Plant Communications
Research article

CellPress
Partner Journal

Evolution of gene regulatory network of C_4 photosynthesis in the genus *Flaveria* reveals the evolutionary status of C_3-C_4 intermediate species

Ming-Ju Amy Lyu[1,3], Qiming Tang[1,2,3], Yanjie Wang[1,2,3], Jemaa Essemine[1,3], Faming Chen[1], Xiaoxiang Ni[1,2], 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]University of Chinese Academy of Sciences

[3]These authors contributed equally to this article.

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

https://doi.org/10.1016/j.xplc.2022.100426

ABSTRACT

C_4 photosynthesis evolved from ancestral C_3 photosynthesis by recruiting pre-existing genes to fulfill new functions. The enzymes and transporters required for the C_4 metabolic pathway have been intensively studied and well documented; however, the transcription factors (TFs) that regulate these C_4 metabolic genes are not yet well understood. In particular, how the TF regulatory network of C_4 metabolic genes was rewired during the evolutionary process is unclear. Here, we constructed gene regulatory networks (GRNs) for four closely evolutionarily related species from the genus *Flaveria*, which represent four different evolutionary stages of C_4 photosynthesis: C_3 (*F. robusta*), type I C_3-C_4 (*F. sonorensis*), type II C_3-C_4 (*F. ramosissima*), and C_4 (*F. trinervia*). Our results show that more than half of the co-regulatory relationships between TFs and core C_4 metabolic genes are species specific. The counterparts of the C_4 genes in C_3 species were already co-regulated with photosynthesis-related genes, whereas the required TFs for C_4 photosynthesis were recruited later. The TFs involved in C_4 photosynthesis were widely recruited in the type I C_3-C_4 species; nevertheless, type II C_3-C_4 species showed a divergent GRN from C_4 species. In line with these findings, a $^{13}CO_2$ pulse-labeling experiment showed that the CO_2 initially fixed into C_4 acid was not directly released to the Calvin–Benson–Bassham cycle in the type II C_3-C_4 species. Therefore, our study uncovered dynamic changes in C_4 genes and TF co-regulation during the evolutionary process; furthermore, we showed that the metabolic pathway of the type II C_3-C_4 species *F. ramosissima* represents an alternative evolutionary solution to the ammonia imbalance in C_3-C_4 intermediate species.

Key words: C_4 photosynthesis, evolution, *Flaveria*, GRN, metabolism

Amy Lyu M.-J., Tang Q., Wang Y., Essemine J., Chen F., Ni X., Chen G., and Zhu X.-G. (2023). Evolution of gene regulatory network of C_4 photosynthesis in the genus *Flaveria* reveals the evolutionary status of C_3-C_4 intermediate species. Plant Comm. **4**, 100426.

INTRODUCTION

The gene regulatory network (GRN) has been regarded as a useful tool for studying the genetic control of development and the evolution of phenotypes (Levine and Davidson, 2005; Peter and Davidson, 2011). The evolution of phenotypes is mainly driven by changes in gene expression (Jacob and Monod, 1961; King and Wilson, 1975; Wohlbach et al., 2009; Romero et al., 2012), and such changes are controlled by the GRN (Fisher and Franz-Odendaal, 2012; Thompson et al., 2015; McQueen and Rebeiz, 2020). Compared with a single effector gene (e.g., enzyme) or regulatory gene (e.g., transcription factor [TF]), the GRN provides higher statistical robustness and biological interpretability in explaining the link between phenotype and genotype (Mitra et al., 2013). Therefore, comparative studies of GRNs across species will help dissect the novel network circuitry formed during the evolution of different taxa, enabling us to reconstruct the evolutionary pathway of phenotypes

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

(Wohlbach et al., 2009; Romero et al., 2012; Necsulea and Kaessmann, 2014).

C_4 photosynthesis is a complex trait that evolved from the ancestral C_3 type (Sage, 2004). The function of C_4 photosynthesis relies on modifications to both metabolic pathways and anatomical structures (Hatch, 1987; McKown and Dengler, 2007). Compared with C_3 photosynthesis, C_4 photosynthesis requires more genes for CO_2 fixation, and these genes show preferential expression in either mesophyll cells (MCs) or bundle sheath cells (BSCs) (Slack and Hatch, 1967; Hatch, 1987). The specialized MCs and BSCs in C_4 species thus form so-called Kranz anatomy (Hatch, 1987). The metabolic and anatomical modifications also enable C_4 photosynthesis to have high light, water, and nitrogen use efficiencies (Zhu et al., 2008; Vogan and Sage, 2011), making C_4 an ideal target for engineering into C_3 crops to increase crop yield and productivity (Hibberd et al., 2008; Sage and Zhu, 2011). Despite the complexity of C_4 traits, C_4 photosynthesis has evolved independently on more than 70 occasions in angiosperms (Sage, 2017), representing a remarkable example of convergent evolution.

All the genes involved in C_4 photosynthesis were recruited from pre-existing genes of the ancestral C_3 species (Christin et al., 2009, 2013; Williams et al., 2012; Moreno-Villena et al., 2018). Therefore, the evolution of C_4 photosynthesis represents a reorganization of pre-existing parts and mechanisms, in which transcriptional regulation plays a pivotal role (Hibberd and Covshoff, 2010; Lyu et al., 2020; Schluter and Weber, 2020). The regulation of a number of C_4 genes has been reported (Gowik et al., 2004, 2017; Williams et al., 2016; Gupta et al., 2020). Genome-scale regulatory TFs for C_4 genes have also been uncovered, either by profiling of TF binding sites (Burgess et al., 2019) or from large-scale ChIP-seq data (Tu et al., 2020), which together have provided a reliable GRN of C_4 photosynthesis. It has been repeatedly documented that the regulation of C_4 genes also existed for their equivalents in C_3 plants (Brown et al., 2011; Kajala et al., 2012; Burgess et al., 2016). Evidence for this notion includes the finding that *cis*-regulatory elements (CREs) of C_4 genes that control cell specificity could be recognized by TFs in C_3 species and had the same cell specificity as in C_4 species (Brown et al., 2011; Gorska et al., 2019; Gupta et al., 2020). Therefore, it has been proposed that the evolution of C_4 photosynthesis was built on the pre-existing genetic network, with minor rewiring (Reyna-Llorens and Hibberd, 2017). However, it is unclear to what extent the regulation of C_4 genes is present in C_3 species, especially in intermediate species. Insights into these questions will inform attempts to engineer C_4 traits into C_3 crops.

To study the evolutionary trajectory of GRNs from C_3 to C_4 species, we used the genus *Flaveria* as a model system because this genus contains species with different photosynthetic types (Powell, 1978; McKown et al., 2005), in which the C_4 species evolved within the last 5 million years (Christin et al., 2011). In addition to being a young C_4 lineage, the metabolic and anatomical features of C_3, type I C_3-C_4, type II C_3-C_4, C_4-like, and C_4 species in this genus have also been extensively studied during the last two decades (Sage, 2004; McKown and Dengler, 2007; Sage et al., 2012; Mallmann et al., 2014). In fact, studies of this genus form the basis for many current theories on C_4 evolution, e.g., the step-wise evolution of C_4-related features (Rumpho et al., 1984; Moore et al., 1989; Sage et al., 2012), in which intermediate species represent transitory states of C_4 photosynthesis (Heckmann et al., 2013; Mallmann et al., 2014).

Here, we constructed genome-wide GRNs for four *Flaveria* species of different photosynthetic types (C_3, type I C_3-C_4, type II C_3-C_4, and C_4). The GRNs were developed based on *de-novo*-generated RNA-seq data using experimental conditions that have previously been reported to affect C_4 photosynthesis, such as low CO_2 (Sage, 2001; Li et al., 2014), high light (Ubierna et al., 2013), and exogenous abscisic acid (ABA) treatments (Ueno, 2001). With these GRNs, we address the following two questions: (1) how the GRN was re-reshuffled during the evolution of C_4 photosynthesisand (2) whether C_3-C_4 intermediate species represent a stage that precedes C_4 photosynthesis.

RESULTS

Transcript assembly and annotation of TFs and C_4 genes in *Flaveria* species

We constructed GRNs using next-generation sequencing (NGS) RNA-seq data from four *Flaveria species* representing four different photosynthetic types: *F. robusta* (C_3; hereafter, *Fro*), *F. sonorensis* (type I C_3-C_4; hereafter, *Fso*), *F. ramosissima* (type II C_3-C_4; hereafter, *Fra*), and *F. trinervia* (C_4; hereafter, *Ftr*) (Edwards and Ku, 1987) (Figure 1A). To confirm the photosynthetic types of the four species under our greenhouse conditions, carbon isotope ratios were analyzed by determining the $\delta^{13}C$ values of dried leaf material. The C_4 species showed less negative $\delta^{13}C$ values than the C_3 species, and the two C_3-C_4 species showed C_3-like $\delta^{13}C$ values (Figure 1B), consistent with previous reports (Monson et al., 1986; Edwards and Ku, 1987; Gowik et al., 2011) and suggesting that the four *Flaveria* species behaved as expected under our growth conditions. RNA-seq datasets were obtained from plants subjected to treatments, such as low CO_2 (Sage, 2001; Li et al., 2014), high light (Ubierna et al., 2013), and exogenous ABA application (Fischer et al., 1986; Duarte et al., 2019), that were previously reported to regulate the expression levels of $C_{4\ \text{photosynthesis-related genes}}$. In total, between 22 and 28 RNA-seq datasets were used to construct the GRN for each species (supplemental Table 1).

Recently, the draft genomes of four *Flaveria* species were reported (Taniguchi et al., 2021), including one of the species used in this study, i.e., *Fro* (C_3). Considering that the completeness of these genomes is relatively low (Taniguchi et al., 2021), some paralogs may be missing in their annotations. To obtain a better gene reference for the *Flaveria* genus and identify the C_4 versions of C_4 metabolic genes, we performed full-length transcript sequencing for one intermediate species that was not sequenced in (Taniguchi et al., 2021). Specifically, we performed transcriptomic sequencing for *Fra* using single-molecular real-time long read (LR) isoform sequencing (Iso-seq) from Pacific Biosciences (PacBio). Iso-seq helps to obtain full-length transcripts and effectively segregates different paralogs. We used the assembled gene sequences from

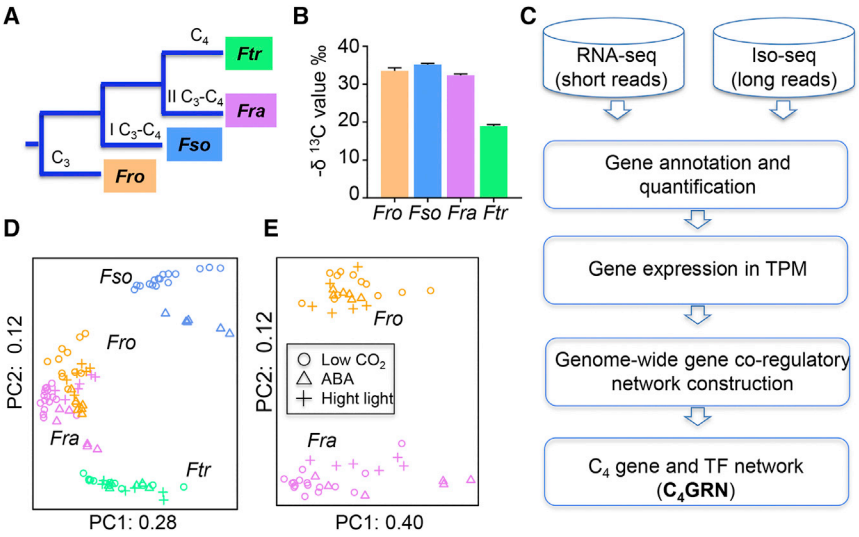


Figure 1. Sample information and flowchart of gene co-regulatory network construction.

(A) A schematic representation of the evolutionary relationships among four *Flaveria* species and their respective photosynthetic types, i.e., *Fro* representing C_3, *Fso* representing type I C_3-C_4 (I C_3-C_4), *Fra* representing type II C_3-C_4 (II C_3-C_4), and *Ftr* representing C_4.

(B) $\delta^{13}C$ values of the four species studied.

(C) Flowchart of GRN construction based on RNA-seq data. Full-length transcript sequences across the transcriptome were generated for *Fra* using Iso-seq to improve gene annotation for the *Flaveria* species. We constructed a genome-wide GRN for each species and then extracted the sub-network that contains C_4 genes and their positively co-regulatory TFs, which we termed the C_4GRN. **(D and E)** Principal component plots of all RNA-seq data sets based on gene expression profiling. *Fra* and *Fro* clustered together when all species were plotted simultaneously.

(D) however, they separated into different clusters when only their RNA-seq data were plotted.

(E) Abbreviations: *Fro*, *F. robusta*; *Fso*, *F. sonorensis*; *Fra*, *F. ramosissima*; *Ftr*, *F. trinervia*; GRN, gene co-regulatory network.

Iso-seq for *Fra*, together with protein sequences from (Taniguchi et al., 2021), as a gene reference dataset to annotate the assembled transcripts from the other three species (Figure 1C).

In total, 9.8 million Iso-seq LRs were obtained for *Fra* (C_3-C_4). Subsequently, we corrected the LRs with next-generation RNA-seq short reads (SRs), which resulted in 276 323 transcripts (supplemental Figure 1A). We then assembled the transcripts by combining both LRs and SRs. In total, 515 464 transcripts were found, among which 99 565 were predicted to be protein-coding transcripts with open reading frames no less than 300 bp in length. After removal of redundant sequences, 34 045 *Fra* genes were preserved. We also included 3292 gene sequences from (Taniguchi et al., 2021) that showed no targets in the 34 045 *Fra* genes for the final gene references (Blastp, E value < 10^{-10}), which resulted in 37 337 genes in total. The final gene references were annotated by searching for their orthologs in UniProt, which comprises 31 708 455 entries from Uniref50 and 558 898 entries from Swiss-Prot. By searching for orthologs at PlantTFDB (Jin et al., 2017), 3023 TFs from 57 families were annotated and characterized in our gene references. Among them, 252 TFs belonged to the bHLH family, the largest TF family in *Fra*, and 216 TFs belonged to the ERF family (supplemental Figure 1B). In line with our findings, bHLH and ERF are also the top two most abundant TF families in many other plants, such as *A. thaliana*, *Nicotiana tabacum*, *Oryza sativa*, and *Zea mays*, as documented at the PlantTFDB website (http://planttfdb.gao-lab.org).

Based on the *de-novo*-assembled transcripts from the SRs for the four species, between 66 234 and 80 154 transcripts were obtained. To confirm the accuracy of the assemblies, we further verified that approximately 80%–96.8% of the SRs could be mapped to the assembled transcripts (supplemental Data 1). Ninety-nine percent of the *de-novo*-assembled transcripts were predicted to encode proteins. In the end, approximately 60%–71% of these protein-coding transcripts possessed orthologs in our gene references (supplemental Table 2).

Principal component analysis (PCA) of the gene expression profiles based on the RNA-seq data (supplemental Data 2) revealed that *Ftr* (C_4) and *Fso* (type I C_3-C_4) differed from the other two species, with the first two components accounting for 40% of the total variance. *Fro* (C_3) and *Fra* (type II C_3-C_4) could not be separated based only on their first two PCA components (Figure 1D). However, if only these two species were analyzed using PCA, they could be distinguished into two clusters using the first two components, which reflected 52% of the total variance (Figure 1E).

