What Matters for C₄ Transporters: Evolutionary Changes of Phosphoenolpyruvate Transporter for C₄ Photosynthesis

Ming-Ju Amy Lyu¹, Yaling Wang¹, Jianjun Jiang², Xinyu Liu¹, Genyun Chen¹ and Xin-Guang Zhu¹*

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

C₄ photosynthesis is a complex trait that evolved from its ancestral C₃ photosynthesis by recruiting pre-existing genes. These co-opted genes were changed in many aspects compared to their counterparts in C₃ species. Most of the evolutionary changes of the C₄ 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 phosphoenolpyruvate (PEP) transporter (PPT) during its recruitment from C₃ to C₄ photosynthesis. Our analysis showed that among the two PPT paralogs PPT1 and PPT2, PPT1 was the copy recruited for C₄ photosynthesis in multiple C₄ lineages. During C₄ 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 C₄ PPT1 as compared to C₃ PPT1, however, the PEP transporting function of PPT1 remained. We found that PPT1 of a Flaveria C₄ species recruited a MEM1 B submodule in the promoter region, which might be related to the increased transcript abundance of PPT1 in C₄ 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 C₄ function.

Keywords: C₄ photosynthesis, evolution, Flaveria, phosphoenolpyruvate transporter

HIGHLIGHTS

During the evolution of C₄ 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 C₄ lineages. Compared to its counterpart in C₃ species, PPT1 in C₄ 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 C₃ photosynthesis, C₄ photosynthesis has higher light, nitrogen, and water using efficiencies (Sage and Zhu, 2011). It achieves these superior properties through a CO₂ concentrating mechanism operating in a specialized leaf anatomical feature termed “Kranz anatomy” (Hatch, 1987). The CO₂ concentrating mechanism involves many enzymes and metabolite transporters, which together pump CO₂ from mesophyll cells (MC) to bundle sheath cells (BSC), creating a localized high CO₂ environment in the BSC around ribulose bisphosphate carboxylase/oxygenase (Rubisco). All genes required for the operation of C₄ photosynthesis are pre-existing in C₃ ancestors and play housekeeping functions (Aubry et al., 2011). The evolution of C₄ photosynthesis is therefore a process of recruiting and re-organizing pre-existing genes to fulfill new functions in C₄ photosynthesis (West-Eberhard et al., 2011; Burgess et al., 2016).

The specific modifications that have occurred on the genes recruited for C₄ photosynthesis are unknown. Comparisons of genes involved in C₄ photosynthesis in C₄ species and their counterparts in C₃ 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 C₄ 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 C₄ photosynthesis focus on enzymes directly related to the carbon shuttle enzymes, such as phosphoenolpyruvate (PEP) carboxylase (PEPC), phosphoenolpyruvate 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 CO₂ concentrating mechanism, C₄ plants recruited these carbon shuttle enzymes and required metabolite transporters, but also recruited a number of proteins responsible for transport of metabolites required for C₄ photosynthesis in both MC and BSC. In fact, compared to C₃ photosynthesis, the extensive usage of transporters is a major feature of C₄ photosynthesis. Quantitatively, to produce one molecule of triose phosphate for the synthesis of sucrose, only one transporter is needed in C₃ photosynthesis, while at least 30 metabolite transport steps are involved in the NADP-ME type C₄ photosynthesis (Weber and Von Caemmerer, 2010).

Furthermore, the flux through the transporters is much higher in C₄ (Wang et al., 2014). In C₃ 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 CO₂ uptake rate. In C₄ photosynthesis, however, the flux of metabolite transport between different compartments is higher than the photosynthetic CO₂ uptake rate due to the leakage of CO₂ from BSC to MC. Furthermore, C₄ plants usually have a higher leaf photosynthetic CO₂ uptake rate compared to C₃ plants. Therefore, C₄ 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 C₄ species than in C₃ species (Emms et al., 2016; Lyu et al., 2018; Moreno-Villena et al., 2018). Identifying the C₄ 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 C₄ metabolism and guide the engineering of C₄ metabolism into a C₃ metabolic background.

