectively. FL represents a full-length sequence from ATG of each promoter, which was consequently truncated into six fragments. The double italic slash represents the truncated site; the numbers 1–6 represent different truncated fragments ranking from the smallest to the largest. The red bold "A" (+1) and "T" (+1) marked the transcription start sites (TSSs), TSS1 and TSS2 represent a upstream sequence from TSSs of *SaPEPC-1* and *SaPEPC-2* promoters, respectively. N, negative control, injected with *A. tumefaciens* EHA105; P, positive control, injected with *A. tumefaciens* EHA105 harboring with pBI121 plasmid. Different lowercase letters indicate a significant difference between different fragments; *, **, ***, ****: Represent a significant difference between light and dark treatments of the same fragment at 0.05, 0.01, 0.001, 0.0001 level, respectively.

PEPC activity varied in consistency with the expression pattern of *SaPEPC* genes, which was significantly increased with the germination time extension (Welch's $F_{3,5.152} = 101.5$, $p < 0.0001$) and the stress enhancement ($F_{3,12} = 24.91$, $p < 0.0001$ for salt stress and $F_{2,9} = 48.49$, $p < 0.0001$ for drought stress). The diurnal variation in light intensity resulted in a gradual increase of PEPC activity in the morning, and reached the highest value (average 574.44 U·mg^{-1} protein) at 12:00, then dramatically decreased thereafter until 22:00 (Welch's $F_{7,9.908} = 75.16$, $p < 0.0001$), appeared as a "unimodal" curve (**Figure 9D**). The protein accumulation of representative photosynthetic enzymes [total PEPC (PEPC-C), phosphorylated PEPC (PEPC-M), PPDK, and RBCL] was also analyzed by the varying light intensity from the morning to the evening (**Figure 9E**). Among them, the amount of

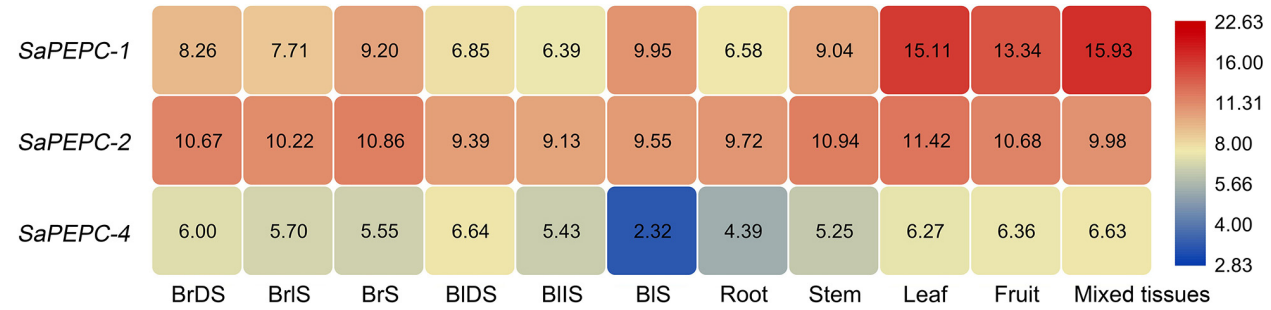

FIGURE 5 | RNA sequencing (RNA-Seq) expression analysis of the three *SaPEPC* genes obtained from the public transcriptomic database of *S. aralocaspica*. BrDS, brown dry seed; BrIS, brown imbibed seed (imbibition for 1 h); BrS, seedling germinated from brown seed (imbibition for 24 h); BIDS, black dry seed; BIIS, black imbibed seed (imbibition for 24 h); BIS, seedling germinated from black seed (germination for 10 days). The fragments per kilobase of exon per million mapped reads (FPKM) values from the RNA-Seq data were $\log^2$ transformed.

RBCL appeared to be abundant and relatively constant; PEPC-C increased with time elongation and reached the highest level from 12:00 onward; PEPC-M and PPDK were remarkably accumulated from 12:00 to 18:00 (the period with the highest light intensity in a day), and decreased significantly from then on. Our results showed that these proteins could apparently be induced by increasing light intensity, especially PEPC phosphorylation, such an expression pattern was corresponding to the diurnal changes of PEPC enzyme activity, suggesting a close relationship between light intensity and PEPC phosphorylation, and the consequent PEPC activity.

Validation of SaPEPCs in Abiotic Stress Tolerance in *E. coli*

Enzyme activity was further determined by the ectopic expression of *SaPEPCs* in *E. coli* under salt (NaCl), drought (PEG 6000), oxidation (MV), temperature, and acid/base (pH) stresses. SaPEPC-1 and SaPEPC-2 proteins were resolved by the SDS-PAGE and detected by an immunoblot analysis, which revealed a recombinant protein with the MW of 110 kDa in accordance with the theoretical values. SaPEPC-1 was expressed in the supernatant and inclusion bodies while SaPEPC-2 was mainly expressed in the inclusion bodies of the cells (**Supplementary Figure 6**). In our previous study, due to the lack of complete coding sequence of *SaPEPC-2*, only *SaPEPC-1* recombinant strain was analyzed with the time course of growth and enzyme activity [under 400 mmol·L^{-1} NaCl, 10% (w/v) PEG, 25°C, 75 μmol·L^{-1} MV, and pH 5.0] (Cheng et al., 2016). In this study, we further supplemented the corresponding data of *SaPEPC-2* recombinant strain, our results showed that the overexpression of *SaPEPC-2* could also significantly enhance cell growth and enzyme activity under different stress conditions, except for the cases of growth at 25°C and the enzyme activity under 75 μmol·L^{-1} MV (compared with *SaPEPC-1* recombinant strain) (**Supplementary Figures 7, 8**). Based on the abovementioned analysis, the enzyme activity of both *SaPEPC* recombinant strains was determined simultaneously under the conditions of 400 mmol·L^{-1} NaCl, 10% (w/v) PEG, 30°C, 25 μmol·L^{-1} MV, and pH 5.0. With an increase in stress time (to a total of 12 h), the PEPC activity of two strains increased significantly and reached the maximum value at 3 h (Welch's $F_{4,34.42} = 69.53$, $p < 0.0001$ for SaPEPC-1 and Welch's $F_{4,34.21} = 95.82$, $p < 0.0001$ for SaPEPC-2), and then reduced but generally remained higher than that of the control (**Figure 10**). In general, the SaPEPC-2 activity was higher than that of SaPEPC-1.

Carboxylase Activity and Influence of Different Effectors on Enzyme Kinetics

Different conc. of PEP, Mg^{2+}, and HCO_3^- were applied in the measurement of the recombinant SaPEPC activity to assess their effects. At pH 8.0 and 25°C, the highest K_m and V_{max} values of SaPEPC-1 or SaPEPC-2 were estimated to be approximately 0.237 or 0.231 mmol·L^{-1} and 33.85 or 57.32 U·mg^{-1} protein, respectively, with different conc. of PEP. On the contrary, the lowest K_m and V_{max} values were present under different conc. of HCO_3^-, the saturation conc. was quickly reached, and the catalytic efficiency (K_{cat}/K_m) was 39.33×10^3 or 66.35×10^3 $(mmol \cdot L^{-1})^{-1} min^{-1}$, respectively, which was two times higher than that with PEP and $MgCl_2$. Moreover, the catalytic efficiency of SaPEPC-2 with PEP or HCO_3^- was about two times as much as that of SaPEPC-1, indicating that SaPEPC-2 may possess a higher substrate affinity compared to SaPEPC-1 (**Table 2** and **Figure 11A**).

The influence of various effectors on the stability of SaPEPC-1 and SaPEPC-2 activities was investigated. The enzymatic activity of two SaPEPCs remained relatively constant at a range of temperature from 15 to 40°C while declining rapidly from 40 to 55°C, and could almost not be detected at 55°C and above, between them, SaPEPC-2 was able to tolerate higher temperature than that of SaPEPC-1. The pH stability test showed that both SaPEPCs presented the highest activity at pH 8.3, whereas the activity decreased significantly when the pH value was less than 8.0 (**Figure 11B**). Al^{3+} ion showed a significant inhibition to the enzyme activity of two SaPEPCs, which was similar to the effect of EDTA (metal ion-chelating agent). When exposed to Cu^{2+} ion, only approximately 50% activity of both SaPEPCs was detected compared to the control. As the metal ion cofactors of PEPC, Mg^{2+}, or Mn^{2+} ions displayed the highest activity

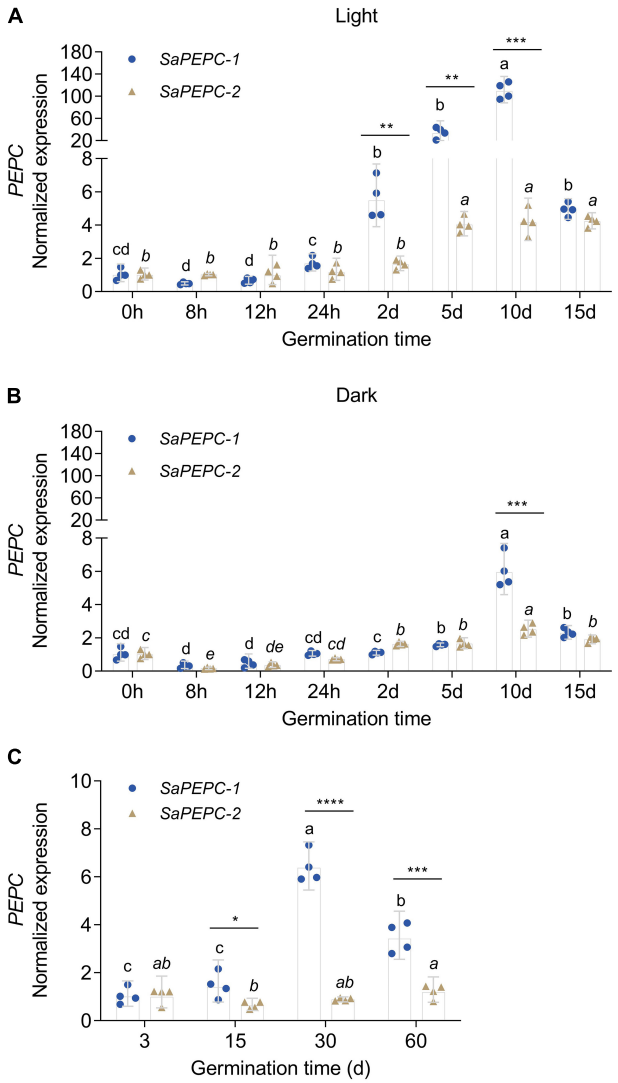

FIGURE 6 | Relative expression of two *SaPEPC* genes in seed germination and the seedling growth of *S. aralocaspica*. **(A,B)** Seed germination under light or in the dark; **(C)** Seedling growth under light. Different lowercase letters indicate a significant difference of the same gene at different time points; *, **, ***, ****: Indicate a significant difference between *SaPEPC-1* and *SaPEPC-2* at the same time point at 0.05, 0.01, 0.001, 0.0001 level, respectively. Values are geometric mean ± 95% CI of four biological replicates.

and had no significant difference compared to the control (**Figure 11C**). The kinetics of the recombinant enzymes were examined in the presence of allosteric activators and inhibitors under standard conditions. As shown in **Figure 11D**, with an increase in the concentration of glucose-6-phosphate and glycine, the activity of both SaPEPCs showed a slight increase while L-malate significantly inhibited their activity.