Because most of the C_4 genes belong to multiple-gene families (Moreno-Villena et al., 2018), it was essential to determine the C_4 versions of the C_4 genes. Hence, we manually selected the C_4 versions of the C_4 genes based on two criteria: (1) the C_4 version gene should show higher transcript abundance compared with other paralogs; and (2) the C_4 version should exhibit higher transcript abundance in C_4 species than in C_3 species. The C_4 versions could be selected mainly based on the first criterion. For instance, two paralogs of β-carbonic anhydrase 1 (CA1) were annotated in the C_4 species *Ftr*, with one paralog (Gene00137) having a transcript abundance of approximately 20 000 TPM (transcripts per kilobase per million mapped reads) and the other (Gene23895) having a transcript abundance of approximately 1000 TPM. In addition, Gene00137 showed higher transcript abundance in the C_4 species *Ftr* than in the C_3 species *Fro* (supplemental Figure 2). Accordingly, Gene00137 was considered to be the C_4 version of CA1 in *Ftr*. *Ftr* uses a typical NADP-ME type C_4 photosynthesis, where only NADP-ME was upregulated in *Ftr* compared with *Fro*, whereas NAD-dependent malic enzyme (NAD-ME) and PEP carboxykinase showed comparable transcript abundances in all four species (supplemental Figure 2). In addition to the enzymes and transporters required for the NADP-ME pathway, we found a significantly higher transcript abundance of alanine aminotransferase (AlaAT) in *Ftr* than in *Fro*; however, aspartate aminotransferase (AspAT) did not show such a pattern across the four species. We therefore included this gene (AlaAT)

in the list of core C_4 metabolic genes. AlaAT is known to transfer ammonia from the BSC to the MC by the glutamate/2-oxoglutarate shuttle to maintain nitrogen homeostasis between the MC and BSC (Mallmann et al., 2014). All of the core C_4 metabolic genes showed higher transcript abundance in the C_4 species *Ftr* than in the C_3 species *Fro*, and such patterns were found to be conserved under the four experimental conditions used in this investigation (supplemental Figure 2). Among the different treatments, the 4-week low-CO_2 treatment had the greatest effect on C_4 gene expression in the four species (supplemental Figure 3).

To confirm the accuracy of transcript quantification via RNA-seq data processing, a qRT–PCR analysis was performed for 14 genes using the same RNA samples used for RNA-seq from plants grown under normal CO_2 conditions in the 4-week low-CO_2 experiment (supplemental Figure 4). We found an average Pearson correlation coefficient (PCC) of 0.78 between transcript abundance from RNA-seq and qRT–PCR. This result shows the reliability and accuracy of our transcript quantification via RNA-seq data processing. The expression patterns of core C_4 metabolic genes obtained from qRT–PCR were consistent with those obtained from the RNA-seq data (Figure 2A). Figure 2B depicts the proposed core C_4 metabolic pathway based on the identified C_4 genes in *Flaveria*.

The type I C_3-C_4 species *Fso* shows the highest similarity with C_4 species in terms of TFs co-regulated with C_4 genes

We constructed a genome-wide GRN based on the RNA-seq data for each species and then built a "C_4GRN," which retained only the C_4 genes and their positively co-regulated TFs. We designated the TFs involved in the C_4GRN as "C_4TFs" (supplemental Data 3). The genome-wide GRN of the type I C_3-C_4 species *Fso* contains the fewest genes (14 677 genes) and the most interactions (7 841 065 interactions). By contrast, that of the C_4 species *Ftr* contains 14 761 genes and 5 769 379 interactions (Figure 3A). Similarly, the C_4GRN of *Fso* includes 220 TFs, followed by those of *Ftr* and *Fro* (C_3) and *Fra* (type II C_3-C_4) with 196, 185, and 114 TFs, respectively (Figure 3A). More than half of the identified C_4TFs were found to be specific to individual species. For instance, 57.9% of the C_4TFs of *Fra* were not shared with any other species, and the percentage of species-specific C_4TFs in *Fro* was 69.2% (Figure 3B). Compared with the high-light and ABA conditions, the 4-week low-CO_2 treatment induced the most differentially expressed C_4TFs in all species, with the exception of *Fra* (supplemental Table 2).

In total, 36.2% of C_4TFs in the C_4 species *Ftr* overlapped with those of at least one of the three other species. The C_4GRN of *Fso* had 37 overlapping TFs with that of *Ftr*, displaying the highest degree of overlap with the C_4 species compared with *Fro* and *Fra* (Figure 3C and supplemental Figure 5). However, the co-regulated C_4 genes of these overlapping C_4TFs varied among the different *Flaveria* species. Five C_4 gene-TF co-regulations are shared between *Fro* and *Ftr*, involving five C_4 genes: NHD1, NADP-MDH, DiT2, BASS2, and PEPC-k. The same number of C_4 gene-TF co-regulations exists between *Fso* and *Ftr*. Notably, *Fra* and *Ftr* share seven C_4 gene-TF co-regulations, four of which involve PEPC1 (Figure 3C and supplemental Table 4).

In the C_4GRN of *Ftr*, PPDK-RP shows the largest number of co-regulated TFs (63 TFs), followed by PEPC1 with 45 TFs (Figure 3C). Interestingly, 74 C_4TFs (37.8%) from *Ftr* show the highest transcript abundance in *Fso* among the four species, even more than the number of C_4TFs that show the highest transcript abundance in *Ftr* (62 TFs) (Figure 3D).

We used an electrophoretic mobility-shift assay (EMSA) to verify the predicted co-regulation of TFs and C_4 genes in the C_4 species *Ftr*. We selected two C_4 enzymes with the maximum number of co-regulated TFs (PEPC1 and PPDK-RP) and two with the minimum number of co-regulated TFs (CA1 and NADP-ME4). For each enzyme, we selected one TF and cognate TF-binding site within 500 bp upstream of the start codon to perform EMSA. Our EMSA experiments provided evidence for the binding of WRKY4 to the promoter of PEPC1, the binding of WRYK75 to the promoter of PPDK-RP, the binding of DOF5.4 to the promoter of CA1, and the binding of OBP3 to the promoter of NADP-ME4 (Figure 3E), providing evidence for the accuracy of the predicted GRN.

Therefore, compared with C_3 and both types of intermediate species, the C_4 species *Ftr* recruited a larger number of TFs, many of which co-regulated PPDK-RP or PEPC1. *Fso* (type I C_3-C_4) shares the largest number of C_4TFs with the C_4 species, whereas *Fra* (type II C_3-C_4) shares the maximum number of C_4 gene and TF co-regulations with the C_4 species, four of which are associated with PEPC1.

C_4 genes were co-regulated with photosynthesis-related genes in all species except *Fra*

To gain insights into the functions of C_4TFs in each of the four species, we first investigated the enriched functions of the C_4TFs. C_4TFs from the four species were consistently enriched in glycolysis (supplemental Data 4). In *Fso* (type I C_3-C_4), C_4TFs were also enriched in histone acetyltransferase activity and the brassinosteroid-mediated signaling pathway (supplemental Data 4), whereas C_4TFs in *Fra* (type II C_3-C_4) were enriched in functions related to glycine metabolism, i.e., glycine dehydrogenase (decarboxylating) activity and glycine catabolic process. However, no C_4TF from any species had over-represented functions directly related to photosynthesis (supplemental Data 4).

We next investigated the related functions of C_4TFs in each species by examining the functions of their co-regulated genes. To this end, we first clustered the genome-wide GRN into multiple sub-networks (modules) and then studied the biological functions of the modules that were enriched in C_4TFs (Figure 4A). Specifically, 222, 363, 642, and 295 modules were identified in *Fro* (C_3), *Fso* (type I C_3-C_4), *Fra* (type II C_3-C_4), and *Ftr* (C_4), respectively (Figure 4B). We found only a single C_4TF-enriched module in *Fro*, *Fso*, and *Ftr*, whereas two C_4TF-enriched modules were characterized in *Fra* (Fisher's test, $P < 0.001$, BH [Benjamini–Hochberg] adjusted); however, only one of the C_4TF-enriched modules in *Fra* showed over-represented gene ontology (GO) functions (Figure 4B). We further examined the top 5 enriched GO terms of genes in the modules enriched in C_4TFs. *Fro* (C_3) exhibited a large proportion of GO terms related to transcriptional and post transcriptional processes, e.g., RNA

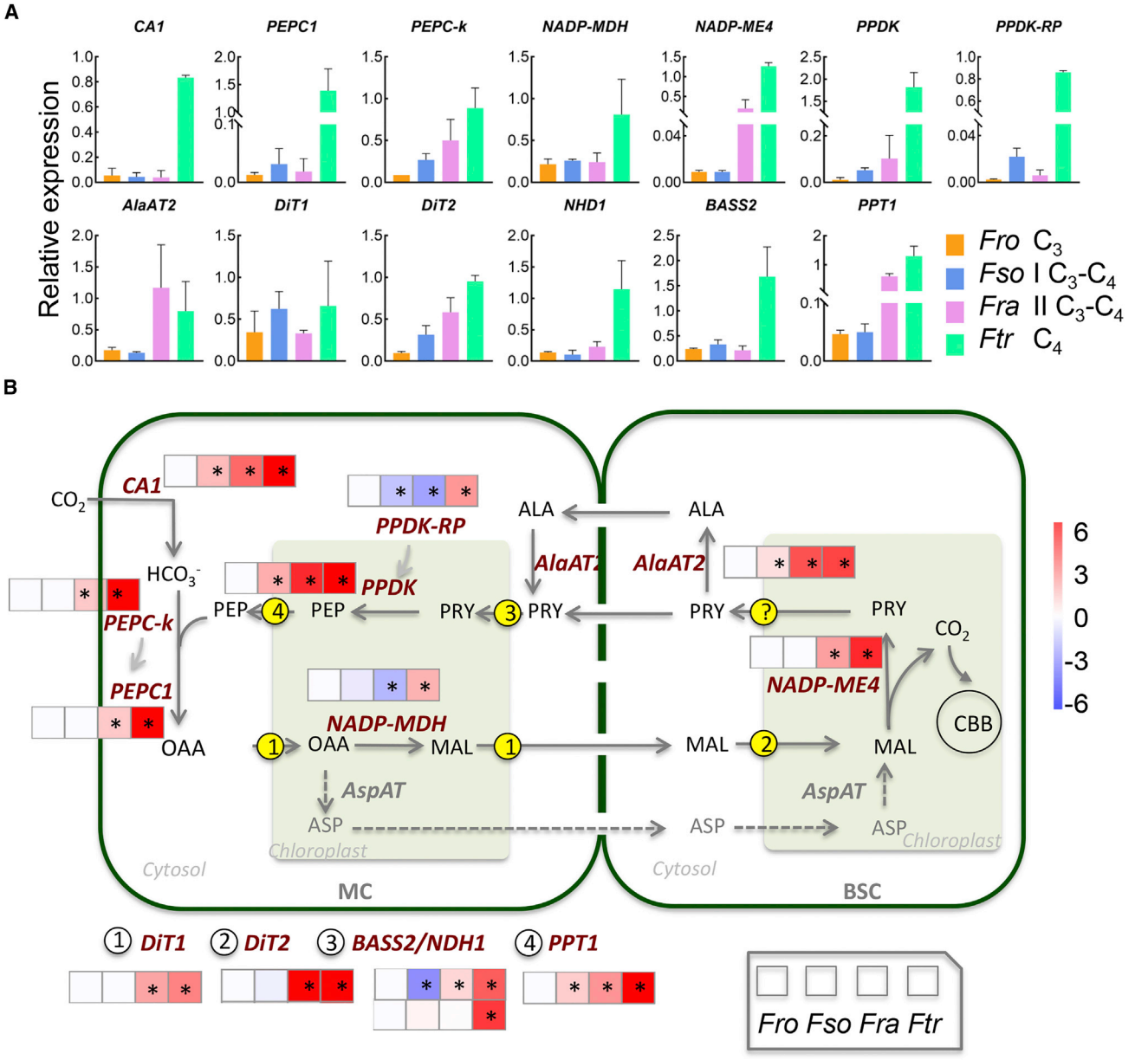


Figure 2. C_4 metabolic pathway and its involved genes.

(A) Histogram plots depicting the quantification of the determined C_4 versions of C_4 genes based on qRT–PCR analysis performed on leaf samples from plants grown at a normal CO_2 concentration (380 ppm) for 4 weeks.

(B) Diagram of the core C_4 pathway and the involved genes in *Flaveria*. *Ftr* has a typical NADP-ME type of C_4 photosynthesis. Heatmaps show the log2-transformed fold changes of transcript levels of C_4 genes assessed as transcripts per kilobase per million mapped reads (TPM) in each species compared with their counterparts in the C_3 species *Fro*. RNA-seq quantification of plants grown under normal CO_2 conditions from the 4-week low-CO_2 experiment. An asterisk (*) represents a significant difference at $P < 0.05$ and FC > 1.5. Aspartate aminotransferase (AspAT) was not included in this study, as the C_4 version failed to be clarified. Abbreviations: CA1, carbonic anhydrase 1; PEPC1, phospho*enol*pyruvate carboxylase 1; PEPC-K, PEPC kinase; NADP-MDH, NADP-dependent malate dehydrogenase; NADP-ME4, NADP-dependent malic enzyme 4; AlaAT2, alanine aminotransferase 2; PPDK, pyruvate orthophosphate dikinase; PPDK-RP, PPDK regulatory protein. Metabolite transporters are displayed in the circled numbers located on the membrane, with the consecutive numbers from 1 to 4 referring to DiT1 [dicarboxylate transporter 1], DiT2 [dicarboxylate transporter 2], BASS2 [bile acid sodium symporter 2]/NHD1 [sodium: hydrogen antiporter 1], and PPT1 [phosphate/phospho*enol*pyruvate translocator 1], respectively. Species abbreviations are as depicted in Figure 1.

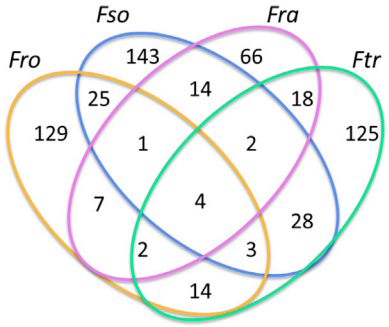
methylation, tRNA metabolic process and translation, and methyltransferase activity (Figure 4C). Notably, there was substantial overlap in enriched GO terms of modules enriched in C_4TFs between *Fso* (type I C_3-C_4), *Fra* (type II C_3-C_4), and *Ftr* (C_4). Many of these overlapping GO terms were involved in photosynthesis, such as carotenoid biosynthetic process, chloroplasts, and response to red light and far-red light (Figure 4C).