In this study, we aim to characterize the evolutionary changes of a C₄ metabolite transporter PPT, which transports PEP (Knappe et al., 2003), a substrate for the first step of C₄ acid formation in C₄ 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 C₄ leaf. Specifically, PEP is the substrate of PEPC, and its carboxylation represents the first step of CO₂ fixation in C₄ 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 C₃ and C₄ 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 C₃ and C₄ 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 C₃ species was constantly recruited for C₄ photosynthesis in multiple C₄ lineages. In an example study in Flaveria, we found that PPT1 from a C₄ species gained a MEM1 B submodule, which might contribute to the changes in transcriptional properties of PPT1 in C₃ species. Comparing PPT1 between C₄ and C₃ 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 C₄ photosynthesis.
MATERIALS AND METHODS

Construction of the Phosphoenolpyruvate 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 Phosphoenolpyruvate Transporters From Published RNA-Seq Data

High transcript abundance is suggested as a major feature of genes recruited to support C₄ 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 C₄ 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.
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 2.0.2) (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 C3 and C4 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 C3 species, and between BSC and MC in C4 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 Phosphoenolpyruvate 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 C4 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 C4 photosynthesis. Specifically, we studied this in five species representing four different photosynthetic types, i.e., F. robusta (C3), Flaveria sonorensis (C3–C4), Flaveria ramosissima (C3–C4), F. trinervia (C4), and Flaveria australasica (C4). 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 PPT1 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 C₃ and C₄ species, which were defined as sites that showed differences between C₃ and C₄ species, but that were conserved within C₃ species and also conserved within C₄ 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 C₄ 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 C₄ species using C₃ 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 C₄ 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 (\( \omega = 1 \), neutral). The maximum likelihood of the alternative hypothesis was calculated under this branch-site model with flexible \( dN/dS \) ratio (\( \omega > 1 \), 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 \( F. bidentis \) \( (C_3) \), and from \( de novo \) synthesized DNA for \( F. robusta \) \( (C_4) \) by Shanghai Personalbio LLC by PCR. A CDS with the 52-amino acid (52 aa) insertion deleted, i.e., \( \Delta Fbid-PPT1 \), was ordered from North American Services Center (NASC) (stock number N3156). Then, we introduced different \( Flaveria \) PPT1-GFP driven by the 35S promoter into \( cuc1-5 \) 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 1/2 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 Phosphoenol/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 \( M. polykarya \) and \( S. moellendorfii \). 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
The Evolution of Phosphoenolpyruvate Transporters in Transcript Abundance and Tissue Specificity of Expression Along the Emergence of C₄ 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 C₄ 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 C₃ species, PPT2 displayed higher transcript abundance than PPT1 except in the two C₃ species in *Heliotropium*, namely, *Heliotropium calcicola* (Hcal) and *Heliotropium karwinsky* (Hkar) (Figure 3A). The higher expression of PPT2 over PPT1 was also shown in C₃–C₄ species of *Mollugo* and *Neurachne*, as
well as in C3–C4 species in Heliotropium, i.e., Heliotropium filiforme (Hfil) and Flaveria i.e., F. sonorensis (Fson) (Figure 3A). During the transition from C3 to C4 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 C4 species.

Though PPT1 in leaves of C3 species did not show higher transcript abundance than PPT2, for the dicot C3 species, PPT1 showed higher transcript abundance than PPT2 in root (Figure 3B); the same pattern was also found in most monocot C3 species with Lasiacis sorgoidea (Lsor) as an exception (Figure 3B). In C4 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 C4 as compared to C3 species (Figure 3B). Therefore, PPT1, the copy recruited to support C4 photosynthesis, did not have higher expression levels than PPT2 in the leaf tissue of C4 plants; however, the gene had higher transcript abundance than PPT2 in root. During the evolution of C4 photosynthesis, the transcript abundance of PPT1 was decreased in root and increased in leaf, implying a major shift in tissue specificity.

Considering that C4 photosynthesis occurs in two cell types, which is a major evolutionary innovation, we further examined the changes in cellular specificity of PPT expression during C4 evolution. For this purpose, we compared the transcript abundance of PPT1 and PPT2 in BSC and whole leaf in one C3 species and that of BSC and MC in four C4 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 C4 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 C4 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 C3 species, i.e., A. thaliana, Oryza sativa, Solanum lycopersicum, and Glycine max, and two C4 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 C3 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 C3 species. In C4 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 Phosphoenolpyruvate Transporter During Evolution From C3 to C4 Photosynthesis