DISCUSSION

Phosphoenolpyruvate carboxylase plays pivotal roles in the carbon fixation of photosynthesis and a variety of metabolic and stress pathways. Therefore, clarifying different *PEPC* isoforms and their properties is necessary for further understanding of their functions. *S. aralocaspica* has evolved a unique SC C_4 pathway (Edwards et al., 2004). As a key photosynthetic enzyme, PEPC has been studied on the light-regulatory characteristics, enzyme activity, responses to multiple stresses, etc., in *S. aralocaspica* (Lara et al., 2006; Cheng et al., 2016; Koteyeva et al., 2016). However, a genome-wide analysis and the comparative study on different PEPC isoforms in *S. aralocaspica* have not been well documented. In this study, we characterized a new BTPC (SaPEPC-4), which presented a lower expression level in germination and all tested tissues compared with other isoforms. In addition to the two PTPCs (SaPEPC-1 and SaPEPC-2) we previously reported, three members of the *PEPC* gene family in *S. aralocaspica* have been classified so far. With the achievement of the complete coding sequences of the two PTPC genes, the comparative study was conducted and the results showed that *SaPEPC-1* and *SaPEPC-2* presented different spatiotemporal expression patterns and distinct subcellular localizations. Compared to *SaPEPC-2, SaPEPC-1* was much more active in the progression of plant development and in response to various stresses. *SaPEPC-1* was especially more responsive to light variations in comparison with *SaPEPC-2* (**Figures 6–8**), which may be partially related to the *cis*-elements distributed on the promoter. The more complicated secondary structure and higher free energy of the 5′-UTR sequence found in *SaPEPC-2* might apply a repressive effect on expression *in vivo*. The expression trend of two SaPTPCs in response to light and abiotic stresses was matched with the PEPC activity in *S. aralocaspica*. The recombinant SaPEPC-2 showed a higher enzymatic activity than SaPEPC-1 with different effectors *in vitro*, and both exhibited a similar pattern in response to various stresses when ectopic expressed in *E. coli*, which might be attributed to being driven by the same T7 RNA polymerase gene promoter (Dubendorff and Studier, 1991). So far, for the limited comparative study on different PEPCs, it is still uncertain whether PEPCs of different plant-type isoforms or from different plants will achieve the same response or not. Our results suggest that *SaPEPC-1* may play a major role in the C_4 photosynthetic pathway in *S. aralocaspica*.

In this study, a genome-wide analysis identified the third *PEPC* isoform from the available genomic data of *S. aralocaspica* (Wang et al., 2019). Only three *PEPC* genes were found so far in *S. aralocaspica,* which were less than that in other plant species, such as *Z. mays* (6) and *G. max* (10) (**Supplementary Table 4**), and which was only one half of the numbers in *C. quinoa*, an allotetraploid species in Chenopodiaceae (Yasui et al., 2016). Similarly, *PEPC* gene number in tetraploid cotton species is about two times as that in diploid cottons, suggesting that the difference of *PEPC* number may be associated with the interspecific hybridization or whole-genome duplication events (Wang et al., 2016; Zhao et al., 2019). PTPC and BTPC are significantly different in gene sequence and molecular structure (O'Leary et al., 2011b). In this study, the newly identified SaPEPC-4 (belonging to the BTPC subfamily) exhibited a more complicated gene structure, e.g., with 20 exons, which are much more than that in PTPCs (**Figure 2**). The most conserved 10 motifs predicted in

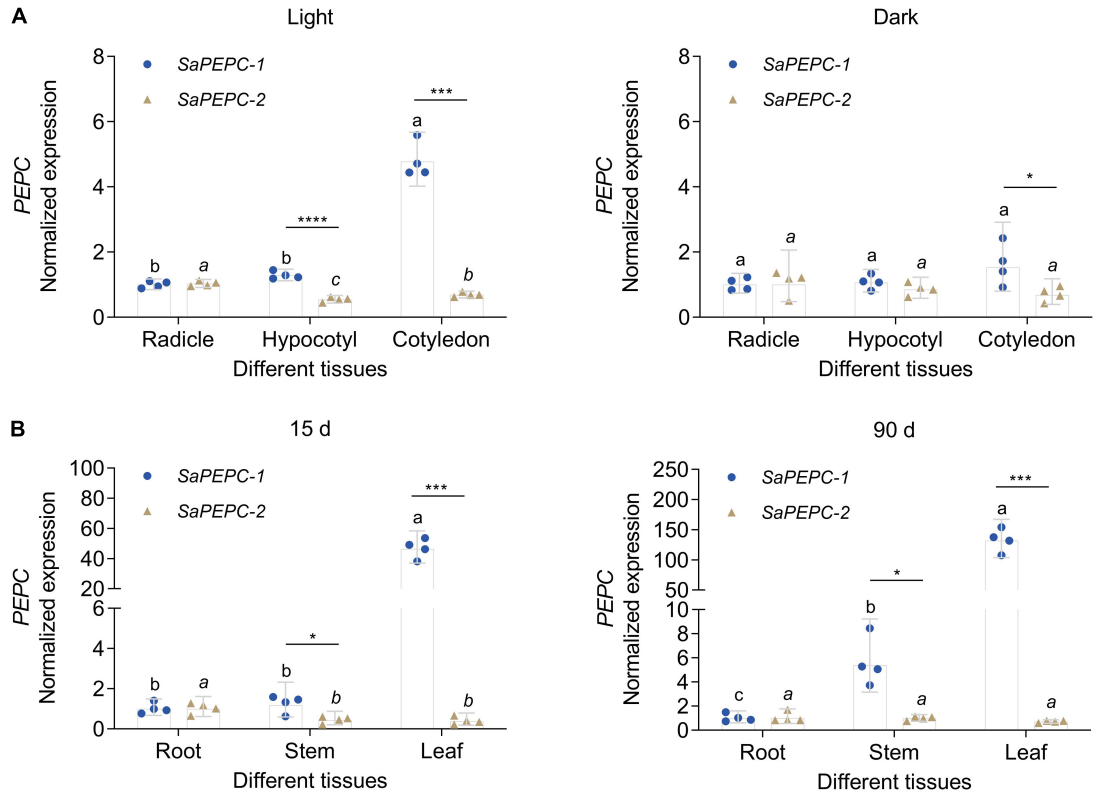

FIGURE 7 | Relative expression of two *SaPEPC* genes in different tissues of *S. aralocaspica*. **(A)** Light and darkness treatments in seed germination; **(B)** Different growth stages of seedlings. Different lowercase letters indicate significant differences of the same gene in different tissues; *, ***, ****: Indicate a significant difference between *SaPEPC-1* and *SaPEPC-2* in the same tissue at 0.05, 0.001, 0.0001 level, respectively. Values are geometric mean ± 95% CI of four biological replicates.

135 PEPCs from 27 different species were all found in the three SaPEPCs, suggesting the common origins and the evolutionary patterns of PEPCs in Viridiplantae (**Supplementary Figure 3**). Notably, however, significant discrepancies on the introns and UTR regions usually exist among various *PEPC* genes, e.g., the third intron of *SaPEPC-2* was about 10 times longer than *SaPEPC-1* in *S. aralocaspica* (**Figure 2** and **Supplementary Figure 2**). Larger introns may have more mutable sites to increase genetic diversity and promote gene evolution through alternative splicing (Kandul and Noor, 2009), whereas genes with shorter introns are selectively favored to reduce the costs of transcription and tend to be highly expressed (Seoighe et al., 2005). In the evolutionary process of multi-gene families, the diversification of gene structure may result in new functions of the gene in adaptation to the changes in the environment. However, the significance of a striking difference in the size of the third intron between *SaPEPC-1* and *SaPEPC-2* still needs to be interpreted.

All plant genomes sequenced to date contain at least one BTPC gene, including the gene of ancestral green algae (O'Leary et al., 2011a), however, the role of BTPC in plant cells remains limited. In developing castor oil seeds, BTPC functions as a catalytic and regulatory subunit appeared to interact with PTPC to form a stable Class-2 PEPC complex, which may facilitate the rapid refixation of respiratory CO_2 to replenish the C-skeletons of tricarboxylic acid cycle (TCA cycle) (O'Leary et al., 2009; Park et al., 2012). In the male gametophyte of *Lilium Longiflorum*, BTPC forms a complex with PTPC and monoubiquitinated PTPC to accelerate the accumulation of storage substances during pollen maturation (Igawa et al., 2010). In *Arabidopsis*, the downregulation of BTPC gene might increase the expression of the other three PTPC genes, indicating a transcriptional interaction between BTPC and PTPC in vascular plants (Wang et al., 2012). In this study, the transcriptional levels of three *SaPEPC* genes in *S. aralocaspica* were analyzed based on the database. Considering its detection in five tissues and at different developmental stages of the dimorphic seeds, *SaPEPC-4* (BTPC) gene showed a lower expression level in all combinations (**Figure 5**), which is consistent with that in *Arabidopsis*, soybean, and foxtail millet (Sánchez and Cejudo, 2003; Wang et al., 2016; Zhao et al., 2020). So far, however, we have not achieved the FL cDNA sequence of *SaPEPC-4* according to the genomic sequence published recently in *S. aralocaspica* (Wang et al., 2019). Therefore, further studies are needed to characterize the *SaPEPC-4* gene and verify its involvement in plant metabolism in *S. aralocaspica*.

The comparative study of two PTPC genes (*SaPEPC-1* and *SaPEPC-2*) was performed (for the lack of the complete coding sequence of *SaPEPC-4*), and they presented different subcellular localization and different expression patterns (**Figures 3**, **5**).