We next investigated the functions of the C_4 orthologous genes in different species by examining the over-represented GO terms in

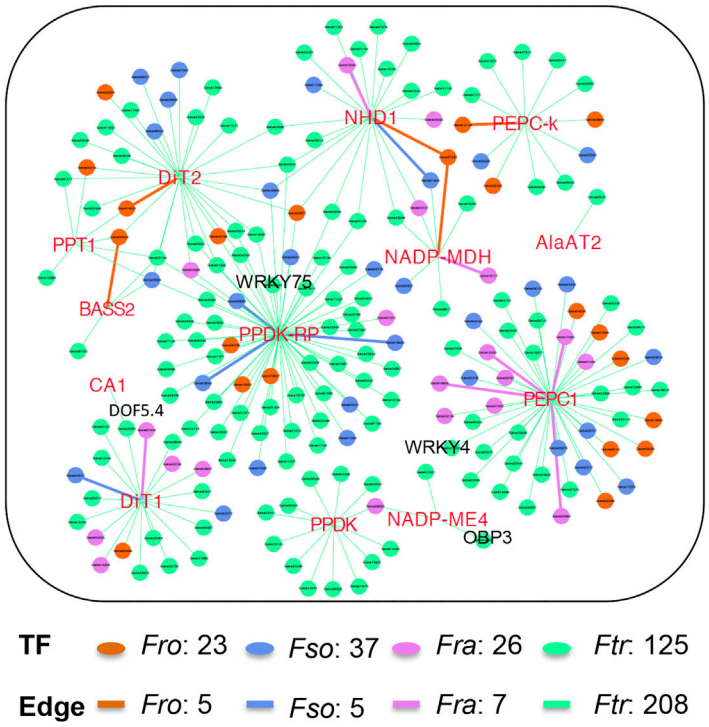
A

Species	*Fro* (C_3)	*Fso* (I C_3-C_4)	*Fra* (II C_3-C_4)	*Ftr* (C_4)
# Genes	15,005	14,677	17,784	14,761
# Interactions	6,394,639	7,841,065	7,330,127	5,769,397
# TF in C_4GRN	185	220	114	196

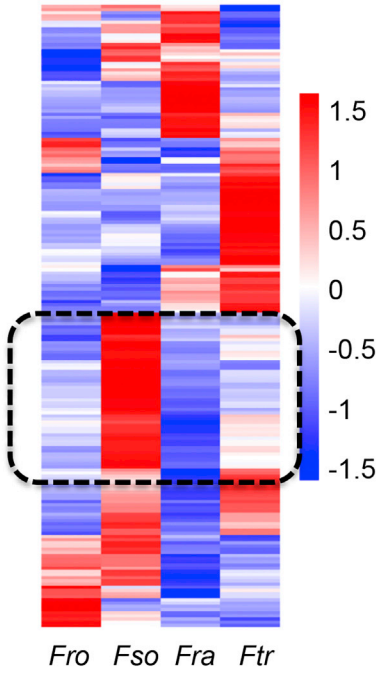
B



C



D



E

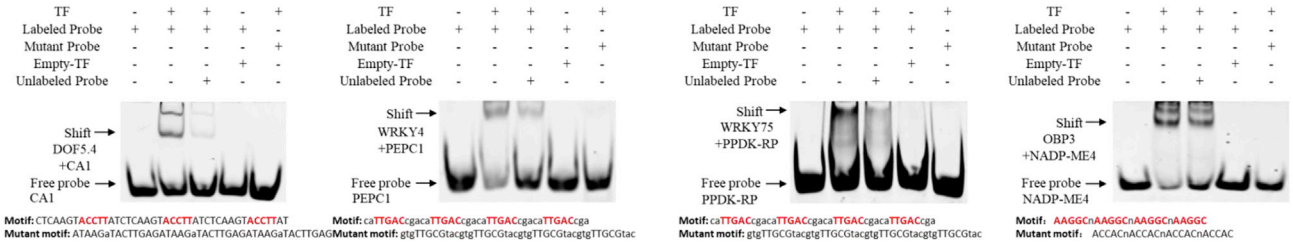


Figure 3. The type I C_3-C_4 species *Fso* shows the largest number of shared C_4TFs with the C_4 species *Ftr*.

(A) Basic information on the genome-wide GRNs and C_4GRNs of the four *Flaveria* species.

(B) Overlapping TFs in the C_4GRNs among the four investigated *Flaveria* species.

(C) The C_4GRN in the C_4 species *Ftr* is displayed, and only the interactions between the C_4 genes and their positively co-expressed TFs are shown. The C_4 genes are depicted in red font, and the TFs are shown in circles with four different colors. The overlapping TFs shared with the other three species are shown in different colors as depicted in the "TF" legend below the panel. The overlapping C_4 gene–TF interactions are displayed in the same colors used for the overlapping TFs, as reflected by "edge" in the legend.

(D) Expression patterns of the 196 co-regulated TFs with the C_4 genes in *Ftr* exhibited as a heatmap constructed using RNA-seq data from plants grown under normal CO_2 conditions from the 4-week low-CO_2 experiment. Colors represent relative transcript abundance, with red, white, and black corresponding to high, median, and low expression levels, respectively. The TFs with the highest transcript abundance in *Fso* are shown in the black dashed rectangle on the heatmap.

(E) EMSA was performed with the TFs labeled in **(C)** and cy5-labeled partial DNA sequences (probes) of their co-regulated C_4 gene promoters individually. Labeled probes were incubated with TF protein-10×His. Mutant-probe represents the mutant probe at predicted core binding sites.

(legend continued on next page)

modules that contained the maximum number of C_4 orthologous genes. In *Fso* (type I C_3-C_4), the module that was enriched in C_4TFs and the module that included the maximum number of C_4 orthologous genes were the same; this module included five C_4 genes, CA1, DiT2, NADP-ME4, NADP-MDH, and PPDK. Two modules that included the maximum number of C_4 orthologous genes were found in *Ftr* (C_4), one of which coincided with the module enriched in C_4TFs. A considerable fraction of the over-represented GO terms of modules that included the maximum number of C_4 orthologous genes were shared among *Fro*, *Fso*, and *Ftr*, and many were related to photosynthesis and chloroplasts (Figure 4D), such as photosynthesis, carotenoid biosynthetic processes, thylakoid, and chloroplast membrane and envelope.

Therefore, the orthologs of key C_4 metabolic genes in the genus *Flaveria* were already co-regulated with photosynthesis-related genes in the C_3 species, whereas the TFs specific to C_4 photosynthesis were mostly recruited during the most recent stages of C_4 evolution. Most of the enriched GO terms in the modules that included the maximum number of C_4 orthologous genes in *Fro* (C_3), *Fso* (type I C_3-C_4), and *Ftr* (C_4), were not shared with those in *Fra* (type II C_3-C_4) (Figure 4D), suggesting that the counterparts of C_4 genes in *Fra* may not function in photosynthesis.

Presence and absence of C_4TFs in two intermediate species compared with the C_4 species

We compared the GRNs between the two intermediate species and the C_4 species. First, 37 C_4TFs (supplemental Data 5) shared between *Fso* (type I C_3-C_4) and *Ftr* (C_4) were co-regulated with 12 of the 13 C_4 genes; PEPC1 was not co-regulated (Figure 5A), consistent with the lower transcript abundance of PEPC1 in *Fso* (Figure 2B and supplemental Figure 2). In the type II C_3-C_4 species *Fra*, 26 C_4TFs were found to be shared between *Fra* and *Ftr* and were co-expressed with 8 of 13 C_4 genes in *Fra* (Figure 5B). Interestingly, four of these 26 TFs were co-regulated with PEPC1 (Figure 3C and supplemental Table 3).

Increased expression of PEPC1 is a major feature of C_4 photosynthesis. A mesophyll expression module 1 (MEM1) located in the PEPC1 promoter region of the C_4 species *F. bidentis* was reported to control the mesophyll-specific high expression of PEPC1 (Gowik et al., 2004). We examined whether the MEM1 was present in the PEPC1 promoter region of *Fra* (type II C_3-C_4). Indeed, the MEM1 was of the C_4 type in *Fra* but of the C_3 type in *Fro* and *Fso* (Figure 5C and supplemental Data 6), consistent with the upregulation of PEPC1 expression in *Fra* compared with *Fro* and *Fso* (Figure 2B and supplemental Figure 2). These results thus demonstrate the coordinated recruitment of TFs and CREs during the evolution of PEPC1 expression regulation.

Taken together, these findings indicate that the mechanisms underlying the mesophyll-specific high expression of PEPC1 are not yet recruited in *Fso* (type I C_3-C_4). However, in *Fra* (type II C_3-C_4), the mechanisms that govern the cell-specific high expression of PEPC1 appear to be at least partially recruited. Nevertheless, five C_4 genes (PPDK, NADP-ME4, PPDK-RP, AlaAT2, and PPT1) were exempt from co-regulation by the shared C_4TFs between *Fra* and *Ftr* (C_4), further suggesting that the counterparts of some C_4 genes in *Fra* may not function in photosynthesis (Figure 4D).

Type II C_3-C_4 species *Fra* represents an alternative to C_4 photosynthesis in metabolism

The above results suggest that *Fra* may represent a parallel rather than a preceding stage to C_4 photosynthesis during evolution. Although 42% CO_2 is initially fixed into C_4-acid (malate or aspartate) (Rumpho et al., 1984; Vogan and Sage, 2011), the $\delta^{13}C$ of *Fra* (type II C_3-C_4) is comparable to that of C_3 species (Figure 1B) (Gowik et al., 2011); furthermore, the expression levels of other C_4 enzymes and transporters are much lower compared with C_4 species. We thus hypothesized that CO_2 fixed by PEPC in *Fra* may not be transported to C_4 metabolism. To test this hypothesis, we performed a $^{13}CO_2$ pulse-labeling experiment on *Fra* to trace the metabolic pathways of CO_2 after its fixation by PEPC. During $^{13}CO_2$ labeling, the time course of changes in isotopic abundance of primary metabolites was measured using high-performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS). We measured all 40 primary metabolites according to our previous dynamic systems model (Zhao et al., 2021) and obtained changes in ^{13}C of 27 metabolites. As for the other 13 metabolites, either they were not labeled by ^{13}C or we failed to quantify their abundance (supplemental Data 7). The 27 metabolites are involved in the Calvin–Benson–Bassham (CBB) cycle, photorespiration, C_4 photosynthesis, nitrogen assimilation, and the tricarboxylic acid (TCA) cycle, which are reactions related to C_4 evolution (Mallmann et al., 2014).

Among CBB-cycle-related metabolites such as ribulose-1,5-bisphosphate and phosphoglycerate, more than 50% were labeled after feeding with $^{13}CO_2$ for 100 s; approximately 10% of C_4 photosynthesis-related metabolites such as oxaloacetate (OAA) (11.7% ± 3.7% SD), malate (MAL) (6.8% ± 0.2% SD), aspartate (Asp) (14.8% ± 1.5% SD), and alanine (Ala) (12.1% ± 9.2% SD) were labeled within 100 s (supplemental Figure 6), consistent with a previous report (Monson et al., 1986). In addition, photorespiration-related metabolites such as glycine (10.1% ± 1.2% SD) and serine (16.7% ± 1.7% SD) were labeled with proportions of 10%–20% at 100 s (supplemental Figure 6), again in line with a previous report (Monson et al., 1986). Considering that the pool sizes of some metabolites are large, the changes in ^{13}C proportion are minor, but the absolute changes may be significant. We thus calculated the absolute changes of ^{13}C of each metabolite at each time point by multiplying the changed proportion by the total concentration (supplemental Data 7). Remarkably, we found that two metabolites involved in the TCA cycle, i.e., fumarate (FUM) (M0 = 93% ± 0.6% SD) and succinate (SUC) (M0 = 94.7% ±

Empty-TF represents GST-FLAG-10×His protein without TF protein. For competition analysis, the binding reaction was performed with addition of a two-fold amount of the corresponding unlabeled probe. + and − indicate the presence and absence of the respective component in the reaction system. Bands corresponding to DNA–protein complexes (shift) or free probes are indicated by arrows. Species abbreviations are the same as those in Figure 1.

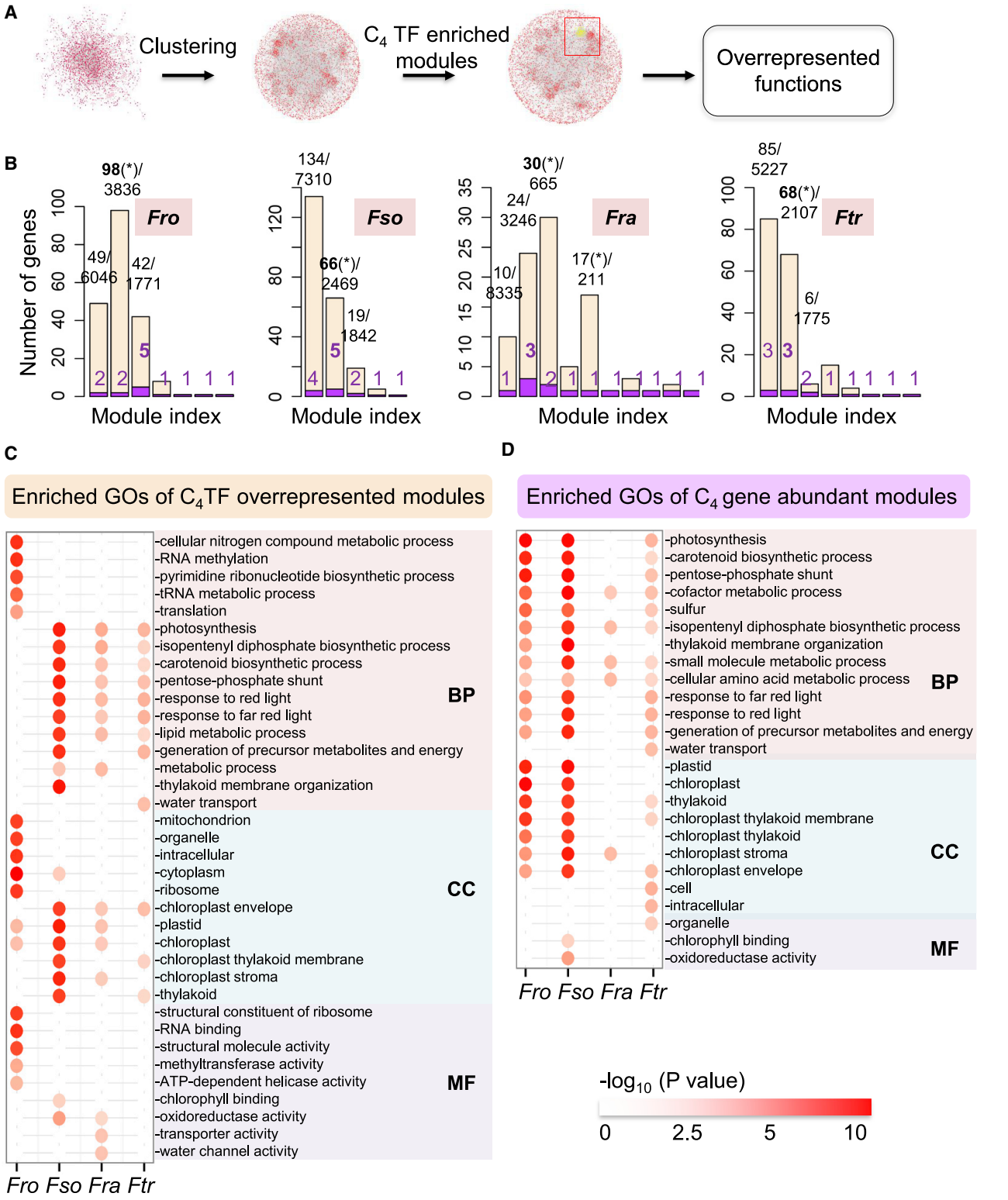


Figure 4. Biological functions of genes from modules enriched in C_4TFs and modules with the maximum number of C_4 genes.

(A) A predictive diagram displaying the analysis pipeline used to identify the potential functions of C_4TFs. A genome-wide GRN was first clustered into modules. The modules enriched in C_4TFs ($P < 0.001$, Fisher's test, BH adjusted) were then determined. Over-represented GO functions of genes in modules that were enriched in C_4TFs were calculated using Fisher's test ($P < 0.001$, BH adjusted).