The mechanism by which PPT1 gained new expression patterns to support C4 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 C4 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 C4 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., C3 photosynthesis, F. robusta; type I C4–C4 species, F. sonorensis; type II C3–C4 species, F. ramosissima; and C4 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 C3 to C3–C4 intermediate to C4 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 C3 F. robusta, whereas significant up-regulation of PPT1 transcript abundance was observed at 2 h in C3–C4 species. In the C4 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 C4 evolution, PPT1 acquired novel 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 C3 to C4 species. Type I C3–C4 species showed the maximal PPT1 transcript abundance at 2 h under illumination, while the transcript abundance of PPT1 in type II C3–C4 and C4 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 C4 evolution. Specifically, in the C3 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 C3–C4 species and the C4 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 C4 species F. trinervia, in another C4 species F. australasica, the transcript abundance of PPT2 showed no significant up-regulation under the illumination. Therefore, during evolution
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.

**C₄ PPT1 Promoter Acquired MEM1 B Submodule But Not in C₃ and C₃–C₄ Species**

Changes in transcriptional responses to external stimuli can be driven by changes in gene regulatory mechanisms. We tested whether C₄ 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. ramosissima (C₃–C₄, II) and F. trinervia (C₄), and one copy in F. robusta (C₃) and F. sonorensis (C₃–C₄, 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 of the start codon) and that from F. ramosissima. The conserved 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* C₄ species, therefore, the same MEM1 B cis-element in the promoter of C₄ 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 C₄ 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 C₄ PPT1 to realize the specific expression of PPT1 in MC.

**Changes in the Physiological Functions of Phosphoenolpyruvate Transporters During C₄ Evolution**

We finally examined the functional changes of PPTs between C₃ and C₄ 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 C₃, C₃–C₄, C₄-like and C₄ 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 C₄ species but differ with those in C₃ species. The results show that PPT1 had more consistent amino acid modifications than PPT2 when the sequences from C₄ and C₃ species were compared.
Specifically, the amino acid sequence of PPT1 had 19 consistent amino acid modifications between C₃ and C₄ species; in contrast, PPT2 exhibited eight consistent amino acid modifications (Figure 6). To test whether these modifications were specific adaptations gained during evolution of C₄ photosynthesis, we performed a positive selection test in protein coding sequences of C₄ species against that of C₃ species in the genus Flaveria. PPT2 showed a signal of positive selection in C₄ species; however, the two predicted positive selected sites of PPT2 were neither C₄ specific nor C₄ consistent modifications (Figure S3). In contrast, PPT1 showed no signal of positive selection in C₄ species, suggesting that the consistent mutations observed in C₄ 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 C₄-like and C₄ 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 C₃ and C₄ species, we compared amino acid
sequences of PPT1 from other C₃ and C₄ species, including three C₃ species, namely, *A. thaliana*, *T. hassleriana* and *F. robusta*, and six C₄ 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* C₄ 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 C₄ evolution, we tested whether the function of PPT1 was conserved between C₃ and C₄ species using a genetic approach by expressing *Flaveria* PPT1 in a C₃ *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 C₃ species *Flaveria cronquistii*, two intermediate species *F. ramosissima* (C₃–C₄) and *F. palmeri* (C₄-like), and one C₄ species *F. bidentis* (C₄). 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.
(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 C₄ and C₃-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 (C₄) (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 C₄ 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 C₃ and C₄ 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 C₃ to C₄ 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 C₄ photosynthesis. The analysis showed that though PPT1 had lower transcript abundance in leaf compared to PPT2, it was recruited to support C₄ photosynthesis in multiple C₄ lineages. During C₄ evolution, PPT1 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 C₄ species. PPT1 also shows major changes in amino acid sequences during C₄ evolution, though they do not change the PEP transporting function. In this section, we discuss these findings in terms of their implications for C₄ photosynthesis.

**Factors Contributing to Recruitment of PPT1 Instead of PPT2 for C₄ Function**

**Potential of high expression:** studies on the evolution of C₄ genes identified a number of properties associated with the recruited paralogs for C₄ 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 C₄ photosynthesis without jeopardizing its native role in C₃ plants. In fact, gene duplication and neofunctionalization have been recognized as major factors contributing to evolution of C₄ 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 C₄-like and C₄ species in Flaveria clade A species, but this element was not observed in PPT1 from other C₃ species, such as G. gynandra and Z. mays (Figure S5). Accumulation of new mutations may contribute to the suitability of PPT1 for C₄ photosynthesis. The observation that PPT1 from either C₃ or C₄ Flaveria rescued the reticulate leaf phenotype of cue1 showed that the PEP transport function of PPT1 was not altered between C₃ and C₄ plants (Figure 7). In the future, more detailed functional studies of catalytic properties of PPT1 from different C₃ and C₄ 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 C₄ context, e.g., increased specificity or transport rate.