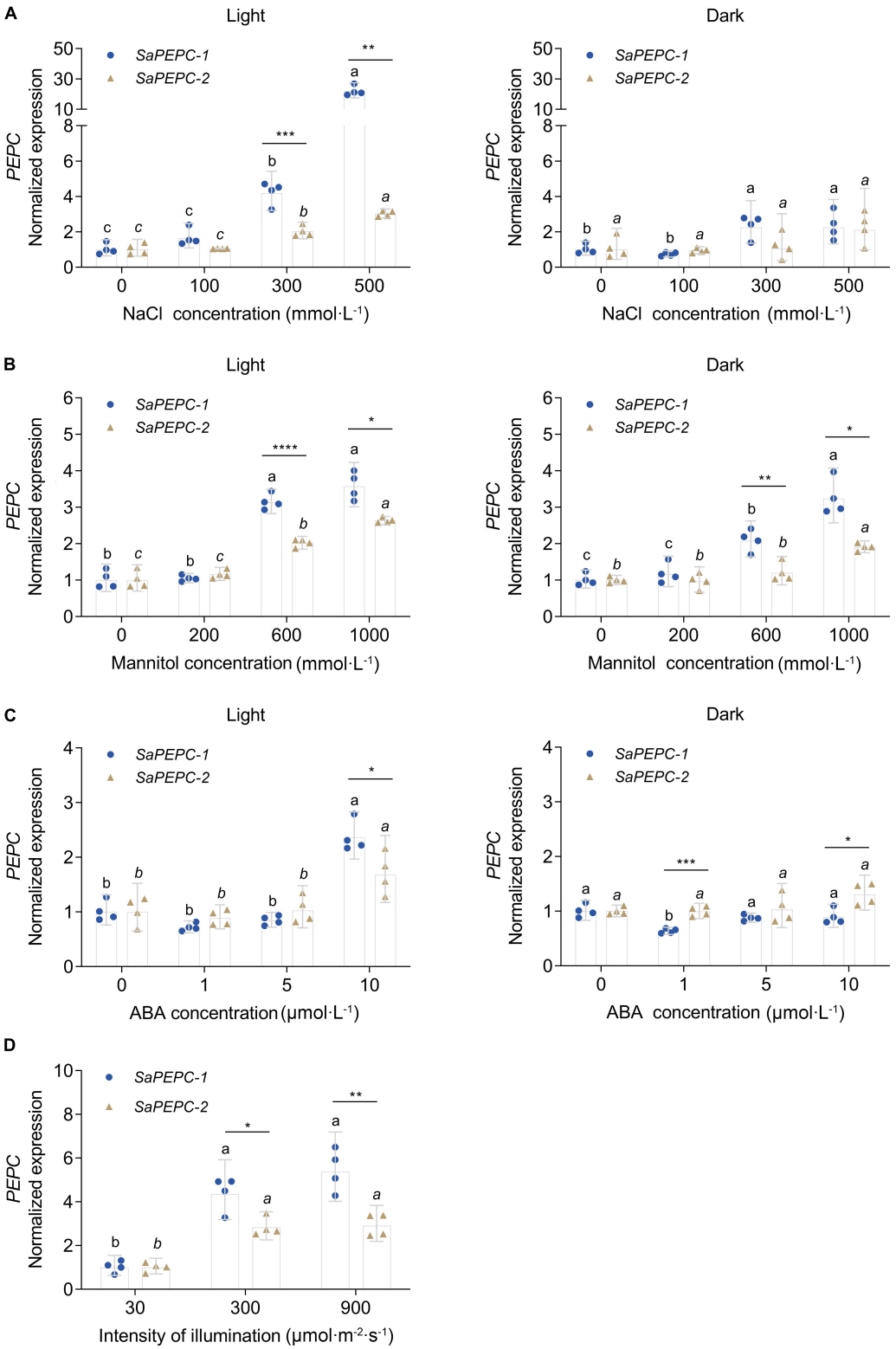

FIGURE 8 | Relative expression of two *SaPEPC* genes under different light intensity and abiotic stresses. **(A–C)** Salt, drought, ABA stresses under light and darkness; **(D)** Treatment under different light intensity. Different lowercase letters indicate a significant difference of the same gene at different concentrations (conc.) or illumination; *, **, ***, ****: Indicate a significant difference between *SaPEPC-1* and *SaPEPC-2* at the same concentration or illumination at 0.05, 0.01, 0.001, 0.0001 level, respectively. Values are geometric mean ± 95% CI of four biological replicates.

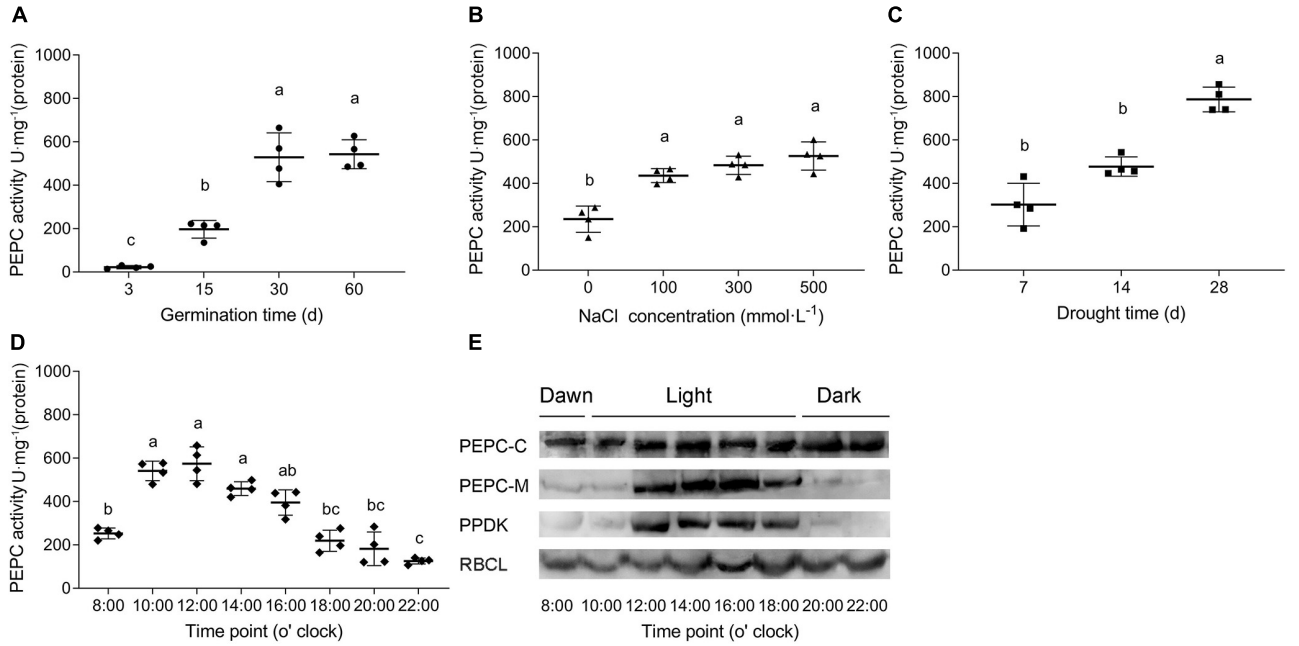

FIGURE 9 | PEPC activity in *S. aralocaspica* leaves under various treatments. **(A)** Different developmental stages of seedlings; **(B)** NaCl stress; **(C)** Natural drought stress (days of withholding water); **(D)** Activity variations within a whole day; **(E)** Variations of protein amounts of key photosynthetic enzymes. PEPC-C, total PEPC; PEPC-M, phosphorylated PEPC; RBCL, Rubisco large subunit. Different lowercase letters indicate significant difference at different treatments. Values are means ± SD of four replicates.

PEPCs are ubiquitous cytosolic enzymes in higher plants, e.g., tomato SlPEPC1, SlPEPC2, and SlPEPC3 (Waseem and Ahmad, 2019), except for rice Osppc4 targeting to the chloroplast (Chollet et al., 1996; Masumoto et al., 2010). In *S. aralocaspica*, the immunolabeling of PEPC-C protein in the fully mature chlorenchyma cells showed an even distribution throughout the cytoplasm (Koteyeva et al., 2016). In this study, SaPEPC-1 exhibited a strong fluorescent signal in the cytoplasm, whereas SaPEPC-2 appeared to exhibit nuclear localization in tobacco epidermal cells. For the potential disadvantage of any transient expression system, i.e., the overexpression or the saturation of the protein may alter the subcellular distribution (Sparkes et al., 2006), especially with the strikingly different background of a SC C_4 photosynthesis system, it is more necessary to employ a similar singular chlorenchyma cell system to determine the subcellular localization of different PEPCs. The protoplast system for the transient gene expression from the chlorenchyma cells of *B. sinuspersici* (another SC C_4 species) has been established (Lung et al., 2011), which should be a reliable system for the determination of the subcellular localization of the PEPC isoforms of SC C_4 species in the future.

Photosynthetic PEPC is highly expressed in C_4 plant leaves, whereas non-photosynthetic PEPC may have no expression specificity (Westhoff and Gowik, 2004). In our previous study, the expression level of *SaPEPC-1* was significantly higher than that of *SaPEPC-2* in cotyledons, and with seed germination progression (from dry seed to germination for 15 days) (Cheng et al., 2016). In this study, by emphasizing the effects of light/darkness and the extended developmental period (from seed germination to a 90-day adult plant), we found that *SaPEPC-1* was mainly expressed in chlorenchyma tissues (cotyledons and leaves), in which light and progressive development significantly induced its expression (**Figures 6**, **7**, **8D**), our results are consistent with that of C_4-type *PEPC* described in sorghum and maize (Crétin et al., 1991; Schäffner and Sheen, 1992). However, the expression level of *SaPEPC-2* altered in a limited range with an increase in light intensity, progressive development, and tested tissues, which is more like the *ppc-aL2* – a housekeeping gene isoform of *PEPC* in sugarcane (Besnard et al., 2003). The performance of *SaPEPC-1* was similar to *Arabidopsis AtPPC2*, which is the only *PEPC* gene expressed in green tissues and participates in carbon fixation in C_3 plants (Li et al., 2014; You et al., 2020), and both belong to the PTPC IV subgroup (**Figure 1**). In our previous study, *SaPEPC-1* shared a high homology with the *PEPC* members of C_4 species and was clustered into C_4 clade while *SaPEPC-2* was located in the C_3 cluster (Cheng et al., 2016). According to Rosnow et al. (2014), C_4 species commonly recruit *ppc-1* gene (an ortholog of *SaPEPC-1*) for use in C_4 photosynthesis. Furthermore, with the supporting evidence of the strong light activation of *SaPEPC-1* expression and significantly increased PEPC activity with the progression of development (**Figures 4B**, **6C**, **9A**), all these clues suggest that SaPEPC-1 might be a C_4-like PEPC isoform to participate in the C_4 photosynthetic pathway in *S. aralocaspica*.