(B) Bar plots show the number of C_4TFs and C_4 genes in the modules that encompass C_4 genes. Bars in beige represent the number of C_4TFs, and bars in purple represent the number of C_4 genes. The numbers given in gray and black are the numbers of C_4TFs (upper value) and the total numbers of genes

(legend continued on next page)

0.1% SD), were labeled after feeding with $^{13}CO_2$ for 100 s. Correspondingly, an absolute value of around 140 nmol mg Chl^{-1} of FUM was labeled at 100 s, more than that of OAA or MAL, although this only accounted for 2% of the total amount of FUM. However, the other two TCA metabolites, i.e., citrate (CIT) and α-ketoglutarate, showed few or no signs of labeling, even after 600 s (Figure 6A). Because our labeling time was short, we suggest that the ^{13}C-labeled FUM and SUC are derived from PEPC-fixed ^{13}C flux rather than from the TCA cycle. Our labeling results are consistent with a previously published ^{13}C labeling study in *Fra* (Borghi et al., 2022), in which a large proportion of photorespiratory metabolites, such as glycine and serine, were labeled 40 min after feeding with ^{13}C-labeled CO_2, whereas α-ketoglutarate showed no signs of labeling. The percentage of labeled malate was small, whereas the absolute amount was relatively large (Borghi et al., 2022).

In primary metabolism, OAA can be generated from PEP (catalyzed by PEPC). OAA may be used to synthesize CIT catalyzed by citrate synthase, MAL catalyzed by malate dehydrogenase (MDH), and Asp catalyzed by AspAT. As CIT was not labeled in our data (Figure 6A), we only considered the generation of MAL and Asp from OAA in *Fra*. Therefore, OAA is involved in three metabolic reactions, which are catalyzed by PEPC, MDH, and AspAT, respectively. For the metabolism of FUM, ^{13}C in FUM can be generated from SUC catalyzed by succinate dehydrogenase (SDH) through the TCA cycle, from MAL catalyzed by fumarate hydratase (FH), or from Asp catalyzed by aspartate ammonia lyase. Therefore, FUM is also involved in three reactions, which are catalyzed by SDH, FH, and aspartate ammonia lyase, respectively (Figure 6B).

We next used flux balance analysis to determine the flux and reaction directions of the local metabolic network involving OAA and FUM mentioned above by fitting ordinary differential equations (Methods). Our results showed that 41 nmol mg Chl^{-1} s^{-1} Asp and 72 nmol mg Chl^{-1} s^{-1} MAL were generated from OAA. Asp was also generated from FUM with a net flux of 262 nmol mg Chl^{-1} s^{-1}. For the measured FUM, 110 nmol mg Chl^{-1} s^{-1} was from SUC and, accordingly, the remaining 152 nmol mg Chl^{-1} s^{-1} was from MAL (152 = 262 − 110). For MAL, except for that from OAA, 80 nmol mg Chl^{-1} s^{-1} MAL was from other unknown reactions. Therefore, the inorganic carbon fixed by PEPC in *Fra* all went to Asp eventually, either directly through OAA or indirectly through OAA→MAL→FUM→Asp (Figure 6B). This suggested that there was no flux from Asp to PYR, as occurs in the C_4 pathway. The metabolic flux of MAL generated by MDH was insufficient for FUM generation; therefore, there was no flux from MAL to PYR for C_4 photosynthesis. As ammonia was fixed when Asp was generated from FUM, ammonia could therefore be transported from BSCs to MCs through Asp transport. Hence, the metabolic pathway in *Fra* may be an alternative solution in parallel to C_4 photosynthesis for redressing the ammonium imbalance between MCs and BSCs that is caused by restriction of glycine decarboxylase in the BSCs, which is termed the C_2 CO_2-concentrating mechanism (C_2 CCM) (Figure 6C).

DISCUSSION

Comparative analysis of GRNs among species can shed light on the driving force and the mechanisms that underlie their adaptation and evolutionary processes (Thompson et al., 2015). In this study, we provide a comprehensive survey of changes in the GRNs of four *Flaveria* species representing four different stages in C_4 evolution, i.e., *Fro* (C_3), *Fso* (type I C_3-C_4), *Fra* (type II C_3-C_4), and *Ftr* (C_4) (Edwards and Ku, 1987; Sage et al., 2012; Lyu et al., 2015). Considering that the transition from C_3 to C_4 photosynthesis spans an evolutionarily short divergence time (Christin et al., 2011), comparison among the GRNs of these four species could shed new light on the evolution of C_4 photosynthesis, not only on regulation but also on metabolic pathways (Figure 7). GRNs enable us to make a more comprehensive comparison of species of different photosynthetic types; however, it is important to note that the regulation of C_4 genes was only inferred based on co-expression, and further verification of the regulation of $C_{4\ genes\ by}$ TFs is required.

Co-option of the photosynthesis-related GRN facilitated multiple evolutions of C_4 photosynthesis

It has been reported that the origin of novel features arises from co-option of a pre-existing GRN that governs other phenotypes (Tomoyasu et al., 2009; Monteiro, 2012; McQueen and Rebeiz, 2020). The multiple independent occurrences of C_4 photosynthesis follow this rule. Burgess et al. (2016) demonstrated this by showing that the counterparts of key enzymes of C_4 photosynthesis were already co-regulated with photosynthesis in C_3 species. Specifically, they showed that the counterparts of C_4 genes in *A. thaliana* were controlled by either direct or indirect light, phytochrome HY5, phytochrome interacting factors, and chloroplast-to-nucleus signaling (Burgess et al., 2016). Here, comparative study of GRNs among different *Flaveria* species also showed that C_4 genes and their counterparts in non-C_4 species, with the exception of *Fra*, were co-regulated with photosynthesis-related genes (Figure 4D).

To achieve the upregulation and cell-specific expression of C_4 genes, the GRN was rewired and new TFs were recruited during the evolution of C_4 photosynthesis. Two interesting questions regarding GRN rewiring are: when were C_{4TFs} recruited, and what functions were governed by the newly recruited C_4TFs in non-C_4 species? Our results show that most of the C_4TFs were enlisted in C_4 species (Figures 3B and 3C), whereas a considerable portion of C_4TFs were recruited in the type I C_3-C_4 species *Fso* (Figure 3B), and these TFs were co-regulated with photosynthesis-related genes (Figure 4C). Co-opting the GRN that already controls photosynthesis by C_4 genes enabled these C_4 genes to coordinate with other components of photosynthesis. Therefore, our study provides evidence for the hypothesis that co-option of the photosynthetic GRN has facilitated

(bottom value), and the numbers in purple are the numbers of C_4 genes. An asterisk (*) indicates that the module was significantly enriched in the C_4TFs. Over-represented GO terms of genes from modules enriched in C_4TFs **(C)** and genes from modules encompassing the maximum number of C_4 genes **(D)** are shown. Abbreviations: BP, biological process; CC, cellular component; MF, molecular function. The species abbreviations are as depicted in Figure 1. BH, Benjamini–Hochberg.

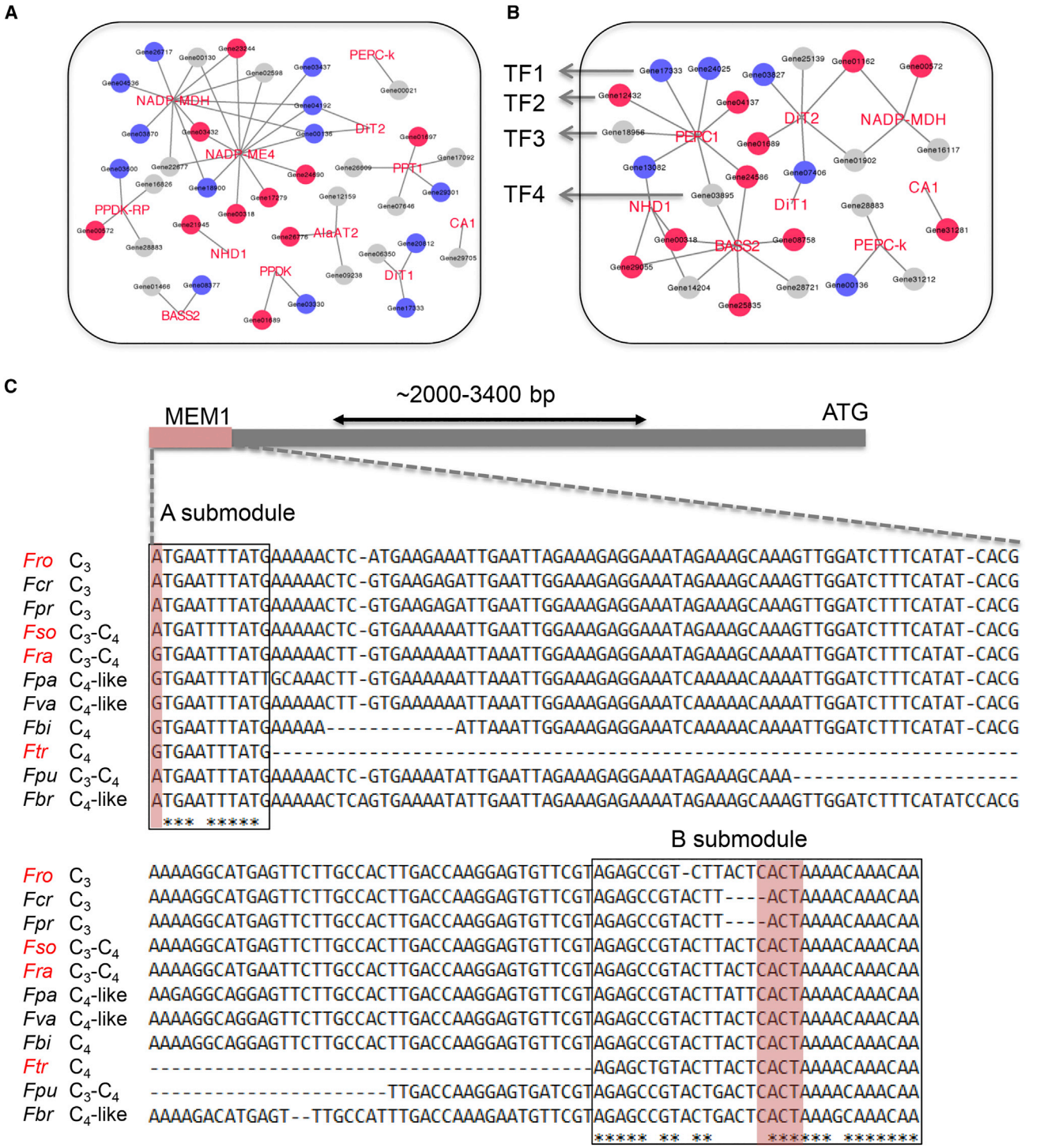


Figure 5. Overlapped C_4GRNs between the two intermediate species and the C_4 species.

Overlapped C_4GRN between the C_4 species *Ftr* and the type I C_3-C_4 species *Fso* **(A)** and between *Ftr* and the type II C_3-C_4 *Fra* **(B)** are shown. The C_4 genes are depicted in red font. TFs are shown in circles, with red/blue/gray colors representing TFs that showed increased/decreased/similar transcript abundance in the C_4 species compared with the C_3-C_4 species (edgeR, $P < 0.05$ and FC > 1.5). Four TFs co-regulating PEPC1 were found to be conserved between *Fra* and *Ftr* and were labeled TF1–4; they are from the MYB_related, NF-YA, GRAS, and DBB TF families, respectively.

(C) The MEM1 sequences of the PEPC1 promoters from 11 *Flaveria* species, including the 4 species used in this study (labeled in red) and 7 species referenced in (Akyildiz et al., 2007). The A and B submodules are highlighted in boxes. Asterisks show identical nucleotides in the two modules. Red zones indicate the single nucleotide difference in the A submodule and the required tetranucleotide CACT in the B submodule. The full promoter sequences of *Fro*, *Fso*, *Fra*, and *Ftr* are available in supplemental Data 5. Enzyme abbreviations are the same as those given in Figure 2. Species abbreviations: *Fro*, *F. robusta*; *Fcr*, *F. cronquistii*; *Fpr*, *F. pringlei*; *Fso*, *F. sonorensis*; *Fra*, *F. ramosissima*; *Fpa*, *F. palmeri*; *Fva*, *F. vaginata*; *Fbi*, *F. bidentis*; *Ftr*, *F. trinervia*; *Fpu*, *F. pubescens*; and *Fbr*, *F. brownii*.

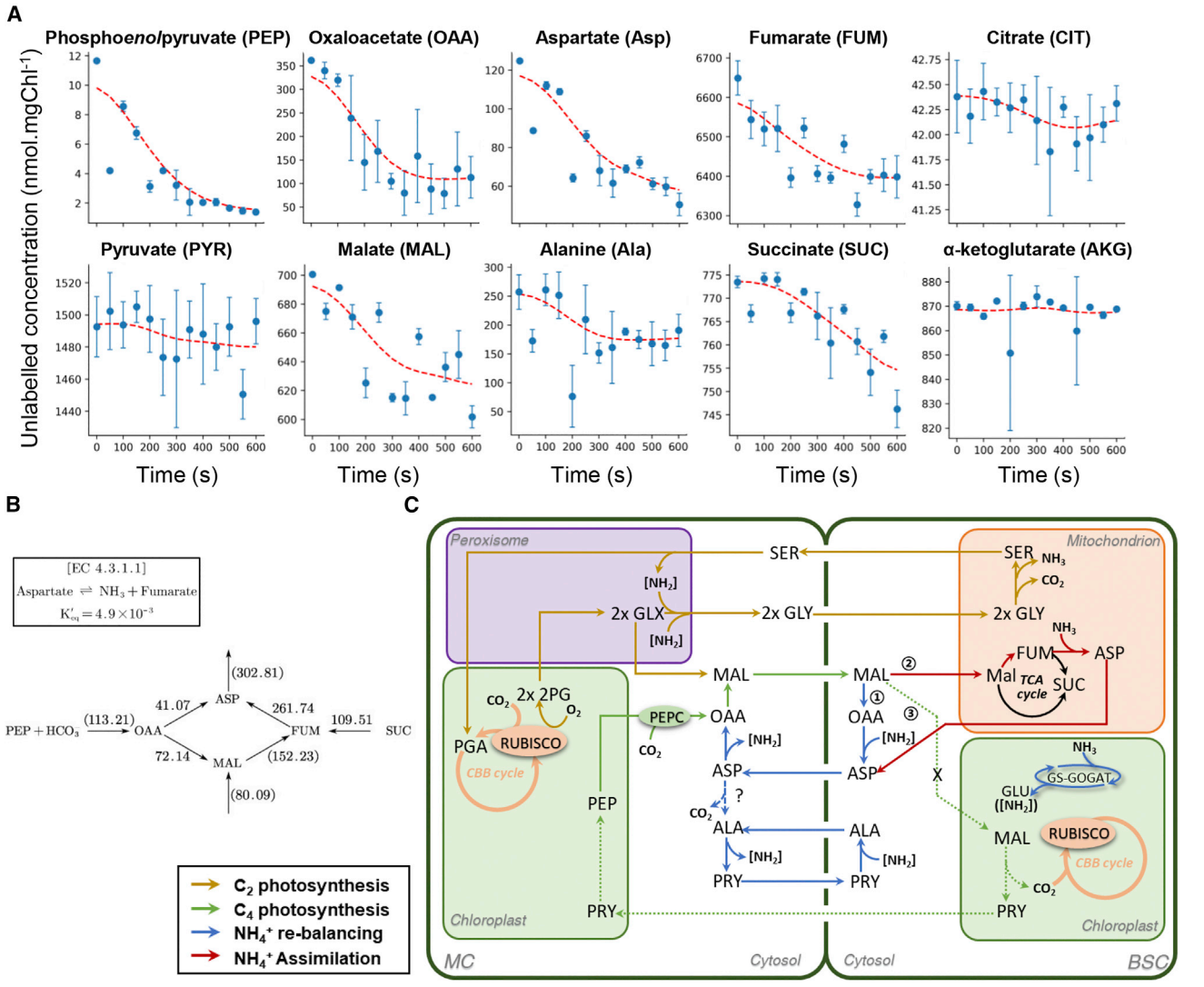


Figure 6. The metabolic pathway in *Fra* represents an alternative solution to C_4 photosynthesis in balancing ammonia between mesophyll and bundle sheath cells.