**Mechanisms Underlying Establishment of New Expression Patterns of PPT1**

**Gaining new cis-elements:** the function of C₄ photosynthesis requires high expression of required genes upon light induction in specific cell types (Sheen, 1999). Reports show that the expression levels of C₄ 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 correlated with 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.

**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 (31870124, 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. Daijin 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 classification of a cis-regulatory module for mesophyll-specific gene expression in the C4 dicot Flaveria trinervia. Plant Cell 19, 3391–3402. doi: 10.1093/pcp/pcp0332

Aubry, S. N., 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

Aubry, S., Kelly, S., Kumpers, B. M., Smith-Urna, 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 C4 photosynthesis in the grass genus Allotropis. J. Exp. Bot. 69, 1967–1980. doi: 10.1093/jxb/ery029

Blasing, O. E., Westhoff, P., and Svensson, P. (2000). Evolution of C4 phosphoenolpyruvate carboxylase in Flaveria, a conserved serine residue in the carboxyl-terminal part of the enzyme is a major determinant for C4-specific characteristics. J. Biol. Chem. 275, 27917–27923. doi: 10.1074/jbc.M009823199

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

Burgess, S. J., Granero-Moya, I., Grange-Guermonte, J., Boursnell, C., Terry, M. J., and Hibberd, J. M. (2016). Ancestral light and chloroplast regulation form modules between two Cleomaceae C3 and C4 plant species. Plant J. 86, 165–167. doi: 10.1111/tpj.130810

Christin, P. A., and Benard, G. (2009). Two independent C4 origins in Aristidioideae (Poaceae) revealed by the recruitment of distinct phosphoenolpyruvate carboxylase genes. Am. J. Bot. 96, 2234–2239. doi: 10.3732/ajb.0800111

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 cells by transcriptomic analysis. Plant Physiol. 160, 165–177. doi: 10.1104/pp.17.003810

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

Fischer, K., Kammerer, B., Gutensohn, M., Arbing, B., Weber, A., Hauler, R., 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 C3 to C4 photosynthesis. Plant Physiol. 155, 56–63. doi: 10.1104/pp.110.163308

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: 10.1093/tpc/167.9.1077

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

Grabher, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Armit, 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 C4 Ppc 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/tpi.14532

Hatch, M. D., and Osmond, C. B. (1976). Compartmentation and transport in C4 photosynthesis. Encyclopedia Plant Physiol. 3, 144–184. doi: 10.1007/978-3-642-66417-5_5

Hatch, M. D. (1987). C4 photosynthesis - a unique blend of modified biochemistry, anatomy and ultrastructure. Biochem. Et Biophys. Acta 985, 81–106. doi: 10.1016/S0300-4437(97)00009-9

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 C4 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 syntic 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 C4 grasses. Plant Physiol. 165, 62–75. doi: 10.1104/pp.14.238667

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

Knapp, S., Lottgent, T., Schneider, A., Voll, L., Flugge, U. I., and Fischer, K. (2003). Characterization of two functional phosphoenolpyruvate/phosphate translocator (PPT) genes in Arabidopsis--aHPT1 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

Kreb's, H. A. (1940). The citric acid cycle and the szent-gyorgyi cycle in pigeon breast muscle. Biochem. J. 34, 775–779. doi: 10.1046/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). gkh340

Lyu et al. Evolution of C4 Transporters
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. *Plant Physiol.* 164, 2231–2246. doi: 10.1104/pp.119.013384

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

Matassa, 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 C4 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

Paulas, 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). Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C4-related processes. *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/er1179

Sage, R. F., Sage, T. L., and Kocacinar, F. (2014). Photosorption and evaporation of C4 photosynthesis. *Annu. Rev. Plant Biol.* 63, 19–47. doi: 10.1146/annurev-arplant-042811-105511

Sheen, J. (1999). C4 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/btl346

Suyama, M., Torres, 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

Tautz, 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

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.119.013384

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

Wesseler, 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. *Annu. Bot.* 93, 13–23. doi: 10.1093/ambio/mch003

Williams, R. 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 C4 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.

**Copyright © 2020 Lyu, Wang, Jiang, Liu, Chen and Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.**