Phosphoenolpyruvate carboxylase participates in plant response to various stresses and hormone signal transduction (Zhao et al., 2019; Gallego-Tévar et al., 2020). In our previous study, both *SaPTPCs* showed a significant upregulation under

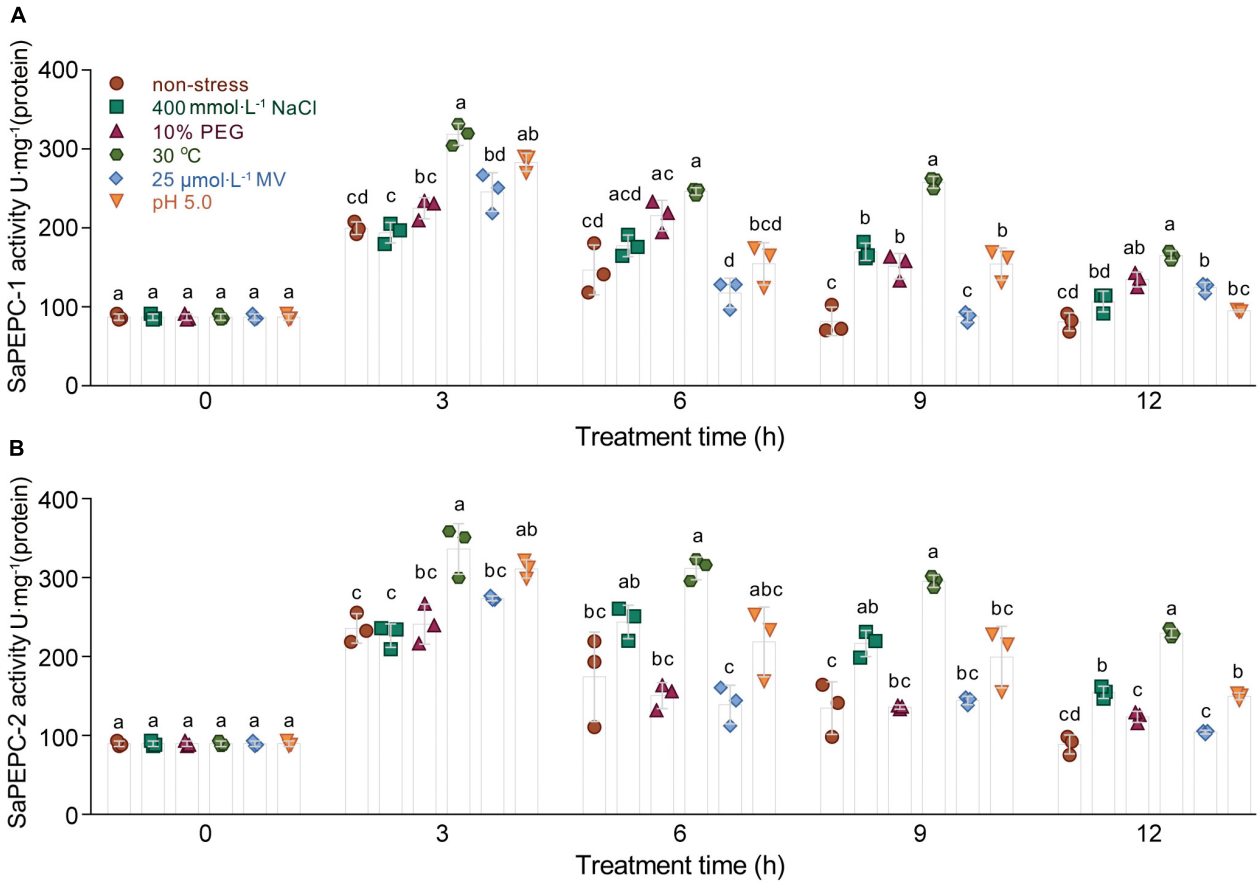

FIGURE 10 | Ectopic expression of two *SaPEPC* genes confers abiotic stress tolerance to *Escherichia coli*. **(A,B)** Activity analysis of recombinant SaPEPC-1 and SaPEPC-2 under abiotic stresses [400 mmol·L^{-1} NaCl, 10% PEG, 25 μmol·L^{-1} methyl viologen (MV), 30°C and pH 5.0]. Different lowercase letters indicate significant difference among different stresses at the same time. Values are means ± SD of three replicates.

salinity while *SaPEPC-1* accumulated much more transcripts than that of *SaPEPC-2* (Cheng et al., 2016). To explore their difference in transcriptional regulation, we examined their promoter sequences and found 15 varieties of stress-responsive *cis*-elements, surprisingly, *SaPEPC-2* contains about three times more stress- and hormone-responsive elements than *SaPEPC-1* (**Figure 4A** and **Supplementary Table 7**). Based on the predictions, the expression profiles of two *SaPTPCs* were further analyzed under the stresses of salt, drought, ABA, and high light intensity, the results indicated that both genes could positively respond to abiotic stresses, but the transcript abundance of *SaPEPC-1* was much greater than that of *SaPEPC-2* (**Figure 8**). In addition, the enzyme activity of recombinant SaPEPC-1 and SaPEPC-2 was simultaneously analyzed in comparison with the result of SaPEPC-1 only in Cheng et al. (2016), both of which were increased in response to various stresses, and consequently, the growth advantage of the recombinant strains was enhanced (**Figure 10** and **Supplementary Figure 7**). The ectopic expression of peanut *AhPEPC2* may confer more osmotic stress resistance to the recombinant strains compared to that of *AhPEPC1* and *AhPEPC5* (Tu et al., 2021). The overexpression of pearl millet C_4-specific *PEPC* in *E. coli* also displays a positive effect against abiotic stresses by increasing PEPC activity (Singh et al., 2012). It has been reported that salinity mainly applies osmotic stress on *E. coli* (Liu et al., 2019) and enhanced PEPC activity can catalyze the synthesis of malate, besides as an osmolyte and also to potentially regulate intracellular pH balance and counteract the excess toxic ions to help cells to tolerate stresses (Martinoia and Rentsch, 1994).

To further compare the difference between the two SaPTPC isoforms, the recombinant proteins of SaPEPC-1 and SaPEPC-2 were produced to analyze their enzymatic kinetics *in vitro*, which may avoid the effect of post-translational modification *in vivo* (Rao et al., 2008). Purified SaPEPC-1 and SaPEPC-2 exhibited a specific carboxylation enzymatic activity (32.437 and 54.927 U·mg^{-1} protein, respectively) (**Table 2**), which is comparable with the PEPC from other plant species and algae (ranging from ∼20 to 35 U·mg^{-1} protein) (Mamedov et al., 2005; Chang et al., 2014). PEPC activity is affected by substrates, ions, conc. of metabolites, activators/inhibitors, temperature, pH, etc., significant differences in enzymatic properties are observed among the different types and sources of PEPCs (Paulus et al., 2013). Similar to *Arabidopsis* AtPPC3, both SaPTPCs exhibited an optimal activity at pH 8.3 while SaPTPCs had a higher heat

stability by retaining 50% activity at 45°C in the presence of a bivalent metal cofactor Mg^{2+} (**Figures 11B,C**), in comparison with AtPPC3 showing the loss of 90% activity (O'Leary et al., 2009). Glucose-6-phosphate and glycine are able to activate maize C_4-PEPC activity by more than 2-fold, and its root-PEPC is more sensitive to the feedback inhibitor L-malate (Dong et al., 1998). In this study, only a slight increase in two SaPTPC activities was detected by applying activators while L-malate significantly inhibited their activity (approximately 4.6- and 2.5-fold for SaPEPC-1 and SaPEPC-2, respectively) (**Figure 11D**), which were similar to the non-C_4 or root-type PEPC in maize. The HuPPC3 in pitaya is also more sensitive to malate but involved in the initial fixation of atmospheric CO_2 in crassulacean acid metabolism (CAM) photosynthesis (Nomura et al., 2020). The catalytic efficiency values (K_{cat}/K_m, $(\text{mmol}\cdot\text{L}^{-1})^{-1}\text{min}^{-1}$) for the substrates of PEP and $HCO_3{}^-$ of *E. coli* PEPC are 1.5×10^3 and 9.0×10^4, respectively (Kai et al., 1999; Lee et al., 2013). The K_{cat}/K_m values of purified wild-type and N-terminal truncated PEPCs in *Phaeodactylum tricornutum* for PEP are about 2–4 times higher than that of *E. coli* PEPC, while for $HCO_3{}^-$ the values are about 15–34% (Chang et al., 2014). In this study, the K_{cat}/K_m values of recombinant SaPEPC-1 and SaPEPC-2 for PEP were 10 and 18 times of that of *E. coli* PEPC, and the values for $HCO_3{}^-$ were only 44 and 74%, respectively (**Table 2**), suggesting that two SaPTPCs may be more efficient in CO_2 utilization.

In this study, the K_m values of two recombinant SaPTPCs were lower compared to those native PEPCs in *S. aralocaspica* (Rosnow et al., 2015), which is also observed in *Pennisetum glaucum* (Singh et al., 2012). PTPCs commonly contain a serine residue at N-terminus, which can be phosphorylated by PEP carboxylase kinase (PPCK) and dephosphorylated by protein phosphatase 2A (PP2A) (O'Leary et al., 2011a). The higher K_m value of native PEPCs might be involved in the phosphorylation activation *in vivo*. In Kranz-type and SC C_4 species, the phosphorylation of PEPC is triggered by light and consequently leads to an increase in catalytic activity of the enzyme (Ping et al., 2018). According to Lara et al. (2006), the phosphorylation of PEPC is enhanced by the increasing light intensity from 7:00 to 17:00, and dramatically reduced by the decreasing light intensity from 20:30 to 00:30, with the major proportion of PEPC-M at 13:00 and 17:00 in *S. aralocaspica*. In this study, the diurnal activity changes of different photosynthetic enzymes were further analyzed although the total amount of SaPEPC protein remained basically constant, the accumulation of phosphorylated SaPEPC and SaPPDK was initiated from dawn (8:00), reached the maximum value by midday (14:00, in winter), and drastically decreased before the sunset (20:00), the key period for PEPC phosphorylation occurred between 12:00 and 18:00, the dephosphorylation of SaPEPC was almost complete by night (22:00) (**Figure 9E**), which is in accordance with the observations by Lara et al. (2006), and in maize (Ueno et al., 2000; Fukayama et al., 2003), indicating that the higher light intensity stimulates PEPC phosphorylation. In correspondence to such changes, PEPC activity also fluctuated within a whole day and reached the maximum value at 12:00 in *S. aralocaspica* (**Figure 9D**). Not only did the light intensity enhance the PEPC activity, but also other factors could apply effects on, e.g., PEPC

TABLE 2 | Enzyme activity of purified SaPEPC proteins and enzymatic kinetics parameters of PEP, $HCO_3{}^-$ and $MgCl_2$.