(A) Dot plots show time-course changes in unlabeled concentrations of metabolites during $^{13}CO_2$ labeling in *Fra* (type II C_3-C_4). Data shown are mean ± SD (n = 4 independent experiments). Dotted lines show smoothed values based on the measured data.

(B) PEPC-coupled carbon metabolic flux in *Fra*. Values represent the fitted net flux (nmol mg Chl^{-1} s^{-1}), and values in brackets indicate the metabolic flux calculated according to the flux of related metabolic fluxes.

(C) Schematic view of the proposed three-carbon coupled with four-carbon ammonia re-balancing pathways in *Fra*. The reaction directions were determined according to total net fluxes. C_2 CCM is indicated with orange arrows, the C_4 cycle is indicated with green arrows, and the ammonia-transporting pathway from the bundle sheath cell (BSC) to the mesophyll cell (MC) is indicated with blue arrows. Ammonia is transported from the MC to the BSC as a result of C_2 CCM and is transported back to the MC through three-carbon alanine and four-carbon aspartate. The carbon skeleton of aspartate is derived from malate by two pathways, i.e., ① oxaloacetate (blue arrows) and ② fumarate (red arrows). In C_4 photosynthesis, malate is decarboxylated in chloroplasts of the BSC (indicated by ③), and CO_2 is released for the Calvin–Benson–Bassham cycle. However, this reaction may not occur in *Fra*, as suggested by our flux analysis (indicated in dotted green lines with an "X"). Abbreviations: GLX, glyoxylate; Gly, glycine; Ser, serine; 2PG, 2-phosphoglycolate; PGA, phosphoglycerate; PEP, phosphoenolpyruvate; OAA, oxaloacetate; MAL, malate; PYR, pyruvate; Ala, alanine; Asp, aspartate; Glu, glutamine; FUM, fumarate; SUC, succinate; MC, mesophyll cell; BSC, bundle sheath cell; CBB, Calvin–Benson–Bassham cycle; RUBISCO, ribulose-1,5-bisphosphate-carboxylase/oxygenase.

multiple evolutions of C_4 photosynthesis (Reyna-Llorens and Hibberd, 2017).

Fso recruited a large number of C_4TFs in two ways

Fso (type I C_3-C_4) shares the highest number of C_4TFs (37) with *Ftr* (C_4) (Figure 3B), regulating 12 of the 13 core C_4 metabolic genes (Figure 5). This finding agrees with the current notion that C_3-C_4 species, which perform C_2 CCM, might represent a metabolic pre-condition for the emergence of C_4 photosynthesis, as many of the relevant enzymes and transporters in C_2 CCM overlap with those of C_4 photosynthesis (Mallmann et al., 2014). Moreover, *Fso* represents the first stage that shows major modifications during the evolution of C_4 photosynthesis

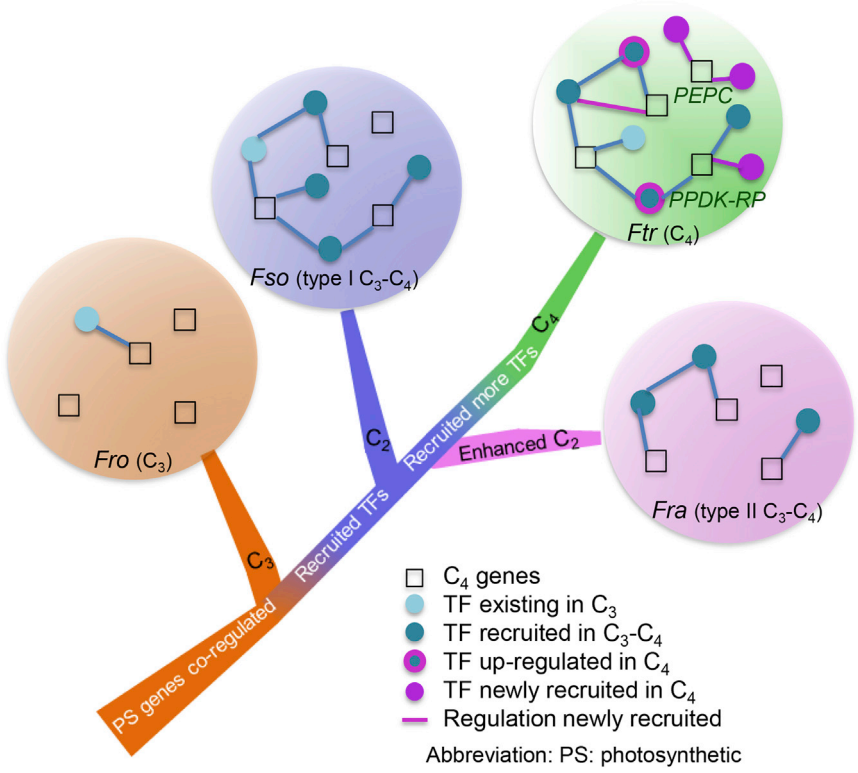


Figure 7. A proposed model for C_4GRN- and C_4-related metabolic evolution in *Flaveria*.
Two key stages are proposed during the evolution of the C_4GRN in the genus *Flaveria*. The first stage was the transition from C_3 photosynthesis to C_2 CCM. C_4 gene orthologs were co-regulated with photosynthesis-related genes in the C_3 species *Fro*and during the evolution of the type I C_3-C_4 species *Fso*, a large number of C_4TFs were recruited, but not for all C_4 genes. The second stage was the transition from C_2 CCM to C_4 photosynthesis as performed in *Ftr*, in which a complete C_4GRN was established. This C_4GRN involved newly recruited TFs, enhanced transcript abundance of C_4TFs that had already been recruited in type I C_3-C_4 species, and formation of new regulatory relationships among pre-existing TFs and other C_4 genes. PEPC and PPDK-RP show more co-regulated TFs than do other core C_4 metabolic genes in the C_4 species *Ftr*. The C_4GRN of the type II C_3-C_4 species *Fra* is different from those of type I C_3-C_4 and C_4 species. C_2 CCM was further enhanced in *Fra* compared with *Fso*. The metabolic pathway in *Fra* may represent an alternative, parallel solution to C_4 photosynthesis for balancing ammonia between mesophyll and bundle sheath cells. Abbreviations: PEPC, phosphoenolpyruvate carboxylase; PPDK-RP, pyruvate orthophosphate dikinase-regulatory protein.

(Lyu et al., 2021). The large number of recruited C_4TFs in *Fso* may reflect recruitment of pre-existing CREs into C_4 genes and/or rewiring of pre-existing TFs to gain new functions for the recruited CREs (Brown et al., 2011; Reyna-Llorens et al., 2018).

In this study, we found that 30 of the recruited 37 C_4TFs in *Fso* are not represented in the C_4GRN of the C_3 species *Fro* (2 + 28 = 30, Figure 3B). These TFs may have been recruited as a result of enlisting pre-existing CREs into C_4 genes from other genomic regions. For example, PEPC1, CA3 (named CA1 in our study), and PPT1 harbor MEM1 and MEM1-like motifs that determine MC expression specificity in C_4 *Flaveria* species (Akyildiz et al., 2007; Gowik et al., 2017; Lyu et al., 2020). CA and PPDK share a MEM2 motif in the 5′ UTR in the C_4 species *Gynandropsis gynandra* that controls expression specificity in the MC (Williams et al., 2016). Furthermore, a comparable CRE was recruited into PEPC in three grass species (Gupta et al., 2020). These CREs in C_4 genes can be recognized by TFs in leaves of C_3 species and generate the same cell-specific expression as in C_4 species (Ku et al., 1999; Nomura et al., 2000; Fukayama et al., 2001; Akyildiz et al., 2007). For example, in the case of NAD-ME, the CRE that controls BS-specific expression in the C_4 species *G. gynandra* shows the same cell specificity in the C_3 species *A. thaliana* (Brown et al., 2011). Altogether, these studies indicate that a pre-existing CRE repository in the ancestral C_3 species could be recruited into C_4 genes for C_4 photosynthesis (Burgess et al., 2019); this has occurred rather frequently during the conversion from C_3 to the type I C_3-C_4 stage.

We also found that 23 of the 37 C_4TFs exhibited higher transcript abundance in *Fso* than in the C_3 species *Fro* (Figure 3D), suggesting that enhanced expression levels of these TFs also occurred during C_4 evolution. In this context, it has also been reported that CREs from PPDK, CA (Kajala et al., 2012), and NAD-ME (Brown et al., 2011) from C_3 species can drive C_4-type cell-specific accumulation of GUS in C_4 species. Therefore, the TFs in C_4 species may have increased their expression or altered their cellular expression patterns to use pre-existing CREs of C_4 genes and thereby express their target genes. Studies of this possibility have already revealed some TFs that acquired conserved MC- or BSC-preferential expression in two divergent C_4 lineages, *G. gynandra* and *Z. mays* (Aubry et al., 2014).

The metabolic pathways of *Fra* represent a parallel solution to C_4 photosynthesis for balancing ammonia between MCs and BSCs

The current notion of C_4 evolution in the genus *Flaveria* is that the C_3-C_4 intermediates are classified into two types, with *Fra* (type II C_3-C_4) reflecting the type II C_3-C_4 intermediate, which shows greater C_4-ness than the type I C_3-C_4 species (McKown and Dengler, 2007; Vogan and Sage, 2011). Therefore, it is striking to find that *Fra* shows less C_4-ness in its C_4GRN than *Fso* does, as reflected by the absence of a number of critical C_4 genes in the overlapping C_4GRNs between *Fra* and *Ftr* (Figure 5A). Furthermore, in *Fra*, the genes in the C_4 gene-enriched modules display no enriched GO functions for photosynthesis (Figure 4D). This raises the question of whether *Fra* constitutes a transitional state from type I C_3-C_4 to C_4 photosynthesis. Compared with the type I C_3-C_4 intermediate, *Fra* indeed shows an increased vein density (McKown and Dengler, 2007), increased cyclic electron transport capacity (Vogan and Sage, 2011; Nakamura et al., 2013), lower CO_2 compensation point

(Ku et al., 1983, 1991; Nakamoto et al., 1983; Rumpho et al., 1984), and higher expression levels of some C_4 genes (Gowik et al., 2011; Mallmann et al., 2014). In addition, 42% of CO_2 can initially be fixed into malate/aspartate in *Fra* (Rumpho et al., 1984). However, *Fra* also shows C_3 characteristics in a number of other C_4-related features, such as CO_2 assimilation rate (Ku et al., 1991) and $\delta^{13}C$ and water use efficiency (Vogan and Sage, 2011). In particular, the lack of change in $\delta^{13}C$ implies that there is no substantial increase in the C_4 flux of *Fra* compared with the C_3 *Fro.*

What may underlie the increased level of C_4 acid synthesis noted previously (Rumpho et al., 1984)? We did find that four C_4TF and PEPC1 co-regulations were conserved between *Fra* and *Ftr* (Figure 5B). *Fra* includes the C_4-type MEM1 motif in its promoter region (Figure 5C), which should underlie increased transcript abundance of PEPC1 in *Fra* compared with *Fro* and *Fso* (Gowik et al., 2011; Mallmann et al., 2014) and higher protein abundance of PEPC in MCs than in BSCs in *Fra* (Taniguchi et al., 2021). The enhanced function of PEPC1 in MCs can support the increased CO_2 fixation in this species (Rumpho et al., 1984); however, the fixed CO_2 is not released to ribulose-1,5-bisphosphate-carboxylase/oxygenase in BSC as occurs in typical C_4 metabolic pathways. If that were the case, an increase in $\delta^{13}C$ would be expected (Farquhar et al., 1989).

If there is no increased C_4 metabolic flux in *Fra*, what might underlie its enhanced C_4-ness? We hypothesized that the high C_4-ness of *Fra* most probably resulted from an increased C_2 CCM following the reallocation of glycine decarboxylase from MCs to BSCs. An increased PEPC1 function would result in a low CO_2 concentration in the MCs, which could increase photorespiratory fluxes. Consistent with this notion, *Fra* has relatively higher expression levels of photorespiratory genes compared with both C_3 and C_4 species (Mallmann et al., 2014). The increased photorespiratory flux could enhance the glycine shuttle from the MC toward the BSC, where glycine would be converted to serine, thereby leading to an increased release of CO_2 and ammonia in the BSC. The $^{13}CO_2$ pulse-labeling results further suggest that CO_2, either from the atmosphere or released from glycine decarboxylation, can be fixed by PEPC for the formation of Asp; this step may occur in the BSC (Figures 5B and 5C). Following its formation, Asp can be transported back to the MC and thus solve the imbalance of ammonia between MCs and BSCs caused by enhanced C_2 CCM (Figure 6C). Therefore, the metabolic features of *Fra* represent an alternative solution to C_4 photosynthesis in balancing ammonia between MCs and BSCs. This also explains why the "presumed" transition from type II C_3-C_4 *Fra* to the common ancestor of the C_4 and C_4-like species required the largest modifications during the evolution of C_4 photosynthesis (Lyu et al., 2021). Considering that *Fra* shows a C_3-like CO_2 assimilation rate (Ku et al., 1991), such a metabolic pathway may be an alternative to C_4 photosynthesis in balancing ammonia but may not be an ideal target for C_4 engineering.

A previous systems-level model of the C_4 adaptive evolutionary trajectory showed that *Flaveria* intermediate species represent transitory states of C_4 photosynthesis, whereas *Fra* deviates from the trajectory from C_3 to C_4 photosynthesis (Heckmann et al., 2013). In this regard, earlier studies showed that the intermediate species *Alloteropsis semialata* of the grass family (Lundgren et al., 2016), C_3-C_4 intermediate species from *Moricandia* such as *M. arvensis* and *M. suffruticosa* (Schluter et al., 2017), and the C_3-C_4 species *Salsola divaricata* (Lauterbach et al., 2017) displayed δ^{13C} values comparable to those of C_3 species, which may be due to the leakiness of the BS (Peisker, 1986). However, we suggest that the metabolic status of C_3-C_4 intermediate species deserves further detailed studies.