Protein	**Specific activity ($U\cdot mg^{-1}$)**	**Enzymatic kinetics parameters**											
		PEP				**$HCO_3{}^-$**				**$MgCl_2$**			
		V_{max} ($U\cdot mg^{-1}$)	**K_m ($mmol\cdot L^{-1}$)**	**K_{cat} (min^{-1})**	**K_{cat}/K_m [$(mmol\cdot L^{-1})^{-1}$ min^{-1}]**	**V_{max} ($U\cdot mg^{-1}$)**	**K_m ($mmol\cdot L^{-1}$)**	**K_{cat} (min^{-1})**	**K_{cat}/K_m [$(mmol\cdot L^{-1})^{-1}$ min^{-1}]**	**V_{max} ($U\cdot mg^{-1}$)**	**K_m ($mmol\cdot L^{-1}$)**	**K_{cat} (min^{-1})**	**K_{cat}/K_m [$(mmol\cdot L^{-1})^{-1}$ min^{-1}]**
SaPEPC-1	32.437 ± 1.21	33.85 ± 0.76	0.237 ± 0.036	3.76×10^3	15.86×10^3	31.87 ± 3.21	0.090 ± 0.15	3.54×10^3	39.33×10^3	32.26 ± 3.88	0.149 ± 0.18	3.58×10^3	24.03×10^3
SaPEPC-2	54.927 ± 3.17	57.32 ± 1.25	0.231 ± 0.032	6.37×10^3	27.57×10^3	50.73 ± 6.06	0.085 ± 0.17	5.64×10^3	66.35×10^3	56.59 ± 6.04	0.301 ± 0.25	6.29×10^3	20.90×10^3

The reaction conditions are pH 8.0, 25°C. The data are represented as means with standard deviations from three independent experiments.

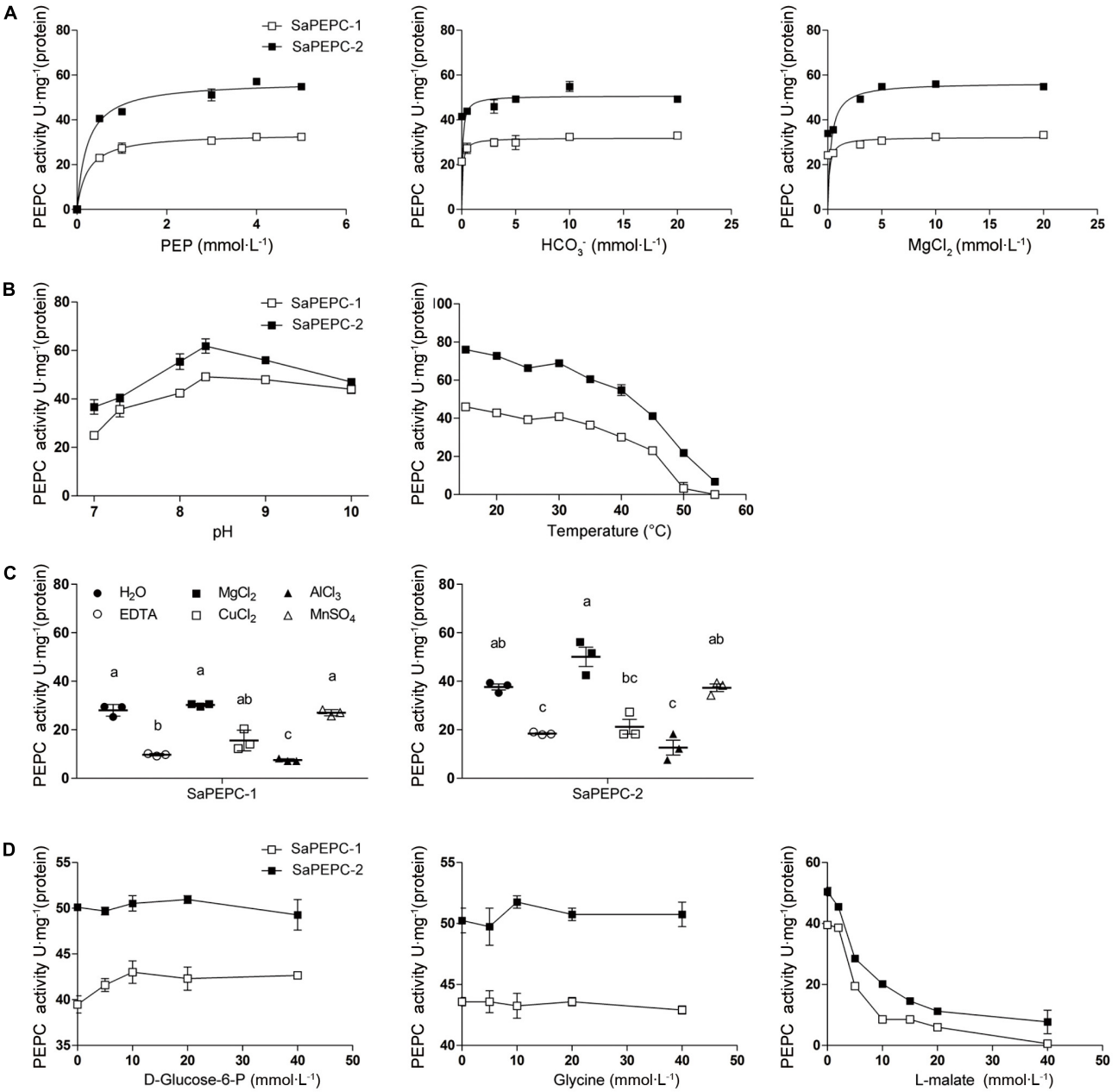

FIGURE 11 | Effects of different factors on enzyme activity of purified recombinant SaPEPC proteins. **(A)** PEP, HCO_3^-, and $MgCl_2$ saturation curves; **(B)** Test of pH and temperature stability; **(C)** Test of metal ion stability; **(D)** Test of allosteric activator and inhibitor stability. Different lowercase letters indicate significant difference among different metal ion treatments. Values are means ± SD of three replicates.

activity changed from a couple of tens to several hundred ($U \cdot mg^{-1}$ protein) with the progression of development and salt or drought stress strengthening (**Figures 9A–C**), which was also matched with *PEPC* gene expression trends at the transcriptional level (**Figures 6**, **8**). Phosphorylated PEPC exhibits low malate sensitivity while the dephosphorylated PEPC is strongly inhibited by malate to avoid a futile photosynthetic cycle (Nomura et al., 2020). This might be a cue for the sensitive inhibition effect of malate on SaPEPC-1 activity *in vitro*. It is worth noting that, in our previous study, the two patterns of the daily variations of PEPC activity were present as a "double peak" (outdoor) or "unimodal" (greenhouse) in *S. aralocaspica* (Liu et al., 2020), which were closely associated with the cultivation conditions, e.g., illumination hours, temperature, etc. In this study, the plants were cultivated and analyzed in a greenhouse, so the PEPC activity presented a "unimodal" trend, which is also an evidence to support that PEPC activity varies with a change in the light intensity.

SaPEPC-1 recombinant protein presented lower activity and enzymatic kinetics compared to SaPEPC-2 (**Table 2**), while the transcriptional level of *SaPEPC-1* in *S. aralocaspica* was significantly higher than that of *SaPEPC-2* (**Figures 6**, **8**). What a possible regulatory mechanism is behind the expression of these *SaPEPCs*? With the investigation of the promoter activity of *SaPEPC-1* and *SaPEPC-2* by driving *GUS* gene, we found no significant difference in the promoter activity at the transcriptional level *in vitro* (t_6 = 1.655, p = 0.1491 under light and t_6 = 2.415, p = 0.0522 under darkness) while two *SaPEPC* promoters similarly enhanced *GUS* expression at the translational level when the 5′-UTR sequence was deleted (**Figures 4B,C**), suggesting that two *SaPEPCs* might have experienced more complicated regulations *in vivo*, e.g., the posttranscriptional modulation by the 5′-UTR sequence. The secondary structures of 5′-UTR have been characterized as the negative regulators of gene expression at both transcriptional and translational levels (Stefanovic et al., 2000; Bunimov et al., 2007). In the tomato pollen-specific promoter of *LAT59* gene, the stem-loop structure of 5′-UTR dramatically decreased the messenger RNA (mRNA) accumulation of the reporter gene without affecting the translation rate and mRNA stability (Curie and McCormick, 1997). A more complicated 5′-UTR secondary structure is located upstream of nitrate reductase (NR) gene in *Chlorella vulgaris*, which accelerates the degradation of NR transcripts (Cannons and Cannon, 2002). In this study, the predicted secondary structure of the 5′-UTR of *SaPEPC-1* was relatively simple with less than −4 kcal/mol free energy (ΔG) while in *SaPEPC-2,* that was complicated and stable with a maximum ΔG value of −51.62 kcal/mol at the RNA level (**Supplementary Figure 5**). It is proposed that every −10 kcal/mol of ΔG is sufficient to reduce the translation efficiency by about 50%, and −50 kcal/mol of ΔG may inhibit more than 85% of the translation efficiency (McCarthy, 1998; Brunn et al., 2012). Therefore, the higher ΔG of 5′-UTR secondary structure might inhibit the *SaPEPC-2* expression *in vivo*; alternatively, *SaPEPC-2* might affect the stability of mRNA with the secondary structure of 5′-UTR, which consequently resulted in the abundance of a lower transcript in *S. aralocaspica*.

CONCLUSION

This study characterized the members of a genome-wide *PEPC* gene family and comparatively analyzed the two PTPC isoforms in SC C_4 species *S. aralocaspica.* With the *in silico* analysis of *S. aralocaspica* genomic database, a new bacterial-type *SaPEPC-4* gene was identified based on the previously reported two PTPC genes (*SaPEPC-1* and *SaPEPC-2*, the latter with a partial sequence), while its function is still unknown. Three *SaPEPC* genes were differentially expressed in roots, stems, leaves, fruits, and heteromorphic seed development, and presented distinct subcellular localization patterns. The transcript abundance and enzyme activity (native or recombinant) of the two PTPC genes (*SaPEPC-1* and *SaPEPC-2*) were stimulated by light and abiotic stresses. In *S. aralocaspica*, the transcript copies of *SaPEPC-1* were significantly higher than that of *SaPEPC-2* under various stresses, but the enzymatic kinetics (V_{max} and K_{cat}/K_m for different substrates) and biochemical properties (heat stability, activator and inhibitor responses) of the latter were higher than that of the former *in vitro*. The 5′-UTR regions of the two *SaPEPC* promoters might apply the repression effect on the expression of *PEPC* genes at transcriptional, posttranscriptional, and/or translational levels. In terms of phylogenetic relationship, spatiotemporal expression pattern, light sensitivity, *SaPEPC-1* gene is more likely to be recruited as a C_4-type PEPC; whereas *SaPEPC-2* behaves like a non-photosynthetic housekeeping gene. However, the details for the regulation of different *PEPC* genes *in vivo* still need to be interpreted with more efforts. Our findings may lead to decipher the exact roles of PEPC isoforms in C_4 photosynthesis, plant growth/development, and tolerance against stresses in *S. aralocaspica* and the similar species.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

HL, JC, and GC designed the experiments and methodology. JC and HL wrote the manuscript. JC, GC, LW, and TM conducted the experiments and collected the data. JC analyzed the data. All authors contributed critically to the manuscript and gave final approval for publication.

FUNDING

This work was supported by the National Natural Science Foundation of China (31460043, 31260037, and 31060027) and Project for Training Young Talents of Xinjiang Uygur Autonomous Region (2013721013).

ACKNOWLEDGMENTS

The authors thank the reviewers and all editors for their helpful comments and suggestions on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.665279/full#supplementary-material

Supplementary Figure 1 | Constructs of plant expression vectors with a series of 5′-end deletion of *SaPEPC-1* and *SaPEPC-2* promoter sequences. **(A–D)** *SaPEPC-1*; **(E–H)** *SaPEPC-2*; **(A,E)** A schematic diagram of deletions at the 5′-end of two promoters; **(B,F)** A schematic diagram of promoter fragments fused

with β-glucuronidase (GUS) gene; **(C,G)** PCR amplification of truncated promoter fragments; **(D,H)** Identification of the plant expression vector pBI121 replaced by different phosphoenolpyruvate carboxylase (PEPC) promoter fragments for CaMV35S promoter. Red box represents the light-response elements; light green box represents the stress-related elements; yellow box represents the endosperm-expressing elements; grass green box represents the root-specific expression and mesophyll cell development-related elements; purple box represents a high-level transcriptional element of 5′-UTR; the double italic slash represents the truncated site. M1, DL 5000 Marker; M2, DL 15000 Marker; M3, DL 15000 + 5000 Marker.

Supplementary Figure 2 | Analysis of exons and introns in *PEPC* genes of different plant species. **(A)** Plant-type PEPC (PTPC) genes of dicots; **(B)** PTPC genes of monocots, mosses, and ferns; **(C)** Bacterial-type PEPC (BTPC) genes in plants.

Supplementary Figure 3 | Distribution of conserved motifs in *PEPC* genes of different plant species. **(A)** PTPC genes of dicots; **(B)** PTPC genes of monocots, mosses, and ferns; **(C)** BTPC genes in plants; **(D)** Schematic diagram of base enrichment of top 10 motifs. The five-pointed star represents the active site of PEPC enzyme.

Supplementary Figure 4 | Subcellular localization of SaPEPC-2 in tobacco epidermal cells. CBL1, calcineurin B-like protein 1, membrane marker control; ABI5, abscisic acid insensitive 5, nucleus marker control; GFP, green fluorescent protein. Bar = 20 μm.

Supplementary Figure 5 | Analysis of 5′-UTR sequence of two *SaPEPC* promoters. **(A,B)** *SaPEPC-1*; **(C,D)** *SaPEPC-2*. **(A,C)** Secondary structure of DNA; **(B,D)** Secondary structure of RNA. The free energy (Δ*G*) of respective structure is shown.

Supplementary Figure 6 | Ectopic expression of *SaPEPC-1* and *SaPEPC-2* in *Escherichia coli* and the immunoblotting analysis. **(A)** Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) analysis of recombinant SaPEPC proteins; **(B)** Immunoblotting analysis of the recombinant SaPEPC proteins with anti-His antibody. NI, non-induced total protein; Total, total crude protein; Sol, soluble protein; Ins, insoluble protein; lane M1, protein molecular weight (MW) marker; and lane M2, prestained protein MW marker. The arrowhead indicates the specific protein band of approximately 110 kDa corresponding to PEPC derivatives.

Supplementary Figure 7 | Time courses of the growth of *SaPEPC-1* **(I)** and *SaPEPC-2* **(II)** recombinant strains under different abiotic stresses. **(A,A′)** Non-stressed condition; **(B,B′)** 400 mmol·L^{-1} NaCl; **(C,C′)** 10% PEG; **(D,D′)** 25°C (30°C for *SaPEPC-2*); **(E,E′)** 75 μmol·L^{-1} methyl viologen (MV); and **(F,F′)** pH 5.0. The culture was sampled at an interval of 2 h to a total of 12 h. Values are means ± SD of three replicate.

Supplementary Figure 8 | Growth and enzyme activity of recombinant *SaPEPC* strains under temperature and MV treatments for 12 h. **(A)** Growth of recombinant *SaPEPC-1* and *SaPEPC-2* strains under different temperatures. **(B)** Enzyme activity of recombinant *SaPEPC-1* and *SaPEPC-2* strains under MV stress. *, ***, ****: Indicate a significant difference between control strain and recombinant *SaPEPC* strains at the same temperature at 0.05, 0.001, 0.0001 level, respectively. Different lowercase letters indicate significant differences between control and stress treatment at different MV concentrations. Values are means ± SD of three replicates.

REFERENCES

Arnon, D. I., and Hoagland, D. R. (1940). Crop production in artificial culture solutions and in soils with special reference to factors influencing yields and absorption of inorganic nutrients. *Soil Sci.* 50, 463–485.

Batistic, O., Waadt, R., Steinhorst, L., Held, K., and Kudla, J. (2010). CBL-mediated targeting of CIPKs facilitates the decoding of calcium signals emanating from distinct cellular stores. *Plant J.* 61, 211–222. doi: 10.1111/j.1365-313X.2009.04045.x

Bensmihen, S., Giraudat, J., and Parcy, F. (2005). Characterization of three homologous basic leucine zipper transcription factors (bZIP) of the ABI5 family during *Arabidopsis* thaliana embryo maturation. *J. Exp. Bot.* 56, 597–603. doi: 10.1093/jxb/eri050

Besnard, G., PinÇon, G., D'Hont, A., Hoarau, J. Y., Cadet, F., and Offmann, B. (2003). Characterization of the phosphoenolpyruvate carboxylase gene family in sugarcane (*Saccharum* spp.). *Theor. Appl. Genet.* 107, 470–478. doi: 10.1007/s00122-003-1268-2

Bradford, M. M. (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

Brunn, N. D., Sega, E. G., Kao, M. B., and Hermann, T. (2012). Targeting a regulatory element in human thymidylate synthase mRNA. *Chembiochem.* 13, 2738–2744. doi: 10.1002/cbic.201200603

Bunimov, N., Smith, J. E., Gosselin, D., and Laneuville, O. (2007). Translational regulation of PGHS-1 mRNA: 5′ untranslated region and first two exons conferring negative regulation. *Biochim. Biophys. Acta* 1769, 92–105. doi: 10.1016/j.bbaexp.2007.01.004

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55, 611–622. doi: 10.1373/clinchem.2008.112797

Cannons, A. C., and Cannon, J. (2002). The stability of the Chlorella nitrate reductase mRNA is determined by the secondary structure of the 5′-UTR: implications for posttranscriptional regulation of nitrate reductase. *Planta* 214, 488–491. doi: 10.1007/s00425-001-0679-z

Cao, J., Lv, X. Y., Chen, L., Xing, J. J., and Lan, H. Y. (2015). Effects of salinity on the growth, physiology and relevant gene expression of an annual halophyte grown from heteromorphic seeds. *AoB Plants* 7:lv112. doi: 10.1093/aobpla/plv112

Cao, J., Wang, L., and Lan, H. Y. (2016). Validation of reference genes for quantitative RT-RCR normalization in Suaeda aralocaspica, an annual halophyte with heteromorphism and C4 pathway without Kranz anatomy. *Peer J.* 4:e1697. doi: 10.7717/peerj.1697

Chang, K. S., Jeon, H., Seo, S., Lee, Y., and Jin, E. (2014). Improvement of the phosphoenolpyruvate carboxylase activity of Phaeodactylum tricornutum PEPCase 1 through protein engineering. *Enzyme Microb. Technol.* 60, 64–71. doi: 10.1016/j.enzmictec.2014.04.007

Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., et al. (2020). TBtools - an integrative toolkit developed for interactive analyses of big biological data. *J. Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Cheng, G., Wang, L., and Lan, H. Y. (2016). Cloning of PEPC-1 from a C_4 halophyte Suaeda aralocaspica without Kranz anatomy and its recombinant enzymatic activity in responses to abiotic stresses. *Enzyme Microb. Tech.* 83, 57–67. doi: 10.1016/j.enzmictec.2015.11.006

Chollet, R., Vidal, J., and O'Leary, M. H. (1996). Phosphoenolpyruvate carboxylase: a ubiquitous, highly regulated enzyme in plants. *Annu. Rev. Plant Biol.* 47, 273–298. doi: 10.1146/annurev.arplant.47.1.273

Commissione Redactorum Florae Xinjiangensis (1994). *Flora Xinjiangensis*. Urumchi: Xinjiang Science & Technology & Hygiene Publishing House.