MATERIALS AND METHODS

Plant materials and growth conditions

Seeds of *Fro* (C_3) and *Fra* (C_3-C_4, type II) were provided by Prof. Peter Westhoff (Heinrich Heine University, Germany). Seeds of *Fso* (C_3-C_4, type I) and *Ftr* (C_4) were obtained from Prof. Rowan Sage (University of Toronto, Canada). For the low-CO_2 experiment, plants were grown in a growth chamber with a photosynthetic photon flux density (PPFD) of 200 μmol m^{-2} s^{-1}, a temperature of 22°C ± 2°C, 70% relative humidity, and a 16-h light/8-h dark photoperiod. The CO_2 concentration used for the low-CO_2 treatment was 100 ppm, and the ambient (or control) CO_2 level was approximately 380 ppm. Plants for the ABA treatment were grown in the phytotron of the Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences (CAS), with a PPFD of 500 μmol m^{-2} s^{-1}, a temperature of 25°C ± 1°C, 70% relative humidity, and a 16-h light/8-h dark photoperiod. The ABA was first dissolved in distilled water to prepare a 40-μM ABA solution, which was subsequently sprayed onto mature leaves (second or third pair of leaves counting from the top) of 1-month-old plants. During the ABA treatment, the mature leaves of control plants were sprayed with distilled water only (no ABA). For the high-light experiment, plants were grown in a phytotron at the Partner Institute of Computational Biology, CAS, with a PPFD of 1400 μmol m^{-2} s^{-1} for the high-light condition and 500 μmol m^{-2} s^{-1} for the control condition. The high-light condition was achieved by supplementing the light with a lab-made light-emitting diode light source. The light spectrum used in both control and high-light conditions is depicted in supplemental Figure 7. For all experiments, plants were watered twice a week, and fertilizer was applied weekly after being dissolved to a concentration of 1‰ (w/w) (N:P:K = 20:20:20) to avoid plant growth limitation due to nutrient depletion and to obtain healthy plants for use in subsequent measurements.

PacBio Iso-seq and NGS of RNA

To obtain better annotations for *Flaveria* genes and understand their potential biological functions, we sequenced full-length transcripts across the *Fra* transcriptome using Iso-seq (supplemental Figure 1). RNA samples isolated from four different tissues (leaf, stem, root, and flower) were mixed and subjected to the sequencing process to obtain a blueprint of the transcripts. For leaf samples, the youngest fully expanded leaf from 1-month-old plants was used, which was usually located on the second or the third pair of leaves counting from the top. For stem samples, the mid-segment of a stem between the root node and that of the leaf from 1-month-old plants was used. For root samples, fine roots of 1-month-old plants were used. The compacted soil on the roots was removed by floating and gently shaking the roots in water before root samples were prepared for RNA extraction. For flower samples, three to five inflorescences were collected from the same plant. Each type of sample was collected from three different plants. All samples were cut and immediately frozen in liquid nitrogen. Total RNA was then isolated from each tissue following the protocol of the PureLink RNA Kit (Thermo Fisher Scientific, USA). The isolated RNA from the four tissues was mixed in equal amounts for the next step of the analysis. cDNA was synthesized using a SMARTer PCR cDNA Synthesis Kit (Clontech, USA) and then amplified using the KAPA HiFi PCR Kit (Roche, USA). The amplified

products were classified into different fragments according to their sizes (1–2 k, 2–3 k, or >3 k) using BluePippin Size Selection (Sage Science, USA) and were subsequently used to construct separate libraries with the SMRTbell Template Prep Kit 1.0 (PacBio, USA). Libraries were sequenced on the PacBio Sequel platform.

For NGS of the RNA data, the youngest fully expanded leaf (usually situated on the second or third pair of leaves, counting from the top) was obtained from 4-week-old (ABA, high light, and 2-week low CO_2), 6-week-old (4-week low CO_2), and 6.5-month-old (6-month low CO_2) plants of all species studied. The chosen leaves were cut, immediately frozen in liquid nitrogen, and stored at −80°C until further processing. Total RNA was isolated according to the procedure described above. NGS of RNA was performed on the Illumina platform in paired-end mode with a read length of 150 bp.

Transcript assembly, annotation, and quantification

Iso-seq LRs from *Fra* were corrected using proovread (Hackl et al., 2014) with RNA-seq SRs obtained from NGS. In this process, the SRs were mapped to LRs, and sequencing errors of the LRs were corrected by the SR consensus approach (Hackl et al., 2014). The rectified Iso-seq reads were then used for transcript assembly with IDP-denovo (Fu et al., 2018). To define (or characterize) a reference on the gene scale, we used BLAST (version 2.2.31+) (Camacho et al., 2009) to detect transcript groups with high similarity, i.e., those that passed an E value threshold of 10^{-10} and a sequence identity threshold of 90%. The longest transcript in the transcript group was retained as the *Fra* gene reference.

In addition to *Fra* gene references, we also included gene sequences from (Taniguchi et al., 2021) that were not included in the *Fra* gene references. Specifically, we first predicted the open reading frames of the transcript sequences of the four *Flaveria* species provided in Taniguchi et al. (2021), i.e., *Fro*, *F. bidentis*, *F. floridana*, and *F. brownii*, using OrfPredictor (Min et al., 2005) with default parameters. We then kept those genes that showed no hits in the *Fra* gene references using BLAST (version 2.2.31+) (Camacho et al., 2009) with an E value threshold of 10^{-10}, resulting in 973 genes from *Fro*, 802 genes from *F. bidentis*, 687 genes from *F. floridana*, and 830 genes from *F. brownii*, which together with the *Fra* gene references were used as the final gene reference dataset.

The biological functions of the final reference genes were annotated by searching for their orthologs in the UniProt database (https://www.uniprot.org/). In this study, we used a combination of two datasets. The first was the UniProt reference clusters of Uniref50, in which each entry represents a cluster of sequences with at least 50% sequence identity and 80% overlap. Further details for this first dataset are available at https://www.uniprot.org/help/uniref. The second dataset was the Swiss-Prot manually annotated and reviewed database. Orthologous genes were identified using BLAST (version 2.2.31+) (Camacho et al., 2009) with an E value threshold of 10^{-5}. TFs were annotated by searching for orthologs of all plant TFs at PlantTFDB (http://planttfdb.gao-lab.org) (Riano-Pachon et al., 2007) using BLAST (version 2.2.31+) (Camacho et al., 2009) with an E value threshold of 10^{-5} (supplemental Figure 1).

The transcript sequences of *Fro*, *Fso*, and *Ftr* were assembled *de novo* based on SRs, and *de novo* assembly was also performed for the SRs from *Fra*. The Trinity tool (version 2.8.4; Grabherr et al., 2011) was used for *de novo* assembly with default parameters, except that the transcript length was constrained to be no less than 300 bp. The *de-novo*-assembled transcripts were annotated by searching the orthologs from the final gene references using BLAST (version 2.2.31+) (Camacho et al., 2009) with an E value threshold of 10^{-5}.

Transcript abundances were quantified as TPM using the RSEM package (Li and Dewey, 2011), in which Bowtie 2 (version 2.3.4.3) (Langmead and Salzberg, 2012) was used to map the SRs to the *de-novo*-assembled transcripts and other parameters remained at default settings (supplemental Figures 2 and 3). Genes that were differentially expressed between the control and treated samples and between the different species were identified using the edgeR package (Robinson et al., 2010) with a P value < 0.05 and fold change > 1.5.

qRT–PCR analysis

To verify the gene expression measurements based on *de novo* SR assembly and confirm the accuracy of our computational analysis, a qRT–PCR analysis was performed on the same RNA samples used for NGS that came from leaf samples of plants grown under normal CO_2 conditions (380 ppm) in the 4-week low-CO_2 experiment. Between 0.2 and 0.5 μg RNA was used for reverse transcription to synthesize first-strand cDNA with Superscript II Reverse Transcriptase (TransGen Biotech, Beijing). The qRT–PCR mixture was prepared following the instructions of Unicon qPCR SYBR Green Master Mix (Yeasen, China). In brief, the mixture of cDNA, buffer, and enzyme was pipetted into Hard-Shell PCR 96-well plates (Bio-Rad, USA) and then covered with Microseal "B" sealing film (Bio-Rad). qRT–PCR was performed on the Bio-Rad CFX Connect system (Bio-Rad). Each 20-μl PCR reaction volume contained 10 μl SYBR Green PCR Master Mix, 4 μl deionized H_2O, 1 μl primers (0.5 F + 0.5 R), and 5 μl cDNA. The amplification reaction was initiated with a pre-denaturing step at 95°C for 3 min, followed by 39 cycles of denaturing at 95°C for 10 s, annealing at 60°C for 20 s, and extension at 72°C for 30 s. Gene expression relative to that of the housekeeping gene, Actin7, was calculated with the $2^{-\Delta\Delta CT}$ method as described previously (Livak and Schmittgen, 2001). Data were processed using Bio-Rad CFX Maestro software (Bio-Rad). For each gene, two technical replicates and three biological replicates were performed. The primers used for qRT–PCR analysis are listed in supplemental Table 5.

Isolation of the 5′ flanking sequence from the PEPC1 gene of four species

The 5′ flanking sequence of the PEPC1 gene of the four *Flaveria* species (*Fro*, *Fso*, *Fra*, and *Ftr*) was amplified from total DNA by PCR. The DNA was extracted following the protocol of the TIANquick Midi Purification Kit (TIANGEN Biotech, Beijing). The primers used for DNA amplification by PCR are listed in supplemental Table 6. The PCR products were sequenced at Sangon Biotech (Shanghai, China).

GRN construction

GRNs were constructed with the CMIP software package (Zheng et al., 2016), which implements a path consistency algorithm based on the conditional mutual information (CMI) (Zhang et al., 2012). The CMIP package uses a table of gene expression values as input. The method first computes a zero-order network based on mutual information, then eliminates indirect relationships by considering the CMI, thereby leading to a first-order network. We also implemented the P value determination in the CMIP package based on a permutation test. In other words, we shuffled the expression level of each gene 2000 times, generating 2000 null datasets. Thereafter, the P value was defined as N/2000, where N was the number of datasets with a calculated CMI greater than that calculated from the original dataset. We used a P value of 0.001 as the cutoff during genome-wide GRN reconstruction. The C_4GRN was developed by retaining only the co-regulatory network of the core C_4 metabolic genes and their positively co-regulated TFs (hereafter, C_4TFs) from the genome-wide GRN.

Special measures were taken to ensure the development of a reliable and accurate GRN. In the original CMIP package, an adequate cutoff for the gene–gene partial PCC was automatically estimated by fitting the curve of interaction number in response to different cutoffs based on an exponential function, and the cutoff was determined as the slope inflection of the fitted curve (Zheng et al., 2016). Given that we focused mainly on

C_4GRN construction, we slightly modified this computational procedure. Specifically, we compared the relative abundance of all TF families (TF frequency) of the C_4GRN generated under different PCC cutoffs, increasing from 0.1 to 0.9 with a step size of 0.1. Consequently, we found that TF frequency maintained a high degree of similarity under different PCC cutoffs ranging from 0.1 to 0.7 (supplemental Figure 8). We calculated the similarity of TF frequencies under two adjacent PCC cutoffs using Pearson correlation, and we found that the Pearson correlations were high (>0.95) from 0.1 to 0.7 but plummeted from 0.7 to 0.8 (supplemental Figure 9). We thus selected 0.7 as the PCC cutoff for GRN construction. Interestingly, the number 0.7 represents the inflection point on the curve that depicts the edge number versus different PCC cutoffs (supplemental Figure 10).

Genome-wide GRN clustering and GO enrichment analysis

To assess the potential implications of C_4 genes and C_4TFs in the different species, including their biological functions and the molecular relevance of their concurrent co-regulations, the genome-wide GRN was clustered into different modules. The modules enriched in C_4TFs or those that included the maximum number of C_4 genes were selected for functional enrichment analysis. We used the Markov Cluster Algorithm (MCL) package (version 14-137; Dongen, 2008) to identify the modules in the GRN. The MCL is an unsupervised model in which the number of clusters cannot be determined while in progress. The parameter "I" (inflation) represents the main handle affecting the cluster granularity, and it influences the cluster number and clustering efficiency that the algorithm produces (Enright et al., 2002). The clustering efficiency was calculated automatically in the program by considering two factors: the density of all clusters and the percentage of lost interactions. To determine a proper value for the I parameter used in the MCL package, we examined the clustering efficiencies and cluster numbers for five different cutoffs of I: 1.4, 2, 4, 5, and 6. We then chose the ideal value for I that had the maximum clustering efficiency for a minimum cluster number. We found that an I of 2 was best for all four species (supplemental Figure 11A), with clustering efficiencies of 0.21, 0.2, 0.16, and 0.22 for *Fro*, *Fso*, *Fra*, and *Ftr*, respectively (supplemental Figure 11B).

Subsequently, we identified modules enriched in C_4TFs (MET) using Fisher's test computed with an in-house R script. The calculated *P* values from the Fisher's test were adjusted using the BH method (Benjamini and Hochberg, 1995) integrated in R software (version 3.0.2). A threshold of 0.001 for the adjusted *P* value was used. The modules that included the maximum number of C_4 orthologs were termed MMC_4. To investigate the potential involvement of modules enriched in C_4TFs and modules with the maximum number of C_4 genes in specific biological functions, we calculated the over-represented GO terms for each of these modules using Fisher's test with a BH-adjusted $P < 0.001$, using genes in the genome-wide GRN as the background.

Determination of $\delta^{13}C$

For elemental and stable isotope analysis, mature leaves were dried in an oven at 150°C for 2 days. Fine powder of the dried leaves was then analyzed using an Isoprime 100 isotope ratio mass spectrometer coupled with an Elementar elemental analyzer (ISOTOPE cube; Elementar Analysensysteme, Germany) following the manufacturer's recommendations. The $\delta^{13}C$ measurements were calibrated following the two-point methods described by Coplen et al. (2006).

Electrophoretic mobility shift assay

The His-tagged recombinant proteins (DOF5.4-10×His, WRKY4-10×His, WRKY75-10×His, and OBP3-10×His) were developed in *E. coli* Rosetta 2(DE3). Bacterial cells were collected with lysis buffer (20 mM Tris–HCl, 300 mM NaCl, 0.5 mM DTT, 1× protease inhibitor cocktail I). The proteins were released from the collected cells by sonication and purified using a Ni column. DNA fragments were end-labeled with Cy5. The fluorescence-labeled DNA (20 μM) was incubated with purified protein in a 20-μl binding system (50 mM Tris–HCl [pH 8.0], 25 mM KCl, 2.5 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, 0.6 mM BSA, and 10% glycerol) for 30 min at 25°C. For competition assays, 40 μM unlabeled competitor DNA was also added to the reaction. For empty-TF, glutathione-S-transferase (GST)-FLAG-10×His without the TF proteins was developed in *E. coli* Rosetta 2(DE3) as described above. The reaction mixture was electrophoresed at 4°C on a 6% native polyacrylamide gel in 0.5× TBE for 40 min at 100 V in the dark. Fluorescently labeled DNA on the gel was then directly detected with a Typhoon imager (Typhoon, Cytiva).

$^{13}CO_2$ labeling and metabolite measurements

Fra plants were grown in the greenhouse under the conditions described above. One-month-old *Fra* plants were placed individually in custom-built labeling chambers. The labeling chambers were home-made according to Heise et al. (2014). Plants were adapted for 30 min in the chamber (temperature 25°C, humidity 50% ± 5%, CO_2 concentration 380 ppm, light intensity 500 μmol m^{-2} s^{-1}), and $^{13}CO_2$ was then introduced (>99%, Cambridge Isotope Laboratories, USA). After being fed with $^{13}CO_2$ for 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, or 600 s, the whole plant was frozen by pouring on liquid nitrogen. Approximately 30 mg of leaf tissue was taken from the frozen plants for metabolite extraction and analysis. Four biological replicates were performed for each measurement.

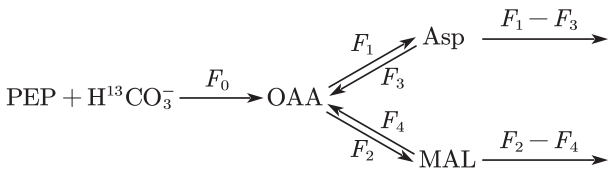
For metabolite extraction, finely ground leaf powder was fully dissolved in 800 μl −20°C pre-cooled extraction buffer (methanol:chloroform = 7:3 [v/v]) and incubated at −20°C for 3 h. After being mixed with 560 μl distilled water (ddH_2O), samples were centrifuged (2200 *g*, 10 min, 4°C), and 800 μl supernatant was collected. Samples were then mixed with another 800 μl −20°C pre-cooled extraction buffer (methanol:ddH_2O = 1:1 [v/v]) and centrifuged (2200 *g*, 10 min, 4°C). Next, 1.6 ml supernatant was collected and filtered through a 0.2-μm nylon filter. Then, 1 ml of the filtered sample was used for MS analysis and 20 μl for quality control. All of the above processes were performed on ice.

For the MS analysis, a 10-μl volume of sample was analyzed by HPLC-MS/MS (QTRAP 6500+ mass spectrometer, AB Sciex, USA). Metabolites were separated with a Luna NH_2 column (3 μm, 100 × 2 mm, Phenomenex, USA). The HPLC solvent gradients were set with eluent A (10 mM ammonium acetate and 5% [v/v] acetonitrile solution, pH adjusted to 9.5 with ammonium hydroxide) and eluent B (acetonitrile) as follows: 0–1 min, 15% A and 85% B; 1–8 min, 15%–70% A and 30%–85% B; 8–20 min, 70%–95% A and 5%–50% B; 20–22 min, 95% A and 5% B; 22–25 min, 15% A and 85% B. Next, MS/MS spectra were analyzed using the MRM model with the parameters described in (Wang et al., 2014; Arrivault et al., 2019) with some modifications, as detailed in supplemental Data 7. From this analysis, the percentages of metabolites labeled at each time point were obtained.

The concentrations of metabolites were calculated based on the "concentration-peak area" curve of analytical-standard samples and converted to chlorophyll content of *Fra* (nmol g Chl^{-1}). Six biological replicates were used for metabolite concentration measurements (supplemental Data 7). The chlorophyll measurements were performed according to (Arnon, 1949). Specifically, weighed leaf samples (second or third counting from the top) were incubated with 1.7 ml 80% acetone at 4°C for 12 h. Samples were centrifuged (1400 *g*, 2 min, 4°C), and 500 μl supernatant was added to a cuvette to measure the absorbance at 663 nm (A663) and 645 nm (A645) with a Cary 50 spectrophotometer (Varian, USA); the absorbance of 80% acetone was used as the blank control. The contents of chlorophyll a (Ca) and chlorophyll b (Cb) were calculated as follows: Ca = 12.21×A663–2.81×A645, Cb = 20.13×A646–5.03×A663. Two technical and five biological replicates were performed (supplemental Data 7).

Flux calculation

We assumed the following two metabolic pathways for OAA consumption:



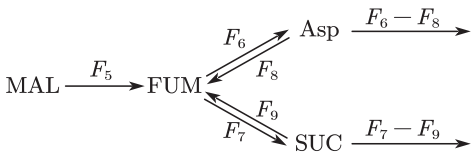
To simplify the calculation, we calculated the change in the abundance of unlabeled metabolite, which was used for subsequent metabolic flux fitting. During the $^{13}CO_2$ labeling, the unlabeled PEP, OAA, and Asp were satisfied according to the following differential equations:

$$\begin{cases} \frac{d[OAA_0]}{dt} = (F_1+F_2)\left(\frac{[PEP_0]}{[PEP]} - \frac{[OAA_0]}{[OAA]}\right) + F_3\left(\frac{[Asp_0]}{[Asp]} - \frac{[PEP_0]}{[PEP]}\right) + F_4\left(\frac{[MAL_0]}{[MAL]} - \frac{[PEP_0]}{[PEP]}\right) \\ \frac{d[Asp_0]}{dt} = F_1\frac{[OAA_0]}{[OAA]} - F\frac{[Asp_0]}{[Asp]_1} \\ \frac{d[MAL_0]}{dt} = F_2\frac{[OAA_0]}{[OAA]} - F_2\frac{[MAL_0]}{[MAL]} \end{cases}$$

With the measured metabolite concentrations of PEP, OAA, and Asp, we then fit the equations and obtained the fluxes as follows:

Reaction	Flux (nmol mg Chl^{-1} s^{-1})
AST forward (F_1)	41.07
MDH forward (F_2)	74.14
AST reverse (F_3)	0
MDH reverse (F_4)	0
Asp generation ($F_1 - F_3$)	41.07
MAL generation ($F_2 - F_4$)	72.14
PEPC ($F_0 = F_1 + F_2 - F_3 - F_4$)	113.21

Assuming that FUM generates Asp and SUC:



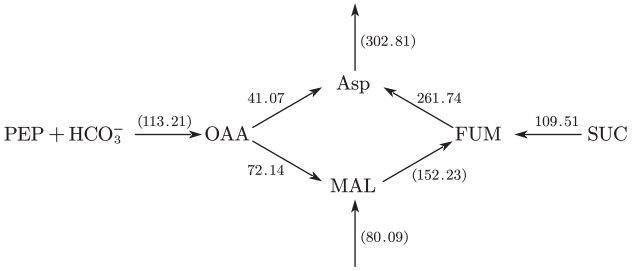
During the $^{13}CO_2$ labeling, the unlabeled MAL, FUM, Asp, and SUC were satisfied according to the following differential equations:

$$\begin{cases} \frac{d[FUM_0]}{dt} = (F_6+F_7)\left(\frac{[MAL_0]}{[MAL]} - \frac{[FUM_0]}{[FUM]}\right) + F_8\left(\frac{[Asp_0]}{[Asp]} - \frac{[MAL_0]}{[MAL]}\right) + F_9\left(\frac{[SUC_0]}{[SUC]} - \frac{[MAL_0]}{[MAL]}\right) \\ \frac{d[Asp_0]}{dt} = F_6\frac{[FUM_0]}{[FUM]} - F_6\frac{[Asp_0]}{[Asp]} \\ \frac{d[SUC_0]}{dt} = F_7\frac{[FUM_0]}{[FUM]} - F_7\frac{[Asp_0]}{[Asp]} \end{cases}$$

With the measured metabolite concentrations of MAL, Asp, FUM, and SUC, we then fit the equations and obtained the fluxes as follows:

Reaction	Flux (nmol mg Chl^{-1} s^{-1})
AspA reverse (F_6)	261.74
SDH forward (F_7)	7505.50
AspA forward (F_8)	0
SDH reverse (F_9)	7615.01
Asp generation ($F_6 - F_8$)	261.74
SUC generation ($F_9 - F_7$)	105.91
FH ($F_5 = F_6 + F_7 - F_8 - F_9$)	152.23

Taken together, the fluxes of the metabolic pathway related to PEPC-fixed CO_2 were as follows (also see Figure 6B).



DATA AVAILABILITY

The RNA-seq and Iso-seq data generated in this study were submitted to the China National Genomics Data Center (NGDC) with accession number PRJCA011102. These data are also available in the Gene Expression Omnibus (GEO) with accession number GSE143470.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This work was supported by the Strategic Priority Research Program of the General Project of the Chinese Academy of Sciences (grant no.

XDB27020105) and by the National Research and Development Program of the Ministry of Science and Technology of the People's Republic of China, MSTC (2019YFA0904600).

AUTHOR CONTRIBUTIONS

X.-G.Z., G.C., and M.-J.L. designed the project and wrote the paper. M.-J.L. performed RNA-seq analysis, GRN comparisons, and qRT–PCR. Q.T. performed the metabolic experiment and analysis. Y.W. performed EMSA. J.E. performed the RNA isolation and wrote the paper. F.C. performed GRN construction. X.N. performed metabolite isolation.

ACKNOWLEDGMENTS

The authors thank and appreciate Profs. Rowan F. Sage and Peter Westhoff for sharing *Flaveria* materials with us. We thank Prof. Haiyang Hu and Mingnan Qu for their fruitful discussions and suggestions. The authors also thank Fenfen Miao, Qingfeng Song, and Xinyu Liu for their assistance in maintaining the growth chamber conditions properly and according to the recorded instructions. No conflict of interest is declared.

Received: March 15, 2022
Revised: June 16, 2022
Accepted: August 11, 2022
Published: August 19, 2022

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. https://doi.org/10.1105/tpc.107.053322.

Arnon, D.I. (1949). Copper enzymes in isolated chloroplasts. Polyphenoloxidase in Beta vulgaris. Plant Physiol. **24**:1–15.

Arrivault, S., Alexandre Moraes, T., Obata, T., Medeiros, D.B., Fernie, A.R., Boulouis, A., Ludwig, M., Lunn, J.E., Borghi, G.L., Schlereth, A., et al. (2019). Metabolite profiles reveal interspecific variation in operation of the Calvin-Benson cycle in both C4 and C3 plants. J. Exp. Bot. **70**:1843–1858. https://doi.org/10.1093/jxb/erz051.

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. https://doi.org/10.1371/journal.pgen.1004365.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B **57**:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Borghi, G.L., Arrivault, S., Günther, M., Barbosa Medeiros, D., Dell'Aversana, E., Fusco, G.M., Carillo, P., Ludwig, M., Fernie, A.R., Lunn, J.E., et al. (2022). Metabolic profiles in C3, C3-C4 intermediate, C4-like, and C4 species in the genus Flaveria. J. Exp. Bot. **73**:1581–1601. https://doi.org/10.1093/jxb/erab540.

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

Burgess, S.J., Granero-Moya, I., Grangé-Guermente, M.J., Boursnell, C., Terry, M.J., and Hibberd, J.M. (2016). Ancestral light and chloroplast regulation form the foundations for C_4 gene expression. Native Plants **2**:16161. https://doi.org/10.1038/nplants.2016.161.

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 C_3 and C_4 grasses. Plant Cell **31**:2297–2314. https://doi.org/10.1105/tpc.19.00078.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST plus : architecture and applications. BMC Bioinf. **10**. https://doi.org/10.1186/1471-2105-10-421.

Christin, P.A., Petitpierre, B., Salamin, N., Büchi, L., and Besnard, G. (2009). Evolution of C_4 phosphoenolpyruvate carboxykinase in Grasses, from genotype to phenotype. Mol. Biol. Evol. **26**:357–365. https://doi.org/10.1093/molbev/msn255.

Christin, P.A., Osborne, C.P., Sage, R.F., Arakaki, M., and Edwards, E.J. (2011). C4 eudicots are not younger than C4 monocots. J. Exp. Bot. **62**:3171–3181. https://doi.org/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 C4 photosynthesis. Genome Biol. Evol. **5**:2174–2187. https://doi.org/10.1093/gbe/evt168.

Coplen, T.B., Brand, W.A., Gehre, M., Gröning, M., Meijer, H.A.J., Toman, B., and Verkouteren, R.M. (2006). New guidelines for delta^{13}C measurements. Anal. Chem. **78**:2439–2441. https://doi.org/10.1021/ac052027c.

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.

Van Dongen, S. (2008). Graph clustering via a discrete uncoupling process. SIAM J. Matrix Anal. Appl. **30**:121–141. https://doi.org/10.1137/040608635.

Duarte, K.E., de Souza, W.R., Santiago, T.R., Sampaio, B.L., Ribeiro, A.P., Cotta, M.G., da Cunha, B.A.D.B., Marraccini, P.R.R., Kobayashi, A.K., and Molinari, H.B.C. (2019). Identification and characterization of core abscisic acid (ABA) signaling components and their gene expression profile in response to abiotic stresses in Setaria viridis. Sci. Rep. **9**:4028. https://doi.org/10.1038/s41598-019-40623-5.

Edwards, G.E., and Ku, M.S.B. (1987). Biochemistry of C_3-C_4 intermediates. In The Biochemistry of Plants, M.D. Hatch and N.K. Boardman, eds. (Academic Press), pp. 275–325.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. **30**:1575–1584. https://doi.org/10.1093/nar/30.7.1575.

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. https://doi.org/10.1146/annurev.pp.40.060189.002443.

Fischer, E., Raschke, K., and Stitt, M. (1986). Effects of abscisic-acid on photosynthesis in whole leaves - changes in CO_2 assimilation, levels of carbon-reduction-cycle intermediates, and activity of ribulose-1, 5-bisphosphate carboxylase. Planta **169**:536–545. https://doi.org/10.1007/Bf00392104.

Fisher, S., and Franz-Odendaal, T. (2012). Evolution of the bone gene regulatory network. Curr. Opin. Genet. Dev. **22**:390–397. https://doi.org/10.1016/j.gde.2012.04.007.

Fu, S., Ma, Y., Yao, H., Xu, Z., Chen, S., Song, J., and Au, K.F. (2018). IDP-denovo: de novo transcriptome assembly and isoform annotation by hybrid sequencing. Bioinformatics **34**:2168–2176. https://doi.org/10.1093/bioinformatics/bty098.

Fukayama, H., Tsuchida, H., Agarie, S., Nomura, M., Onodera, H., Ono, K., Lee, B.H., Hirose, S., Toki, S., Ku, M.S., et al. (2001). Significant accumulation of C(4)-specific pyruvate, orthophosphate dikinase in a C(3) plant, rice. Plant Physiol. **127**:1136–1146.

Górska, A.M., Gouveia, P., Borba, A.R., Zimmermann, A., Serra, T.S., Lourenço, T.F., Margarida Oliveira, M., Peterhänsel, C., and Saibo, N.J.M. (2019). ZmbHLH80 and ZmbHLH90 transcription factors act antagonistically and contribute to regulate PEPC1 cell-specific gene

expression in maize. Plant J. **99**:270–285. https://doi.org/10.1111/tpj.14323.

Gowik, U., Bräutigam, A., Weber, K.L., Weber, A.P.M., and Westhoff, P. (2011). Evolution of C4 photosynthesis in the genus Flaveria: how many and which genes does it take to make C4? Plant Cell **23**:2087–2105. https://doi.org/10.1105/tpc.111.086264.

Gowik, U., Burscheidt, J., Akyildiz, M., Schlue, U., Koczor, M., Streubel, M., and Westhoff, P. (2004). cis-Regulatory elements for mesophyll-specific gene expression in the C_4 plant Flaveria trinervia, the promoter of the C_4 phosphoenolpyruvate carboxylase gene. Plant Cell **16**:1077–1090. https://doi.org/10.1105/tpc.019729.

Gowik, U., Schulze, S., Saladié, M., Rolland, V., Tanz, S.K., Westhoff, P., and Ludwig, M. (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. https://doi.org/10.1093/jxb/erw475.

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

Gupta, S.D., Levey, M., Schulze, S., Karki, S., Emmerling, J., Streubel, M., Gowik, U., Paul Quick, W., and Westhoff, P. (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. https://doi.org/10.1111/tpj.14532.

Hackl, T., Hedrich, R., Schultz, J., and Förster, F. (2014). proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics **30**:3004–3011. https://doi.org/10.1093/bioinformatics/btu392.

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

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

Heise, R., Arrivault, S., Szecowka, M., Tohge, T., Nunes-Nesi, A., Stitt, M., Nikoloski, Z., and Fernie, A.R. (2014). Flux profiling of photosynthetic carbon metabolism in intact plants. Nat. Protoc. **9**:1803–1824. https://doi.org/10.1038/nprot.2014.115.

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

Hibberd, J.M., Sheehy, J.E., and Langdale, J.A. (2008). Using C_4 photosynthesis to increase the yield of rice-rationale and feasibility. Curr. Opin. Plant Biol. **11**:228–231. https://doi.org/10.1016/j.pbi.2007.11.002.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. **3**:318–356. https://doi.org/10.1016/s0022-2836(61)80072-7.

Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J., and Gao, G. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. **45**:D1040–D1045. https://doi.org/10.1093/nar/gkw982.

Kajala, K., Brown, N.J., Williams, B.P., Borrill, P., Taylor, L.E., and Hibberd, J.M. (2012). Multiple Arabidopsis genes primed for recruitment into C_4 photosynthesis. Plant J. **69**:47–56. https://doi.org/10.1111/j.1365-313X.2011.04769.x.

King, M.C., and Wilson, A.C. (1975). Evolution at two levels in humans and chimpanzees. Science **188**:107–116. https://doi.org/10.1126/science.1090005.

Ku, M.S., Agarie, S., Nomura, M., Fukayama, H., Tsuchida, H., Ono, K., Hirose, S., Toki, S., Miyao, M., and Matsuoka, M. (1999). High-level expression of maize phosphoenolpyruvate carboxylase in transgenic rice plants. Nat. Biotechnol. **17**:76–80. https://doi.org/10.1038/5256.

Ku, M.S.B., Nakamoto, H., Monson, R.K., Littlejohn, R.O., Fisher, D.B., and Edwards, G.E. (1983). Photosynthetic characteristics of Flaveria species intermediate between C-3 and C-4 plants. Plant Physiol. **72**:43.

Ku, M.S., Wu, J., Dai, Z., Scott, R.A., Chu, C., and Edwards, G.E. (1991). Photosynthetic and photorespiratory characteristics of Flaveria species. Plant Physiol. **96**:518–528. https://doi.org/10.1104/pp.96.2.518.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods **9**:357–359. https://doi.org/10.1038/Nmeth.1923.

Lauterbach, M., Schmidt, H., Billakurthi, K., Hankeln, T., Westhoff, P., Gowik, U., and Kadereit, G. (2017). De novo transcriptome assembly and comparison of C_3, C_3-C_4, and C_4 species of tribe salsoleae (Chenopodiaceae). Front. Plant Sci. **8**:1939. https://doi.org/10.3389/fpls.2017.01939.

Levine, M., and Davidson, E.H. (2005). Gene regulatory networks for development. Proc. Natl. Acad. Sci. USA **102**:4936–4942. https://doi.org/10.1073/pnas.0408031102.

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. https://doi.org/10.1186/1471-2105-12-323.

Li, Y., Xu, J., Haq, N.U., Zhang, H., and Zhu, X.G. (2014). Was low CO2 a driving force of C4 evolution: Arabidopsis responses to long-term low CO2 stress. J. Exp. Bot. **65**:3657–3667. https://doi.org/10.1093/jxb/eru193.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. Methods **25**:402–408. https://doi.org/10.1006/meth.2001.1262.

Lundgren, M.R., Christin, P.A., Escobar, E.G., Ripley, B.S., Besnard, G., Long, C.M., Hattersley, P.W., Ellis, R.P., Leegood, R.C., and Osborne, C.P. (2016). Evolutionary implications of C3 -C4 intermediates in the grass Alloteropsis semialata. Plant Cell Environ. **39**:1874–1885. https://doi.org/10.1111/pce.12665.

Lyu, M.J.A., Wang, Y., Jiang, J., Liu, X., Chen, G., and Zhu, X.G. (2020). What matters for C4 transporters: evolutionary changes of phosphoenolpyruvate transporter for C4 photosynthesis. Front. Plant Sci. **11**:935. https://doi.org/10.3389/fpls.2020.00935.

Lyu, M.J.A., Gowik, U., Kelly, S., Covshoff, S., Hibberd, J.M., Sage, R.F., Ludwig, M., Wong, G.K.S., Westhoff, P., and Zhu, X.G. (2021). The coordination of major events in C4 photosynthesis evolution in the genus Flaveria. Sci. Rep. **11**:15618. https://doi.org/10.1038/s41598-021-93381-8.

Lyu, M.J.A., Gowik, U., Kelly, S., Covshoff, S., Mallmann, J., Westhoff, P., Hibberd, J.M., Stata, M., Sage, R.F., Lu, H., et al. (2015). RNA-Seq based phylogeny recapitulates previous phylogeny of the genus Flaveria (Asteraceae) with some modifications. BMC Evol. Biol. **15**:116. https://doi.org/10.1186/s12862-015-0399-9.

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

McKown, A.D., and Dengler, N.G. (2007). Key innovations in the evolution of Kranz anatomy and C4 vein pattern in Flaveria

(Asteraceae). Am. J. Bot. **94**:382–399. https://doi.org/10.3732/ajb.94.3.382.

McKown, A.D., Moncalvo, J.-M., and Dengler, N.G. (2005). Phylogeny of Flaveria (Asteraceae) and inference of C4 photosynthesis evolution. Am. J. Bot. **92**:1911–1928. https://doi.org/10.3732/ajb.92.11.1911.

McQueen, E., and Rebeiz, M. (2020). On the specificity of gene regulatory networks: how does network co-option affect subsequent evolution? Curr. Top. Dev. Biol. **139**:375–405. https://doi.org/10.1016/bs.ctdb.2020.03.002.

Min, X.J., Butler, G., Storms, R., and Tsang, A. (2005). OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res. **33**:W677–W680. https://doi.org/10.1093/nar/gki394.

Mitra, K., Carvunis, A.R., Ramesh, S.K., and Ideker, T. (2013). Integrative approaches for finding modular structure in biological networks. Nat. Rev. Genet. **14**:719–732. https://doi.org/10.1038/nrg3552.

Monson, R.K., Moore, B.D., Ku, M.S., and Edwards, G.E. (1986). Co-function of C_3-and C_4-photosynthetic pathways in C_3, C_4 and C_3-C_4 intermediate Flaveria species. Planta **168**:493–502. https://doi.org/10.1007/BF00392268.

Monteiro, A. (2012). Gene regulatory networks reused to build novel traits: co-option of an eye-related gene regulatory network in eye-like organs and red wing patches on insect wings is suggested by optix expression. Bioessays **34**:181–186. https://doi.org/10.1002/bies.201100160.

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. https://doi.org/10.1093/molbev/msx269.

Nakamoto, H., Ku, M.S.B., and Edwards, G.E. (1983). Photosynthetic characteristics of C3-C4 intermediate Flaveria species .2. Kinetic-properties of phosphoenolpyruvate carboxylase from C-3, C-4 and C3-C4 intermediate species. Plant Cell Physiol. **24**:1387–1393. https://doi.org/10.1093/oxfordjournals.pcp.a076659.

Nakamura, N., Iwano, M., Havaux, M., Yokota, A., and Munekage, Y.N. (2013). Promotion of cyclic electron transport around photosystem I during the evolution of NADP-malic enzyme-type C4 photosynthesis in the genus Flaveria. New Phytol. **199**:832–842. https://doi.org/10.1111/nph.12296.

Necsulea, A., and Kaessmann, H. (2014). Evolutionary dynamics of coding and non-coding transcriptomes. Nat. Rev. Genet. **15**:734–748. https://doi.org/10.1038/nrg3802.

Nomura, M., Sentoku, N., Nishimura, A., Lin, J.H., Honda, C., Taniguchi, M., Ishida, Y., Ohta, S., Komari, T., Miyao-Tokutomi, 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. https://doi.org/10.1046/j.1365-313x.2000.00726.x.

Peisker, M. (1986). Models of carbon metabolism in C-3-C-4 intermediate plants as applied to the evolution of C-4 photosynthesis. Plant Cell Environ. **9**:627–635. https://doi.org/10.1111/j.1365-3040.1986.tb01620.x.

Peter, I.S., and Davidson, E.H. (2011). Evolution of gene regulatory networks controlling body plan development. Cell **144**:970–985. https://doi.org/10.1016/j.cell.2011.02.017.

Powell, A.M. (1978). Systematics of Flaveria (flaveriinae Asteraceae). Ann. Mo. Bot. Gard. **65**:590–636. https://doi.org/10.2307/2398862.

Reyna-Llorens, I., and Hibberd, J.M. (2017). Recruitment of pre-existing networks during the evolution of C4 photosynthesis. Philos. Trans. R. Soc. Lond. B Biol. Sci. **372**. https://doi.org/10.1098/rstb.2016.0386.

Reyna-Llorens, I., Burgess, S.J., Reeves, G., Singh, P., Stevenson, S.R., Williams, B.P., Stanley, S., and Hibberd, J.M. (2018). Ancient duons may underpin spatial patterning of gene expression in C_4 leaves. Proc. Natl. Acad. Sci. USA **115**:1931–1936. https://doi.org/10.1073/pnas.1720576115.

Riano-Pachon, D.M., Ruzicic, S., Dreyer, I., and Mueller-Roeber, B. (2007). PlnTFDB: an integrative plant transcription factor database. BMC bioinformatics **8**:42. https://doi.org/10.1186/1471-2105-8-42.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**:139–140. https://doi.org/10.1093/bioinformatics/btp616.

Romero, I.G., Ruvinsky, I., and Gilad, Y. (2012). Comparative studies of gene expression and the evolution of gene regulation. Nat. Rev. Genet. **13**:505–516. https://doi.org/10.1038/nrg3229.

Rumpho, M.E., Ku, M.S., Cheng, S.H., and Edwards, G.E. (1984). Photosynthetic characteristics of C3-C4 intermediate Flaveria species .3. Reduction of photorespiration by a limited-C4 pathway of photosynthesis in Flaveria-ramosissima. Plant Physiol. **75**:993–996. https://doi.org/10.1104/pp.75.4.993.

Sage, R.F. (2001). Environmental and evolutionary preconditions for the origin and diversification of the C4 photosynthetic syndrome. Plant Biol. **3**:202–213. https://doi.org/10.1055/s-2001-15206.

Sage, R.F. (2004). The evolution of C4 photosynthesis. New Phytologist, 341–370.

Sage, R.F. (2017). A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. J. Exp. Bot. **68**:4039–4056. https://doi.org/10.1093/jxb/erx005.

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

Sage, R.F., Sage, T.L., and Kocacinar, F. (2012). Photorespiration and evolution of C_4 photosynthesis. Annu. Rev. Plant Biol. **63**:19–47.

Schlüter, U., and Weber, A.P.M. (2020). Regulation and evolution of C_4 photosynthesis. Annu. Rev. Plant Biol. **71**:183–215. https://doi.org/10.1146/annurev-arplant-042916-040915.

Schlüter, U., Bräutigam, A., Gowik, U., Melzer, M., Christin, P.A., Kurz, S., Mettler-Altmann, T., and Weber, A.P. (2017). Photosynthesis in C3-C4 intermediate Moricandia species. J. Exp. Bot. **68**:191–206. https://doi.org/10.1093/jxb/erw391.

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. https://doi.org/10.1042/bj1030660.

Taniguchi, Y.Y., Gowik, U., Kinoshita, Y., Kishizaki, R., Ono, N., Yokota, A., Westhoff, P., and Munekage, Y.N. (2021). Dynamic changes of genome sizes and gradual gain of cell-specific distribution of C4 enzymes during C4 evolution in genus Flaveria. Plant Genome **14**:e20095. https://doi.org/10.1002/tpg2.20095.

Thompson, D., Regev, A., and Roy, S. (2015). Comparative analysis of gene regulatory networks: from network reconstruction to evolution. Annu. Rev. Cell Dev. Biol. **31**:399–428. https://doi.org/10.1146/annurev-cellbio-100913-012908.

Tomoyasu, Y., Arakane, Y., Kramer, K.J., and Denell, R.E. (2009). Repeated co-options of exoskeleton formation during wing-to-elytron evolution in beetles. Curr. Biol. **19**:2057–2065. https://doi.org/10.1016/j.cub.2009.11.014.

Tu, X., Mejía-Guerra, M.K., Valdes Franco, J.A., Tzeng, D., Chu, P.Y., Shen, W., Wei, Y., Dai, X., Li, P., Buckler, E.S., et al. (2020). Reconstructing the maize leaf regulatory network using ChIP-seq

data of 104 transcription factors. Nat. Commun. **11**:5089. https://doi.org/10.1038/s41467-020-18832-8.

Ubierna, N., Sun, W., Kramer, D.M., and Cousins, A.B. (2013). The efficiency of C4 photosynthesis under low light conditions in Zea mays, Miscanthus x giganteus and Flaveria bidentis. Plant Cell Environ. **36**:365–381. https://doi.org/10.1111/j.1365-3040.2012.02579.x.

Ueno, O. (2001). Environmental regulation of C(3) and C(4) differentiation in the amphibious sedge Eleocharis vivipara. Plant Physiol. **127**:1524–1532.

Vogan, P.J., and Sage, R.F. (2011). Water-use efficiency and nitrogen-use efficiency of C_3 -C_4 intermediate species of Flaveria Juss. (Asteraceae). Plant Cell Environ. **34**:1415–1430. https://doi.org/10.1111/j.1365-3040.2011.02340.x.

Wang, L., Czedik-Eysenberg, A., Mertz, R.A., Si, Y., Tohge, T., Nunes-Nesi, A., Arrivault, S., Dedow, L.K., Bryant, D.W., Zhou, W., et al. (2014). Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice. Nat. Biotechnol. **32**:1158–1165. https://doi.org/10.1038/nbt.3019.

Williams, B.P., Aubry, S., and Hibberd, J.M. (2012). Molecular evolution of genes recruited into C4 photosynthesis. Trends Plant Sci. **17**:213–220. https://doi.org/10.1016/j.tplants.2012.01.008.

Williams, B.P., Burgess, S.J., Reyna-Llorens, I., Knerova, J., Aubry, S., Stanley, S., and Hibberd, J.M. (2016). An untranslated cis-element regulates the accumulation of multiple C4 enzymes in Gynandropsis gynandra mesophyll cells. Plant Cell **28**:454–465. https://doi.org/10.1105/tpc.15.00570.

Wohlbach, D.J., Thompson, D.A., Gasch, A.P., and Regev, A. (2009). From elements to modules: regulatory evolution in Ascomycota fungi. Curr. Opin. Genet. Dev. **19**:571–578. https://doi.org/10.1016/j.gde.2009.09.007.

Zhang, X., Zhao, X.M., He, K., et al. (2012). Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information **28**:98–104. https://doi.org/10.1093/bioinformatics/btr626.

Zhao, H.L., Chang, T.G., Xiao, Y., and Zhu, X.G. (2021). Potential metabolic mechanisms for inhibited chloroplast nitrogen assimilation under high CO2. Plant Physiol. **187**:1812–1833. https://doi.org/10.1093/plphys/kiab345.

Zheng, G., Xu, Y., Zhang, X., Liu, Z.P., Wang, Z., Chen, L., and Zhu, X.G. (2016). CMIP: a software package capable of reconstructing genome-wide regulatory networks using gene expression data. BMC Bioinf. **17**:535. https://doi.org/10.1186/s12859-016-1324-y.

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. https://doi.org/10.1016/j.copbio.2008.02.004.