Crétin, C., Santi, S., Keryer, E., Lepiniec, L., Tagu, D., Vidal, J., et al. (1991). The phosphoenolpyruvate carboxylase gene family of Sorghum: promoter structures, amino acid sequences and expression of genes. *Gene* 99, 87–94. doi: 10.1016/0378-1119(91)90037-c

Curie, C., and McCormick, S. (1997). A strong inhibitor of gene expression in the 5′ untranslated region of the pollen-specific LAT59 gene to tomato. *Plant Cell.* 9, 2025–2036. doi: 10.1105/tpc.9.11.2025

Dong, L. Y., Masuda, T., Kawamura, T., Hata, S., and Izui, K. (1998). Cloning, expression, and characterization of a root-form phosphoenolpyruvate carboxylase from Zea mays: comparison with the C_4-form enzyme. *Plant Cell. Physiol.* 39, 865–873. doi: 10.1093/oxfordjournals.pcp.a029446

Dubendorff, J. W., and Studier, F. W. (1991). Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* 219, 45–59. doi: 10.1016/0022-2836(91)90856-2

Edwards, G. E., Franceschi, V. R., and Voznesenskaya, E. V. (2004). Single-cell C_4 photosynthesis versus the dual-cell (Kranz) paradigm. *Annu. Rev. Plant Biol.* 55, 173–196. doi: 10.1146/annurev.arplant.55.031903.141725

Edwards, G. E., and Voznesenskaya, E. V. (2011). "C_4 photosynthesis: Kranz forms and single-cell C_4 in terrestrial plants," in *C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms*, eds A. S. Raghavendra and R. F. Sage (Dordrecht, Netherlands: Springer), 29–61. doi: 10.1007/978-90-481-9407-0_4

Fukayama, H., Hatch, M. D., Tamai, T., Tsuchida, H., Sudoh, S., Furbank, R. T., et al. (2003). Activity regulation and physiological impacts of maize C_4-specific phosphoenolpyruvate carboxylase overproduced in transgenic rice plants. *Photosynth. Res.* 77, 227–239. doi: 10.1023/A:1025861431886

Gallego-Tévar, B., Peinado-Torrubia, P., Álvarez, R., Gandullo, J., Grewell, B. J., Figueroa, E., et al. (2020). Changes to the functional traits of phosphoenolpyruvate carboxylase following hybridization in C-4 halophytes. *Physiol. Plant* 169, 83–98. doi: 10.1111/ppl.13053

Geissler, N., Hussin, S., El-Far, M. M., and Koyro, H. W. (2015). Elevated atmospheric CO_2 concentration leads to different salt resistance mechanisms in a C_3 (Chenopodium quinoa) and a C_4 (Atriplex nummularia) halophyte. *Environ. Exp. Bot.* 118, 67–77. doi: 10.1016/j.envexpbot.2015.06.003

Gennidakis, S., Rao, S. K., Greenham, K., Uhrig, R. G., O'Leary, B., Snedden, W. A., et al. (2007). Bacterial- and plant-type phosphoenolpyruvate carboxylase polypeptides interact in the hetero-oligomeric Class-2 PEPC complex of developing castor oil seeds. *Plant J.* 52, 839–849. doi: 10.1111/j.1365-313X.2007.03274.x

He, M. X., Du, X. F., Chen, L., Lv, X. Y., and Lan, H. Y. (2013). Effects of salt, alternating temperature and hormone treatments on seed germination and seedling establishment of Suaeda aralocaspica (Chenopodiaceae) dimorphic seeds (in Chinese). *Chinese. J. Eco.* 32, 45–51.

Igawa, T., Fujiwara, M., Tanaka, I., Fukao, Y., and Yanagawa, Y. (2010). Characterization of bacterial-type phosphoenolpyruvate carboxylase expressed in male gametophyte of higher plants. *BMC Plant Biol.* 10:200. doi: 10.1186/1471-2229-10-200

Izui, K., Matsumura, H., Furumoto, T., and Kai, Y. (2004). Phosphoenolpyruvate carboxylase: a new era of structural biology. *Annu. Rev. Plant Biol.* 55, 69–84. doi: 10.1146/annurev.arplant.55.031903.141619

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. doi: 10.1002/j.1460-2075.1987.tb02730.x

Kai, Y., Matsumura, H., Inoue, T., Terada, K., Nagara, Y., Yoshinaga, T., et al. (1999). Three-dimensional structure of phosphoenolpyruvate carboxylase: a proposed mechanism for allosteric inhibition. *Proc. Natl. Acad. Sci. USA.* 96, 823–828. doi: 10.1073/pnas.96.3.823

Kandul, N. P., and Noor, M. A. (2009). Large introns in relation to alternative splicing and gene evolution: a case study of *Drosophila bruno-3*. *BMC Genet.* 10:67. doi: 10.1186/1471-2156-10-67

Koteyeva, N. K., Voznesenskaya, E. V., Berry, J. O., Cousins, A. B., and Edwards, G. E. (2016). The unique structural and biochemical development of single cell C_4 photosynthesis along longitudinal leaf gradients in Bienertia sinuspersici and Suaeda aralocaspica (Chenopodiaceae). *J. Exp. Bot.* 67, 2587–2601. doi: 10.1093/jxb/erw082

Koteyeva, N. K., Voznesenskaya, E. V., Roalson, E. H., and Edwards, G. E. (2011). Diversity in forms of C_4 in the genus Cleome (Cleomaceae). *Ann. Bot. London* 107, 269–283. doi: 10.1093/aob/mcq239

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lara, M. V., Chuong, S. D. X., Akhani, H., Andreo, C. S., and Edwards, G. E. (2006). Species having C_4 single-cell-type photosynthesis in the Chenopodiaceae family evolved a photosynthetic phosphoenolpyruvate carboxylase like that of Kranz-type C_4 species. *Plant Physiol.* 142, 673–684. doi: 10.1104/pp.106.085829

Lee, H. J., Kim, H. J., Seo, J., Na, Y. A., Lee, J., Lee, J. Y., et al. (2013). Estimation of phosphoenolpyruvate carboxylation mediated by phosphoenolpyruvate carboxykinase (PCK) in engineered *Escherichia coli* having high ATP. *Enzyme Microb. Technol.* 53, 13–17. doi: 10.1016/j.enzmictec.2013.04.001

Li, Y., Xu, J., Haq, N. U., Zhang, H., and Zhu, X. G. (2014). Was low CO_2 a driving force of C_4 evolution: Arabidopsis responses to long-term low CO_2 stress. *J. Exp. Bot.* 65, 3657–3667. doi: 10.1093/jxb/eru193

Liu, A., Xiao, Z., Li, M. W., Wong, F. L., Yung, W. S., Ku, Y. S., et al. (2019). Transcriptomic reprogramming in soybean seedlings under salt stress. *Plant Cell. Environ.* 42, 98–114. doi: 10.1111/pce.13186

Liu, Y. X., Maimaitijiang, T., Zhang, J. H., Ma, Y. L., and Lan, H. Y. (2020). The developmental enhancement of a C_4 system with non-typical C_4 physiological characteristics in Salsola ferganica (kranz anatomy), an annual desert halophyte. *Front. Plant Sci.* 11:152. doi: 10.3389/fpls.2020.00152

Lung, S. C., Yanagisawa, M., and Chuong, S. D. (2011). Protoplast isolation and transient gene expression in the single-cell C_4 species, Bienertia sinuspersici. *Plant Cell. Rep.* 30, 473–484. doi: 10.1007/s00299-010-0953-2

Mamedov, T. G., Moellering, E. R., and Chollet, R. (2005). Identification and expression analysis of two inorganic C- and N-responsive genes encoding novel and distinct molecular forms of eukaryotic phosphoenolpyruvate carboxylase in the green microalga Chlamydomonas reinhardtii. *Plant J.* 42, 832–843. doi: 10.1111/j.1365-313X.2005.02416.x

Martinoia, E., and Rentsch, D. (1994). Malate compartmentation-responses to a complex metabolism. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 45, 447–467. doi: 10.1146/annurev.pp.45.060194.002311

Masumoto, C., Miyazawa, S., Ohkawa, H., Fukuda, T., Taniguchi, Y., Murayama, S., et al. (2010). Phosphoenolpyruvate carboxylase intrinsically located in the chloroplast of rice plays a crucial role in ammonium assimilation. *Proc. Natl. Acad. Sci. USA.* 107, 5226–5231. doi: 10.1073/pnas.0913127107

McCarthy, J. E. (1998). Posttranscriptional control of gene expression in yeast. *Microbiol. Mol. Biol. Rev.* 62, 1492–1553. doi: 10.1128/mmbr.62.4.1492-1553.1998

McDonald, J. H. (2014). *Handbook of Biological Statistics*, 3rd Edn. Baltimore, Maryland: Sparky House Publishing.

Nomenclature Committee of the International Union Of Biochemistry (NC-IUB). (1979). Units of enzyme activity. *Eur. J. Biochem.* 97, 319–320. doi: 10.1111/j.1432-1033.1979.tb13116.x

Nomura, K., Sakurai, Y., and Dozono, M. (2020). Molecular cloning of novel-type phosphoenolpyruvate carboxylase isoforms in pitaya (Hylocereus undatus). *Plants (Basel)* 9:1241. doi: 10.3390/plants9091241

O'Leary, B., Park, J., and Plaxton, W. C. (2011a). 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. doi: 10.1042/BJ20110078

O'Leary, B., Fedosejevs, E. T., Hill, A. T., Bettridge, J., Park, J., Rao, S. K., et al. (2011b). Tissue-specific expression and post-translational modifications of plant- and bacterial-type phosphoenolpyruvate carboxylase isozymes of the castor oil plant, *Ricinus communis* L. *J. Exp. Bot.* 62, 5485–5495. doi: 10.1093/jxb/err225

O'Leary, B., Rao, S. K., Kim, J., and Plaxton, W. C. (2009). Bacterial-type phosphoenolpyruvate carboxylase (PEPC) functions as a catalytic and regulatory subunit of the novel class-2 PEPC complex of vascular plants. *J. Biol. Chem.* 284, 24797–24805. doi: 10.1074/jbc.M109.022863

Pan, L., Zhang, J., Chen, N., Chen, M., Wang, M., Wang, T., et al. (2017). Molecular characterization and expression profiling of the phosphoenolpyruvate carboxylase genes in peanut (Arachis hypogaea L.). *Russ. J. Plant Physiol.* 64, 576–587. doi: 10.1134/S1021443717040100

Park, J., Khuu, N., Howard, A. S., Mullen, R. T., and Plaxton, W. C. (2012). Bacterial- and plant-type phosphoenolpyruvate carboxylase isozymes from developing castor oil seeds interact in vivo and associate with the surface of mitochondria. *Plant J.* 71, 251–262. doi: 10.1111/j.1365-313X.2012.04985.x

Paulus, J., Schlieper, D., and Groth, G. (2013). Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution. *Nat. Commun.* 4:1518. doi: 10.1038/ncomms2504

Ping, C. Y., Chen, F. C., Cheng, T. C., Lin, H. L., Lin, T. S., Yang, W. J., et al. (2018). Expression profiles of phosphoenolpyruvate carboxylase and phosphoenolpyruvate carboxylase kinase genes in Phalaenopsis, implications for regulating the performance of crassulacean acid metabolism. *Front. Plant Sci.* 9:1587. doi: 10.3389/fpls.2018.01587

Rao, S. K., Reiskind, J. B., and Bowes, G. (2008). Kinetic analyses of recombinant isoforms of phosphoenolpyruvate carboxylase from *Hydrilla* verticillata leaves and the impact of substituting a C_4-signature serine. *Plant Sci.* 174, 475–483. doi: 10.1016/j.plantsci.2008.01.010

Rosnow, J. J., Edwards, G. E., and Roalson, E. H. (2014). Positive selection of Kranz and non-Kranz C_4 phosphoenolpyruvate carboxylase amino acids in

Suaedoideae (Chenopodiaceae). *J. Exp. Bot.* 65, 3595–3607. doi: 10.1093/jxb/eru053

Rosnow, J. J., Evans, M. A., Kapralov, M. V., Cousins, A. B., Edwards, G. E., and Roalson, E. H. (2015). Kranz and single-cell forms of C_4 plants in the subfamily Suaedoideae show kinetic C_4 convergence for PEPC and Rubisco with divergent amino acid substitutions. *J. Exp. Bot.* 66, 7347–7358. doi: 10.1093/jxb/erv431

Sánchez, R., and Cejudo, F. J. (2003). Identification and expression analysis of a gene encoding a bacterial-type phosphoenolpyruvate carboxylase from Arabidopsis and rice. *Plant Physiol.* 132, 949–957. doi: 10.1104/pp.102.019653

Schäffner, A. R., and Sheen, J. (1992). Maize C_4 photosynthesis involves differential regulation of phosphoenolpyruvate carboxylase genes. *Plant J.* 2, 221–232. doi: 10.1046/j.1365-313x.1992.t01-44-00999.x

Seoighe, C., Gehring, C., and Hurst, L. D. (2005). Gametophytic selection in *Arabidopsis* thaliana supports the selective model of intron length reduction. *PLoS Genet.* 1:e13. doi: 10.1371/journal.pgen.0010013

Sharpe, R. M., and Offermann, S. (2014). One decade after the discovery of single-cell C_4 species in terrestrial plants: what did we learn about the minimal requirements of C_4 photosynthesis? *Photosynth. Res.* 119, 169–180. doi: 10.1007/s11120-013-9810-9

Singh, J., Reddy, G. M., Agarwal, A., Chandrasekhar, K., Sopory, S. K., Reddy, M. K., et al. (2012). Molecular and structural analysis of C_4-specific PEPC isoform from Pennisetum glaucum plays a role in stress adaptation. *Gene* 500, 224–231. doi: 10.1016/j.gene.2012.03.018

Smith, M. E., Koteyeva, N. K., Voznesenskaya, E. V., Okita, T. W., and Edwards, G. E. (2009). Photosynthetic features of non-Kranz type C_4 versus Kranz type C_4 and C_3 species in subfamily Suaedoideae (Chenopodiaceae). *Funct. Plant Biol.* 36, 770–782. doi: 10.1071/FP09120

Sparkes, I. A., Runions, J., Kearns, A., and Hawes, C. (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat. Protoc.* 1, 2019–2025. doi: 10.1038/nprot.2006.286

Stefanovic, B., Lindquist, J., and Brenner, D. A. (2000). The 5' stem-loop regulates expression of collagen alpha1(I) mRNA in mouse fibroblasts cultured in a three-dimensional matrix. *Nucleic Acids Res.* 28, 641–647. doi: 10.1093/nar/28.2.641

Taylor, S. C., Nadeau, K., Abbasi, M., Lachance, C., Nguyen, M., and Fenrich, J. (2019). The ultimate qPCR experiment: producing publication quality, reproducible data the first time. *Trends Biotechnol.* 37, 761–774. doi: 10.1016/j.tibtech.2018.12.002

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016

Tu, J., Feng, L., Hong, Y., Liu, Q., Huang, X., and Li, Y. (2021). Prokaryotic expression of phosphoenolpyruvate carboxylase fragments from peanut and analysis of osmotic stress tolerance of recombinant strains. *Plants (Basel)* 10:365. doi: 10.3390/plants10020365

Tu, J., Gan, L., Feng, L., Yuan, L., and Li, Y. (2018). Bioinformatics analysis of PEPC gene family in Arachis duranensis (in Chinese). *J. Tropic. Subtropic. Bot.* 26, 107–115. doi: 10.11926/jtsb.3804

Ueno, Y., Imanari, E., Emura, J., Yoshizawa-Kumagaye, K., Nakajima, K., Inami, K., et al. (2000). Immunological analysis of the phosphorylation state of maize C_4-form phosphoenolpyruvate carboxylase with specific antibodies raised against a synthetic phosphorylated peptide. *Plant J.* 21, 17–26. doi: 10.1046/j.1365-313x.2000.00649.x

Voznesenskaya, E. V., Franceschi, V. R., and Edwards, G. E. (2004). Light-dependent development of single cell C_4 photosynthesis in cotyledons of Borszczowia aralocaspica (Chenopodiaceae) during transformation from a storage to a photosynthetic organ. *Ann. Bot.* 93, 177–187. doi: 10.1093/aob/mch026

Voznesenskaya, E. V., Franceschi, V. R., Kiirats, O., Freitag, H., and Edwards, G. E. (2001). Kranz anatomy is not essential for terrestrial C_4 plant photosynthesis. *Nature* 414, 543–546. doi: 10.1038/35107073

Wang, F. L., Liu, R. H., Wu, G. T., Lang, C. X., Chen, J. Q., and Shi, C. H. (2012). Specific downregulation of the bacterial-type PEPC gene by artificial microRNA improves salt tolerance in *Arabidopsis*. *Plant Mol. Biol. Rep.* 30, 1080–1087. doi: 10.1007/s11105-012-0418-6

Wang, L., Huang, Z. Y., Baskin, C. C., Baskin, J. M., and Dong, M. (2008). Germination of dimorphic seeds of the desert annual halophyte Suaeda aralocaspica (Chenopodiaceae), a C_4 plant without Kranz anatomy. *Ann. Bot.* 102, 757–769. doi: 10.1093/aob/mcn158

Wang, L., Ma, G. L., Wang, H. L., Cheng, C., Mu, S. Y., Quan, W. L., et al. (2019). A draft genome assembly of halophyte Suaeda aralocaspica, a plant that performs C_4 photosynthesis within individual cells. *Gigascience* 8:giz116. doi: 10.1093/gigascience/giz116

Wang, L., Wang, H. L., Yin, L., and Tian, C. Y. (2017). Transcriptome assembly in Suaeda aralocaspica to reveal the distinct temporal gene/miRNA alterations between the dimorphic seeds during germination. *BMC Genom.* 18:806. doi: 10.1186/s12864-017-4209-1

Wang, N., Zhong, X., Cong, Y., Wang, T., Yang, S., Li, Y., et al. (2016). Genome-wide analysis of phosphoenolpyruvate carboxylase gene family and their response to abiotic stresses in soybean. *Sci. Rep.* 6:38848. doi: 10.1038/srep38448

Waseem, M., and Ahmad, F. (2019). The phosphoenolpyruvate carboxylase gene family identification and expression analysis under abiotic and phytohormone stresses in Solanum lycopersicum L. *Gene* 690, 11–20. doi: 10.1016/j.gene.2018.12.033

Weissgerber, T. L., Milic, N. M., Winham, S. J., and Garovic, V. D. (2015). Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS Biol.* 13:e1002128. doi: 10.1371/journal.pbio.1002128

Westhoff, P., and Gowik, U. (2004). Evolution of C4 phosphoenolpyruvate carboxylase. Genes and proteins: a case study with the genus Flaveria. *Ann. Bot.* 93, 13–23. doi: 10.1093/aob/mch003

Xu, W., Ahmed, S., Moriyama, H., and Chollet, R. (2006). The importance of the strictly conserved, C-terminal glycine residue in phosphoenolpyruvate carboxylase for overall catalysis: mutagenesis and truncation of gly-961 in the Sorghum C_4 leaf isoform. *J. Biol. Chem.* 281, 17238–17245. doi: 10.1074/jbc.M602299200

Yasui, Y., Hirakawa, H., Oikawa, T., Toyoshima, M., Matsuzaki, C., Ueno, M., et al. (2016). Draft genome sequence of an inbred line of Chenopodium quinoa, an allotetraploid crop with great environmental adaptability and outstanding nutritional properties. *DNA Res.* 23, 535–546. doi: 10.1093/dnares/dsw037

You, L., Zhang, J., Li, L., Xiao, C., Feng, X., Chen, S., et al. (2020). Involvement of abscisic acid, ABI5, and PPC2 in plant acclimation to low CO_2. *J. Exp. Bot.* 71, 4093–4108. doi: 10.1093/jxb/eraa148

Yu, S. L., Pan, L., Yang, Q. L., Cheng, M. N., and Zhang, H. S. (2010). Identification and expression analysis of the phosphoenolpyruvate carboxylase gene family in peanut (Arachis hypogaea L.). *Agricul. Sci. China* 9, 477–487. doi: 10.1016/S1671-2927(09)60120-6

Zhao, J. F., Du, Y. W., Wang, G. H., Li, Y. F., Zhao, G. Y., Wang, Z. H., et al. (2020). Identification of PEPC genes from foxtail millet and its response to abiotic stress (in Chinese). *Acta Agronomica Sinica* 46, 700–711. doi: 10.3724/SP.J.1006.2020.94107

Zhao, Y., Guo, A., Wang, Y., and Hua, J. (2019). Evolution of PEPC gene family in *Gossypium* reveals functional diversification and GhPEPC genes responding to abiotic stresses. *Gene* 698, 61–71. doi: 10.1016/j.gene.2019.02.061

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership