Title: **Distinct C_4 Sub-Types and C_3 Bundle Sheath Isolation In The Paniceae Grasses**

Short Title: C4 Sub-Types In The Paniceae Grasses

Authors: Jacob D. Washburn[1,2,*], Josh Strable[3], Patrick Dickinson[4], Satya S. Kothapalli[2], Julia M. Brose[2], Sarah Covshoff[4], Gavin C. Conant[5], Julian M. Hibberd[4], J. Chris Pires[2]

1) Plant Genetics Research Unit, USDA-ARS, 302-A Curtis Hall, University of Missouri, Columbia, MO 65211 USA
2) Division of Biological Sciences, University of Missouri, 311 Bond Life Science Center, Columbia, MO 65211 USA
3) Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC, USA 27695
4) Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA UK
5) Program in Genetics, Bioinformatics Research Center, Department of Biological Sciences, 356 Ricks Hall, North Carolina State University, Raleigh, NC, USA 27695

*Corresponding Author: Jacob D. Washburn (Jacob.Washburn@USDA.gov)

Abstract

In C_4 plants, the enzymatic machinery underpinning photosynthesis can vary, with, for example, three distinct C_4 acid decarboxylases being used to release CO_2 in the vicinity of RuBisCO. For decades, these decarboxylases have been used to classify C_4 species into three biochemical sub-types. However, more recently the notion that C_4 species mix and match C_4 acid decarboxylases has increased in popularity and, as a consequence, the validity of specific biochemical sub-types has been questioned. Using five species from the grass tribe Paniceae, we show that, while in some species transcripts encoding multiple C_4 acid decarboxylases accumulate, in others, transcript abundance and enzyme activity is almost entirely from one decarboxylase. In addition, the development of a bundle sheath isolation procedure for a close C_3 species in the Paniceae enables the preliminary exploration of C_4 sub-type evolution.

Keywords

C_4, Photosynthesis, C_4 Sub-types, Evolution

Introduction

C_4 photosynthesis is often considered the most productive mechanism by which plants convert sunlight into chemical energy (Sage, 2004; Wang et al., 2012; Niklaus and Kelly, 2019; Kopriva and Weber, 2021). The C_4 pathway leads to increased photosynthetic efficiency because high concentrations of CO_2 are supplied to RuBisCO. Since its discovery in the 1960s (Hatch and Slack, 1966), a unified understanding of the biochemistry underpinning C_4 photosynthesis has emerged. This basic system comprises a biochemical pump that initially fixes HCO_3^- into C_4 acids in mesophyll (M) cells. Subsequently, diffusion of these C_4 acids into a separate compartment, followed by their decarboxylation, generates high concentrations of CO_2 around RuBisCO. In many plants, the release of CO_2 occurs in bundle sheath (BS) cells (Hatch, 1992; Furbank, 2016; von Caemmerer et al., 2017). Although this pump demands additional ATP inputs, in warm environments where RuBisCO catalyzes high rates of oxygenation (and therefore photorespiration), the C_4 pathway increases photosynthetic efficiency compared with the ancestral C_3 state.

Elucidation of the C_4 pathway was initially based on analysis of sugarcane (*Saccharum spp.* L.) and maize (corn, *Zea mays* L.), which both use the chloroplastic NADP-DEPENDENT MALIC ENZYME (NADP-ME) to release CO_2 in BS cells. However, it became apparent that not all species used this chloroplastic enzyme. For example, *Megathyrsus maximus* (formerly *Panicum maximum*), *Urochloa texanum* (formerly *Panicum texanum*), and *Sporobolus poiretti* used the cytosolic enzyme PHOSPHONENOLPYRUVATE CARBOXYKINASE (PEPCK) (Edwards et al., 1971) to release CO_2 in the BS, whereas *Atriplex spongiosa* and *Panicum miliaceum* showed high activities of the mitochondrial NAD-DEPENDENT MALIC ENZYME (NAD-ME) (Hatch and Kagawa, 1974).

These findings led to the consensus that different C_4 species made preferential use of one C_4 acid decarboxylase and resulted in the classification of C_4 plants into one of three distinct biochemical pathways (Edwards et al., 1971; Hatch et al., 1975; Hatch and Kagawa, 1976). According to Furbank (2016), there was some early discussion about whether the sub-types were mutually exclusive or if one species might employ two or more sub-types together, but in general, the sub-types were described as distinct (Hatch, 1987).

For several decades this description of three sub-types has been standard practice (Sheen, 1999; Hibberd and Covshoff, 2010) and even used in taxonomic classification (Brown, 1977). However, more recent work has provided evidence that some C_4 species use multiple C_4 acid decarboxylases. Maize, for example, was traditionally classified as using NADP-ME but evidence has mounted that it and sugar-cane both have high activities of PEPCK (Walker et al., 1997; Wingler et al., 1999; Majeran et al., 2010; Furbank, 2011; Pick et al., 2011; Bellasio and Griffiths, 2013; Sharwood et al., 2014; Wang et al., 2014; Koteyeva et al., 2015; Weissmann et al., 2016; Cacefo et al., 2019). This blurring of the NADP-ME C_4 sub-type coincided with observations that many plants with high amounts of PEPCK also contained either NADP-ME of NAD-ME (Furbank, 2011). Furthermore, computational models of the C_4 pathways suggested that BS energy requirements could not be met in a system with only PEPCK decarboxylation (Wang et al., 2014). It has therefore been suggested that PEPCK may never function on its own as a distinct sub-type (Furbank, 2011; Bräutigam et al., 2014; Wang et al., 2014).

Alternatives to the three sub-type classification have since been proposed and used in a number of recent publications. These include a two sub-type system (based on the use of NADP-ME or NAD-ME), as well as a four sub-type classification placing species into NADP-ME, NAD-ME,

NADP-ME + PEPCK, and NAD-ME + PEPCK sub-types (Wang et al., 2014; Washburn et al., 2015; Rao and Dixon, 2016). At present, none of these classification schemes has been widely adopted by the community. Moreover, convincing experimental evidence (i.e., transcriptomic, or proteomic data) that species traditionally defined as belonging to the PEPCK sub-type actually use another C_4 acid decarboxylation enzyme at a higher level than PEPCK is lacking, while enzyme activity measurements in the older literature indicate strong PEPCK predominance for several species (Gutierrez et al., 1974; Prendergast et al., 1987; Lin et al., 1993).

Only one group of species, the tribe Paniceae (Poaceae) has been documented to contain all three classical biochemical sub-types of C_4 photosynthesis together in a pattern consistent with a single C_4 origin (Sage et al., 2011). The subtribe Cenchrinae consists of species using the classical NADP-ME C_4 sub-type, the subtribe Melinidinae the PEPCK sub-type, and the Panicinae the NAD-ME sub-type (Gutierrez et al., 1974; Prendergast et al., 1987; Lin et al., 1993). The sub-tribes Cenchrinae, Melinidinae, and Panicineae (CMP) form a well-supported phylogenetic clade of C_4 species with many C_3 species sister to the clade (Vicentini et al., 2008; Grass Phylogeny Working Group II, 2012; Washburn et al., 2015). Studies based solely or predominantly on nuclear genes have confirmed this CMP clade, but also placed the sub-tribe Anthephorineae as sister to the CMPA clade. These phylogenies would create a CMPA clade of C_4 species potentially sharing a single C_3 ancestor (Vicentini et al., 2008; Washburn et al., 2017). This clade is here referred to as the CMP(A) clade in order to indicate the incongruence found between nuclear and chloroplast phylogenies (Figure 1). The analyses here performed would be equally valid regardless on the inclusion of Anthephorineae.

How and why C_4 photosynthesis and its sub-types evolved has been investigated for many years (Raghavendra, 1980; Rawsthorne, 1992; Sage, 2001; Sage, 2004; Langdale, 2011; Sage et al., 2012; Schluter and Weber, 2020). Current hypotheses suggest an intermediate C_3-C_4 stage in which a photorespiratory pump operated (Sage, 2004; Sage et al., 2012; Heckmann et al., 2013; Mallmann et al., 2014; Bräutigam and Gowik, 2016; Blätke and Bräutigam, 2019). Each C_4 sub-type would require at least some distinct evolutionary innovations, and the question of how or why multiple sub-types would evolve from the same C_3 ancestor remains unanswered, although some evidence suggests it could be related to light quality and/or nutrient availability (Pinto et al., 2016; Sonawane et al., 2018; Blätke and Bräutigam, 2019; Arp et al., 2021).

To better determine whether the PEPCK pathway represents a true biochemical sub-type and investigate the extent to which C_4 species make use of mixtures of C_4 acid decarboxylases, global patterns of mRNA abundance were assessed from BS and M enriched samples across phylogenetically spaced C_4 plants that were traditionally defined as exclusively using one of each of the C_4 sub-types. These species belong to each subtribe of the CMP(A) described above. The C_3 species *Sacciolepis indica*, another member of the Paniceae and sister to the CMP(A) clade (sister to CMP in chloroplast phylogeny and sister to CMPA in nuclear phylogeny), was included in the analysis to provide insight into the ancestral state and evolutionary transition from C_3 to different C_4 sub-types (Washburn et al., 2015; Washburn et al., 2017). A simple method was developed for isolating bundle sheath cells from the C_3 species *Sacciolepis indica*.

We find that at least one species in the tribe appears to use PEPCK decarboxylation exclusively or nearly so, while the other species examined appear to be of mix subtype. Analysis of the C_3 species *S. indica* shows low levels of C_4 transcripts and an amenability to mechanical bundle

sheath separation procedures not previously seen in C_3 species. These observations lead us to hypothesize that S. indica may lie somewhere on the spectrum of C_3-C_4 intermediates or represent a reversion from an ancestral C_3-C_4 intermediate.

Results

M And BS Extraction And Distribution Of Transcripts Encoding The Core C_4 Cycle

Four C_4 species from the Paniceae tribe were chosen to represent the CMP(A) subtribes in the Paniceae (Figure 1). *Setaria italica* for Cenchrinae (NADP-ME), *Urochola fusca* for Melinidinae (PEPCK), *Panicum hallii for* Panicinae (NAD-ME), and *Digitaria californica* for Anthephorineae (NADP-ME). *Sacciolepis indica* was chosen to represent the closest C_3 relative to the group.

Microscopic examination of leaves of *S. italica, U. fusca, P. hallii* and *D. californica*, from which M cell contents had been extracted, showed bands of cells containing low chlorophyll content (Figure 2A-D) a phenotype consistent with efficient removal of M content (Covshoff et al., 2013; John et al., 2014). In addition, after mechanical isolation of leaves, BS preparations of high purity for all C_4 species were generated (Figure 2A-D). Separation of BS strands was also successful for the C_3 species *Sacciolepis indica,* something that to our knowledge has not been successful in any other C_3 species (Figure 2E). Analysis of transcripts derived from core C_4 genes showed clear differences in abundance between M and BS samples from the C_4 species. For example, transcripts derived from *CARBONIC ANHYDRASE* (*CA*), *PHOSPHOENOLPYRUVATE CARBOXYLASE* (*PEPC*) and *PYRUVATE, ORTHOPHOSPHATE DIKINASE* (*PPDK*) genes preferentially accumulated in M cells (Figure 3A). In contrast, transcripts derived from the *RUBISCO SMALL SUBUNIT* (*RBCS*) and *RUBISCO ACTIVASE* (*RCA*) as well as either *NADP-ME*, *NAD-ME* or *PEPCK* were more abundant in

BS strands (Figure 3B). The abundance of transcripts relating to C_4 photosynthesis in the C_3 species *S. indica* were also consistent with current knowledge of metabolism in the BS of C_3 species. For example, RBCS and RCA were more abundant in whole leaf samples than in the BS.

Some Paniceae Lineages Use Classical Sub-Types and Others Mix C_4 Acid Decarboxylases

Setaria italica, classically considered an NADP-ME sub-type species, showed high transcript levels for *NADP-ME* and *NADP-MDH* in BS and M cells respectively (Figure 4A). In addition, consistent with the *NADP-ME* sub-type, in BS strands of *S. italica* transcripts encoding *PEPCK*, *NAD-ME*, *NAD-MDH*, *ASP-AT*, and *ALA-AT* were detected at low levels. Enzyme activity assays also indicated high levels of NADP-ME in *S. italica* (Figures 5 and 6). Surprisingly high levels of PEPCK enzyme activity were also found in *S. italica,* but this was not the case for *PEPCK* transcripts. This may be explainable by the differences in growth chamber conditions, though slight, between RNA samples and enzyme activity samples or an alternate protein may have generated this activity.

Urochola fusca is classically thought to exclusively use *PEPCK* to release CO_2 in the BS. The patterns of transcript accumulation in M and BS strands of *U. fusca* are consistent with *PEPCK* functioning in this species with very little to no supplemental decarboxylation from either *NADP-ME* or *NAD-ME* (Figure 4B). BS stands contained barely detectable levels of transcripts encoding *NADP-ME* and *NAD-ME*, but very high levels of those encoding *PEPCK*. Enzyme activity in *U. fusca* was consistent with transcript abundances, PEPCK having high levels and the other two decarboxylases very low levels (Figures 5 and 6). In addition, consistent with the cycling of aspartate and alanine between the two cell-types, transcripts derived from genes encoding both *ASP-AT* and *ALA-AT* were detectable in the two cell-types (Figure 4B).

147 In contrast to the above analysis of *S. italica* and *U. fusca* which seem to fit with an exclusive
148 use of one decarboxylation enzyme, analysis of *P. hallii* and *D. californica* indicated they
149 potentially use multiple C_4 acid decarboxylases during photosynthesis (Figure 4C-D). Although *P.*
150 *hallii* is classically considered to use *NAD-ME* in addition to high levels of transcripts encoding
151 *NAD-ME*, *NAD-MDH*, *ASP-AT* and *ALA-AT*, unexpectedly high levels of transcripts encoding *NADP-*
152 *ME* were detected in the BS (Figure 4C). *NAD-ME* transcript levels where still more than twice
153 those of *NADP-ME*, and the enzyme assays found much higher relative levels of NAD-ME than
154 NADP-ME (Figures 5 and 6). In the case of *Digitaria californica* which is thought to belong to the
155 *NADP-ME* sub-type, although transcripts encoding *NADP-ME* and *NADP-MDH* were abundant in
156 BS and M samples respectively, *PEPCK* levels were more than double those of *NADP-ME* in the
157 BS (or perhaps the levels are similar if one considers the possibility that both *NADP-ME* genes
158 here mapped are resulting in similar functional products). Enzyme assay results showed high
159 levels of NADP-ME and much lower levels of PEPCK consistent with the traditional subtype
160 classification of this species (Figures 5 and 6).

161 To further confirm or refute these findings RNA *in situ* hybridizations were undertaken.
162 Transcript accumulation by *in situ* hybridization experiments were consistent in signal strength
163 with the RNA-seq results from above and indicated that the signals clearly localized to the
164 expected anatomical locations (Supplemental Figure 1). Previous enzyme assay results for species
165 closely related to the ones sampled here are also extremely consistent with the RNA-Seq results
166 (Supplemental Figure 2). It should be noted that the conditions for mRNA, *in situ*, and enzyme
167 assay sample collection were not identical (see Materials and Methods and
168 Discussion sections).

169

170 **C_4 Pathway Transporters**

171 The transcript abundance of various transporters related to C_4 photosynthesis were examined.
172 Some transporters had low or undetectable levels such as *DICARBOXYLATE TRANSPORTER 1*
173 *(DIT1).* Others, such *SODIUM BILE ACID SYMPORTER 2 (BASS2)*, *SODIUM:HYDROGEN*
174 *ANTIPORTER (NHD)*, *MITOCHONDRIAL DICARBOXYLATE CARRIER (DIC),* and
175 *PHOSPHATE/PHOSPHOENOLPYRUVATE TRANSLOCATOR (PPT)*, had variable levels of transcript
176 abundance across the species.

177

178 **C_4 Transcript Abundance Levels In The Closest C_3 Relative To The MPC(A) Clade**

179 Transcript abundance levels from *S. indica*, a C_3 species that is part of the most closely related
180 out group to the CMP(A) clade, were generally consistent with expectations for a C_3 species
181 (Figures 3 and 7). RBCS and RCA are more highly expressed in whole leaf tissue than in BS
182 extracts. Transcripts related to C_4 photosynthesis are also expressed at a low level in both whole
183 leaf and BS.

184 Comparisons between the *S. indica* (C_3) BS enriched and whole leaf samples showed significant
185 (adjusted p-value < 0.001) BS over-abundance for 912 gene models and whole leaf (WL) over-
186 abundance for 746 gene models (Supplemental Table 1). Of these over-abundant genes,
187 significant Gene Ontology (GO) enrichment (adjusted p-value < 0.05) was found for 16 different
188 GO terms for the BS, and 14 GO terms for the M (Supplemental Table 1). The different GO terms
189 enriched in BS cells related to diverse processes including cellular transport, molecular binding,

and cell wall and membrane components. The WL GO terms related to photosystems I & II, photosynthesis, and other processes (Supplemental Table 1).

Gene Expression in C_3 versus C_4 Bundle Sheath Cells

The experimental design provides the opportunity to compare the BS expression from a C_3 species against BS expression in four closely related C_4 species. A total of 357 gene models displayed significantly (padj < 0.001, log2 fold change > 2) higher transcript abundance levels over the C_3 BS in all four C_4 species in the analysis (Supplemental Table 2). Many of these genes are expected, such as photosystem I and II subunits, cytochrome b_6f, cyclic electron chain proteins, Calvin cycle proteins, cellulose synthase, *Pectinacetylesterase*, starch synthase, and others. The remaining genes are potential candidates involved in C_4 photosynthesis that deserve further molecular and biochemical investigation (Supplemental Table 2).

Discussion

The PEPCK Sub-Type

The dominance of *PEPCK* transcript and enzyme activity over *NADP-ME* and *NAD-ME* in *U. fusca* provides evidence for the biological relevance of the classical PEPCK sub-type (Figures 4-6, Supplemental Figure 2). These data contrast with proposals that PEPCK cannot function on its own but rather is always ancillary to one of the other two C_4 acid decarboxylases (Furbank, 2011; Bräutigam et al., 2014; Wang et al., 2014). While this notion may be true in some cases, the current results suggest that it is not the case for *U. fusca*. Furthermore, our findings are supported by earlier measurements of enzyme activity made within the sub-tribe Melinidinae, where PEPCK

was also shown to be highly dominant over the other sub-types (Gutierrez et al., 1974; Prendergast et al., 1987; Lin et al., 1993), and also indicate that these differences in activity are due to differences in steady-state transcript abundance rather than post-transcriptional modifications that act to suppress accumulation of NADP-ME and NAD-ME.

C_4 Sub-Type Mixing

Of the four C_4 species examined, *P. hallii* and *D. californica* show the most evidence of sub-type mixing. The potential for mixing has previously been considered in *Panicum virgatum*, a close relative of *P. hallii* (Zhang et al., 2013; Meyer et al., 2014; Rao and Dixon, 2016; Rao et al., 2016). Rao et al. (2016) suggested that the high abundance of *NADP-ME* transcripts may be accounted for by post-transcriptional or translational modifications but experimental evidence for testing that hypothesis is lacking.

Digitaria californica also showed some evidence of sub-type mixing. In this case, *NADP-ME* and *PEPCK* transcripts were both reasonably abundant. Although this species is classically considered to belong to the NADP-ME sub-type, transcripts encoding *PEPCK* were more than double the abundance of those of *NADP-ME*. *ASP-AT* transcript levels, which are also associated with the *PEPCK* sub-type, were high as well. However, enzyme activity data does not support this idea with NADP-ME having much higher levels than PEPCK. Some of these differences between transcript and enzyme levels could be due the enzyme activity assays being carried out on samples from a different growth chamber with somewhat different conditions than the RNA sequencing samples (See Materials and Methods).

Sacciolepis indica and the C_3 Bundle Sheath

Analysis of transcript abundance in M and BS cells from C_3 species that are closely related to C_4 species is critical to understanding how C_4 photosynthesis evolved and how it can be engineered for enhanced crop production. Although transcripts loaded onto ribosomes in the BS of C_3 *Arabidopsis thaliana* have been assessed, and this analysis provided insight into the role of the BS in eudicot plants (Aubry et al., 2014), to our knowledge, there are no equivalent data from a monocotyledonous lineage in which both C_3 and C_4 species are found. The ability to mechanically isolate intact BS cells indicates that *S. indica* has very strong BS cell walls, similar to those found in C_4 species. However, all other currently available data including phylogenetic placement and RNA-seq from this study are consistent with *S. indica* being a C_3 species. The relatively high levels of C_4 related transcripts in the BS of *S. indica* are consistent with previous work on cells around the veins of C_3 plants (Hibberd and Quick, 2002; Brown et al., 2010; Shen et al., 2016). Together, these data support the concept that some C_3 species are pre-adapted to adopt the C_4 mechanism (Gould, 1989; Christin et al., 2009; Brown et al., 2011; Christin et al., 2015; Washburn et al., 2016; Williams et al., 2016; Reyna-Llorens et al., 2018; Burgess et al., 2019). Another interesting hypothesis from this study is that perhaps *S. indica* represents a step on the pathway to becoming a C_3-C_4 intermediate, or maybe it represents a reversion to C_3 photosynthesis from an ancestral C_3-C_4 intermediate (Sage, 2004; Bräutigam and Gowik, 2016).

The *S. indica* C_3 BS shows marked similarities and differences to the BS in other species

Aubry and colleagues (2014) investigated the functions of *Arabidopsis thaliana* BS cells by labeling ribosomes within the cell type and sequencing the mRNA resident in the ribosomes. In

general, our *S. indica* BS cells displayed similar patterns to those seen in Arabidopsis. Of the 912 significantly over-abundant gene models in the *S. indica* BS, 50 of them have Arabidopsis homologs that were significantly over-abundant in BS cells within the Aubry study (Aubry et al., 2014). These genes have annotated functions relating to transport (nucleotide, peptide, amino acid, sulphate, metals, ABS transporters), metal handling, transcription regulation, protein degradation, cell wall modification, amino acid metabolism, hormone metabolism, and ATP synthesis. Other functional annotations present both in the Arabidopsis and *S. indica* upregulated BS gene sets (but not necessarily from homologous genes) included: nitrogen metabolism, glutamine synthetase, tryptophan, ethylene induced signaling and regulation, lipid metabolism, trehalose metabolism, phenylpropanoid metabolism, and sulfur regulation (Supplemental Table 1).

Similarly to Arabidopsis and maize, several trehalose metabolism genes were found to be overexpressed within the *S. indica* BS, supporting the hypothesis that metabolism of trehalose is an ancestral BS function (Chang et al., 2012; Aubry et al., 2014). The data are also consistent with the hypothesis that BS cells play an important role in sulfur transport and metabolism and in nitrogen metabolism (Aubry et al., 2014). We do note, however, that some sulfur metabolism related genes shown to be enriched in Arabidopsis BS were actually found to be depleted in the *S. indica* BS. Overall, the *S. indica* BS samples are highly consistent with previous studies on Arabidopsis and rice BS, indicating both the validity of the mechanical C_3 BS isolation done here, and the conservation of C_3 functions across these divergent species (Aubry et al., 2014; Hua et al., 2021).

The Evolution of Three C4 Sub-types in the MPC(A)

For the majority of C_4 genes examined, all five species appear to use orthologous genes, or at least their transcripts mapped to the same *S. bicolor* gene (Figures 4 and 7). This is based on the common assumption that the highest expressed gene in each species/tissue is the one being used. For CA, PEPC, PPDK, PPT, NHD, ALA-AT, NAD-MDH, PEPCK, and NAD-ME the highest expressed gene in all species was clearly the same, although in some species the gene expression was so low it is likely non-functional. NADP-MDH, NADP-ME, and ASP-AT are less clear with the highest gene being different between some species but also often having high abundance levels for both genes making it hard to conclude that the lower gene is not relevant. The lack of good genomic resources for all species involved makes it difficult to conclude if the same genes are in fact being used by all species, and therefore potentially the result of a single recruitment, or if the genes are simply close homologs and recruited to C_4 separately in different lineages.

Ancestral state reconstructions were also performed based on the transcript abundance and enzyme activity data, however, these analyses were inconclusive and have been excluded due to the low phylogenetic sampling of the clade within this study.

Materials and Methods

Plant Materials

Accessions from five plant species were used in this study: *Setaria italica* yugu1, *Urochloa fusca* LBJWC-52, *Panicum hallii* FIL2, *Digitaria californica* PI 364670, and *Sacciolepis indica* PI 338609. More details on the accessions can be found in Washburn et al. (2015) with exception

of *P. hallii* FIL2, obtained from Thomas Juenger of the University of Texas at Austin with further details at *Panicum hallii* v2.0, DOE-JGI, http://phytozome.jgi.doe.gov/.

Plant materials for RNA sequencing were grown in controlled growth chambers at the University of Missouri in Columbia. Plants were grown under 16 hours of light (from 6:00-20:00) and 8 hours of darkness with temperatures of 23C during the day and 20C at night. Lights were placed between 86-88 cm above the plants. Plantings were grown in 4 replicates in a completely randomized design with 32 plants per replicate (except for *Sacciolepis indica* where plants were smaller and grown with 64 plants per replicate). The third leaf was sampled between 11:00 and 15:00 using established leaf rolling and mechanical BS isolation methods with some modifications (See Supplemental Protocol 1) (Sheen and Bogorad, 1985; Chang et al., 2012; Covshoff et al., 2013; John et al., 2014). Due two time and cost constraints only 3 of the 4 replicates (each based on a pool of plants) were processed for sequencing.

The protocol used for obtaining BS strands in *S. indica* was the same as that used for the C_4 species. Variations on the amount of time for each blending step were investigated, but only resulted in higher levels of contamination as viewed under a microscope. That said, even when microscopic examination indicated higher contamination levels in some *S. indica* BS samples, transcript abundance levels were qualitatively similar to samples with less apparent contamination. One step that may have been key to the isolation of C_3 BS strands, was the use of leaf rolling on the sampled leaves just prior to the BS strand isolation procedure (Furbank et al., 1985; John et al., 2014). This enables the removal of at least some M sap prior to BS isolation.

M enriched samples where not successfully obtained for *S. indica* in this study because of the sensitivity of the C_3 leaves to rolling. Very small amounts of leaf rolling pressure resulted in the

322 leaves becoming damaged to the point that the purity of M sap obtained from them was called
323 into question. Leaves that were rolled gently enough not to damage the BS strands and
324 contaminate the M sap resulted in sap with insufficient quantities of RNA for sequencing. It is our
325 opinion that M sap could be sampled using: 1) low input mRNA extraction and sequencing
326 procedures, 2) a more precise instrument for leaf rolling such as that described by Leegood
327 (1985), and/or 3) further experimentation with the developmental stage at which M sap is
328 extracted.

329 This resulted in 3 replicates of BS and M (or whole leaf tissue for the C_3) for each of the 5
330 species used for a total of 30 samples used in RNA extraction, sequencing, and analysis.

331

332 **Sequencing**

333 RNA was extracted using the PureLink® RNA Mini Kit (Invitrogen, Carlsbad, CA, USA) and
334 mRNA-seq libraries were constructed and sequenced by the University of Missouri DNA Core
335 Facility using the TruSeq Stranded mRNA Sample Prep Kit (Illumina, Inc., San Diego, CA, USA) and
336 the Illumina HiSeq and NextSeq platforms.

337

338 **Analysis**

339 Each mRNA sample was quality trimmed and mapped to the *Sorghum bicolor* genome version
340 3.1.1 (Paterson et al., 2009; DOE-JGI, 2017). All species were mapped to *S. bicolor* because
341 reference genomes do not exist for some of the species in this study and the reference genomes
342 that do exist are of variable quality leading to bias. Mapping all species to S. bicolor, which is
343 equally related to all, also allows for orthology assignment during the mapping step as opposed

to later in the process. Raw sequence was processed using Trimmomatic and Trinity v2.8.4 following the workflows outlined on the Trinity website (Grabherr et al., 2011; Haas et al., 2013; Bolger et al., 2014). This processing included the use of eXpress and Bowtie2 for read mapping and counting as well as edgeR and DESeq2 for differential expression analysis (Robinson et al., 2010; Li and Dewey, 2011; Langmead and Salzberg, 2012; McCarthy et al., 2012; Love et al., 2014). A list of known C_4 photosynthesis genes was compiled based on the literature; a custom script and BLAST were then used to find the appropriate homologous genes and to compare their relative abundance levels (Camacho et al., 2009; Chang et al., 2012; Covshoff et al., 2013; Bräutigam et al., 2014; John et al., 2014; Tausta et al., 2014; Rao et al., 2016). For comparisons across all cell types and species within this study, the Trimmed Mean of M-values (TMM) method described by Robinson and Oshlack (2010) as implemented in DESeq2 was used. All scripts and workflows used in the analysis can be found in a Bitbucket repository at https://bitbucket.org/washjake/paniceae_c4_m_bs_mrna/.

Enzyme assays

Enzyme activity assays were performed based on methods described in (Ashton et al., 1990; Marshall et al., 2007). Samples were grown in growth chambers at the University of Cambridge, UK and growth conditions were matched as closely as possible to those above. The temperature was a constant 20C, 60% humidity, 300 µmol light, and 16 hours of light. Samples were prepared by grinding leaf tissue with a pestle and mortar in extraction buffer then centrifuged at 13000 x g and supernatant taken.

For PEPCK assays, we used the Walker and Leegood method that was developed to reduce proteolysis, and which has been used extensively (Walker et al., 1995; Häusler et al., 2001; Marshall et al., 2007; Sommer et al., 2012; Sharwood et al., 2016). As it has been previously reported that the forward (de-carboxylation) reaction is about 2.6 times faster than the reverse (carboxylation) reaction, the PEPCK activity *in vivo* may be higher than what we measured (Ashton et al., 1990). PEPCK extraction buffer consisted of 200 mM Bicine-KOH, pH 9.0, 20 mM MgCl2 and 5 mM DTT. NAD-ME extraction buffer consisted of 50 mM HEPES-NaOH pH 7.2, 50 mM Tricine, 2 mM MnCl2, 5 mM DTT, 0.25% w/v PVP- 40000 and 0.5% Triton X-100. NADP-ME extraction buffer consisted of 50 mM HEPES-KOH pH 8.3, 50 mM Tricine, 5 mM DTT and 0.1 mM EDTA.

For PEPCK activity, assay buffer contained of 80 mM MES-NaOH pH 6.7, 0.35 mM NADH, 5 mM DTT, 2 mM MnCl2, 2 mM ADP and 50 mM KHCO3. Background rates were measured for five minutes then 1.2 units of malate dehydrogenase was added and rates measured for a further five minutes. Assays were initiated with the addition of 2mM Phosphoenolpyruvate (PEP). For NAD-ME activity, assay buffer contained 25 mM HEPES-NaOH pH 7.2, 5 mM L-malic acid, 2 mM NAD, 5 mM DTT, 0.2 mM EDTA. Background rates were measured for five minutes then 24 mM MnCl2 and 0.1 mM coenzyme A were used to initiate the reaction. For NADP-ME activity, assay buffer contained 25 mM Tricine-KOH pH 8.3, 5 mM L-malic acid, 0.5 mM NADP, 0.1 mM EDTA. Background rates were measured for five minutes, and the assays were initiated with 2mM MgCl2.

All assays were performed in 96 well plates at 25°C in a CLARIOstar plus plate-reader (BMG labtech) in 200 µl reactions with absorbance at 340 nm measured every 60 s until steady states

were reached. Rates were calculated as the rate of reaction from the initial slope of the reaction minus any observed background rate. Rates were normalized to both protein concentration, measured using the Qubit protein assay (Life Technologies), and chlorophyll concentration, extracted using 80% acetone and calculated as in Porra et al. (1989).

RNA *in situ* hybridization

Each of the genotypes were grown in a climate-controlled growth chamber at 50% relative humidity in 16:8 light to dark cycles at 29.4° C and 23.9° C day and night temperature, respectively. These conditions were different from the original mRNA samples due to the logistics of growth chamber availability. However, since the results appear to support those from the RNAseq at a lower temperature it does not appear this temperature difference had a strong impact. Replicates of fully expanded leaf three were harvested from each genotype when plants were at vegetative stage 4 (V4) when the fourth leaf collar was visible. Along the longitudinal length of the leaf blade, mid-sections of blade tissue were dissected in 3.7% FAA at 4˚ C. Samples were vacuum infiltrated and fixed overnight at 4˚ C in 3.7% FAA. Leaf samples were dehydrated through a graded ethanol series (50%, 70%, 85%, 95%, 100%) with 3 changes in 100% ethanol; all changes were 1 hour each at 4˚ C except for the last 100% ethanol, which was overnight at 4˚ C. Samples were then passed through a graded HistoChoice® (Sigma-Aldrich) series (3:1, 1:1, 1:3 ethanol: HistoChoice®) with 3 changes in 100% HistoChoice®; all changes were 1 hour each at room temperature. Samples were then embedded in Paraplast®Plus (McCormick Scientific), sectioned to 10 μm, and hybridized as described previously (Strable and Satterlee, 2021). Two fragments for *PEPCK* consisted of 450 base pairs

(bp) of the CDS (synthesized from JSC4-6 and JSC4-7 primers) and 456 bp of the 3' end that included UTR (JSC4-4 and JSC4-5). Two fragments for *NADP-ME* consisted of 790 base pairs (bp) of the CDS (JSC4-8 and JSC4-9) and 286 bp of the 3' end that included UTR (JSC4-10 and JSC4-11). Fragments were subcloned into pCR 4-TOPO (Invitrogen) and confirmed by Sanger sequencing. Antisense or sense strand digoxygenin-UTP labeled RNA was generated for *PEPCK* and *NADP-ME* using a DIG RNA labeling kit (Roche). For *PEPCK* hybridizations, equal amounts of the two probes for *PEPCK* were mixed prior to hybridization. Similarly, for *NADP-ME* hybridizations, equal amounts of the two probes for *NADP-ME* were mixed prior to hybridization. Primer sequences for RNA in situ probes are provided in Supplemental Table 3.

Accession numbers

Sequence data are available on NCBI SRA (https://www.ncbi.nlm.nih.gov/sra) under accession number ….

Supplemental data files

Supplemental Table 1. Differentially expressed genes between the whole leaf and bundle sheath of *Sacciolepis indica.*

Supplemental Table 2. Differentially expressed genes that are upregulated in all four C4 species bundle sheath cells and down regulated in *Sacciolepis indica* bundle sheath.

Author Contributions

All Authors contributed to drafting and revising the manuscript. JDW, JCP, GCC, SC, and JMH conceived of the work and experimental design. JDW, SC, SSK, and JMB developed and performed the leaf rolling experiments. JDW performed the bioinformatic analysis. JS performed the RNA *in situ* hybridization experiments. PD and JMH designed and performed the enzyme assay experiments. JDW has agreed to serve as the author responsible for contact and communication.

Acknowledgments

This work was supported by grants from the University of Missouri (Mizzou Advantage, MU Research Board, and Molecular Life Science Fellowships), and the U.S. National Science Foundation (Award no. 1501406, 1710618, 1710973). Additional support comes from the U.S. Department of Agriculture, Agricultural Research Service.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Figure Legends

Figure 1. Phylogenetic relationships between a subset of species in the grass tribe Paniceae (Poaceae).
The photosynthetic type (C_3 or C_4) and C_4 sub-type of each species is labeled in the colored triangle next to it. *NADP-ME* = *NADP-DEPENDENT MALIC ENZYME*, *PCK* = *PHOSPHONENOLPYRUVATE CARBOXYKINASE*, *NAD-ME* = *NAD-DEPENDENT MALIC ENZYME*. A) Phylogeny based on nuclear genes (Vicentini et al. 2008 and Washburn et al. 2017). B) Phylogeny based on chloroplast genes (Washburn et al. 2015).

Figure 2. Representative whole leaf and bundle strands.
Images from leaves that have been rolled to remove mesophyll (M) contents or bundle sheath (BS) strands after isolation. A) *Setaria italic*a, B) *Urochloa fusca*, C) *Panicum hallii*, D) *Digitaria californica*, E) *Sacciolepis indica*. All species use the C_4 pathway except E, *Sacciolepis indica* which is a C_3 plant. The bands of cells with low chlorophyll content in M images represent the position of mesophyll cells that have collapsed and had their contents expelled during the rolling procedure. Scale bars are depicted.

Figure 3. Log_2 fold change between mesophyll (M) and bundle sheath (BS) enriched mRNA transcripts.
Species used are *Setaria italica, Urochloa fusca, Panicum hallii* and *Digitaria californica*, and *Sacciolepis indica.* Note that for *S. indica*, a C_3 species, whole leaf data is used in place of M. Genes depicted encode proteins of the core C_4 cycle that are known to be preferentially expressed in either: A) M, or B) BS cells. The number of asterisks in each box represents the p-value. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. *CA* = *CARBONIC ANHYDRASE*, *PEPC* = *PHOSPHOENOLPYRUVATE CARBOXYLASE*, *NADP-MDH* = *NADP-DEPENDENT MALATE DEHYDROGENASE, PPDK* = *PYRUVATE,ORTHOPHOSPHATE DIKINASE*, *NADP-ME* = *NADP-DEPENDENT MALIC ENZYME*, *PEPCK* = *PHOSPHONENOLPYRUVATE CARBOXYKINASE*, *NAD-ME* = *NAD-DEPENDENT MALIC ENZYME*, *NAD-MDH* = *NADP-DEPENDENT MALATE DEHYDROGENASE*, *RCA* = *RUBISCO ACTIVASE*, *RBCS* = *RUBISCO SMALL SUBUNIT*, *ASP-AT* = *ASPARAGINE-AMINOTRANSFERASE,* and *ALA-AT* = *ALANINE-AMINO TRANSFERASE*. DIT = *DICARBOXYLATE TRANSPORTER 1* , BASS2 = *SODIUM BILE ACID SYMPORTER 2*, NHD = *SODIUM:HYDROGEN ANTIPORTER*, DIC = *MITOCHONDRIAL DICARBOXYLATE CARRIER, PPT* = *PHOSPHATE/PHOSPHOENOLPYRUVATE TRANSLOCATOR.* The addition of a space and a number after the enzyme name indicates that multiple genes where mapped that may perform this function.

Figure 4. Relative transcript abundance between core C_4 enzymes within mesophyll (M) and bundle sheath (BS) extracts. Data is displayed for: A) *Setaria italica,* B) *Urochloa fusca,* C) *Panicum hallii,* and D) *Digitaria californica*. The schematics below each histogram indicate the enzyme complement associated with each of the three biochemical sub-types. *CA* = *CARBONIC ANHYDRASE*, *PEPC* = *PHOSPHOENOLPYRUVATE CARBOXYLASE*, *NADP-MDH* = *NADP-DEPENDENT MALATE DEHYDROGENASE, PPDK* = *PYRUVATE,ORTHOPHOSPHATE DIKINASE*, *NADP-ME* = *NADP-DEPENDENT MALIC ENZYME*, *PEPCK* = *PHOSPHONENOLPYRUVATE CARBOXYKINASE*, *NAD-*

ME = *NAD-DEPENDENT MALIC ENZYME*, *NAD-MDH* = *NADP-DEPENDENT MALATE DEHYDROGENASE*, *RCA* = *RUBISCO ACTIVASE*, *RBCS* = *RUBISCO SMALL SUBUNIT*, *ASP-AT* = *ASPARAGINE-AMINOTRANSFERASE,* and *ALA-AT* = *ALANINE-AMINO TRANSFERASE*. DIT = *DICARBOXYLATE TRANSPORTER 1* , BASS2 = *SODIUM BILE ACID SYMPORTER 2*, NHD = *SODIUM:HYDROGEN ANTIPORTER*, DIC = *MITOCHONDRIAL DICARBOXYLATE CARRIER, PPT = PHOSPHATE/PHOSPHOENOLPYRUVATE TRANSLOCATOR.* The addition of a space and a number after the enzyme name indicates that multiple genes where mapped that may perform this function. Error bars are plus or minus the standard error across replicates.

Figure 5. Enzyme activities of C_4 decarboxylases. *NADP-DEPENDENT MALIC ENZYME* (*NADP-ME*), *PHOSPHOENOLPYRUVATE CARBOXYKINASE* (*PEPCK*), and *NAD-DEPENDENT MALIC ENZYME* (*NAD-ME*) for *Setaria italica, Urochloa fusca, Panicum hallii,* and *Digitaria californica*. Error bars represent plus or minus the standard error across replicates.

Figure 6. Relative transcript accumulation and enzyme activities of C_4 decarboxylases. *NADP-DEPENDENT MALIC ENZYME* (*NADP-ME*), *PHOSPHOENOLPYRUVATE CARBOXYKINASE* (*PEPCK*), and *NAD-DEPENDENT MALIC ENZYME* (*NAD-ME*) transcript accumulation and enzyme activities for *Setaria italica, Urochloa fusca, Panicum hallii,* and *Digitaria californica*. Values are represented as a percentage of the total of all three decarboxylase values. Enzyme and mRNA data were not collected at the same time, but under closely matching environmental conditions.

Figure 7. Relative transcript abundance between core C_4 enzymes within whole leaf and bundle sheath (BS) extracts from *Sacciolipis indica*. A) Whole leaf*,* B) Bundle sheath*. CA* = *CARBONIC ANHYDRASE*, *PEPC* = *PHOSPHOENOLPYRUVATE CARBOXYLASE*, *NADP-MDH* = *NADP-DEPENDENT MALATE DEHYDROGENASE, PPDK* = *PYRUVATE,ORTHOPHOSPHATE DIKINASE*, *NADP-ME* = *NADP-DEPENDENT MALIC ENZYME*, *PEPCK* = *PHOSPHONENOLPYRUVATE CARBOXYKINASE*, *NAD-ME* = *NAD-DEPENDENT MALIC ENZYME*, *NAD-MDH* = *NADP-DEPENDENT MALATE DEHYDROGENASE*, *RCA* = *RUBISCO ACTIVASE*, *RBCS* = *RUBISCO SMALL SUBUNIT*, *ASP-AT* = *ASPARAGINE-AMINOTRANSFERASE,* and *ALA-AT* = *ALANINE-AMINO TRANSFERASE*. DIT = *DICARBOXYLATE TRANSPORTER 1* , BASS2 = *SODIUM BILE ACID SYMPORTER 2*, NHD = *SODIUM:HYDROGEN ANTIPORTER*, DIC = *MITOCHONDRIAL DICARBOXYLATE CARRIER, PPT = PHOSPHATE/PHOSPHOENOLPYRUVATE TRANSLOCATOR.* The addition of a space and a number after the enzyme name indicates that multiple genes where mapped that may perform this function. Error bars are plus or minus the standard error across replicates.

Literature Cited

Arp JJ, Kambhampati S, Chu KL, Koley S, Jenkins LM, Mockler TC, Allen DK (2021) Developmental Effects on Relative Use of PEPCK and NADP-ME Pathways of C_4 Photosynthesis in Maize. 2021.2006.2025.449949

Ashton AR, Burnell JN, Furbank RT, Jenkins CLD, Hatch MD (1990) The enzymes in C_4 photosynthesis. *In* PJ Lea, JB Harborne, eds, Enzymes of Primary Metabolism, Vol 3. Academic Press, London, UK, pp 39-72

Aubry S, Smith-Unna RD, Boursnell CM, Kopriva S, Hibberd JM (2014) Transcript residency on ribosomes reveals a key role for the Arabidopsis thaliana bundle sheath in sulfur and glucosinolate metabolism. The Plant Journal **78:** 659-673

Bellasio C, Griffiths H (2013) The operation of two decarboxylases (NADPME and PEPCK), transamination and partitioning of C_4 metabolic processes between mesophylll and bundle sheath cells allows light capture to be balanced for the maize C_4 pathway. Plant Physiology

Blätke MA, Bräutigam A (2019) Evolution of C_4 photosynthesis predicted by constraint-based modelling. Elife **8**

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**

Bräutigam A, Gowik U (2016) Photorespiration connects C_3 and C_4 photosynthesis. Journal of Experimental Botany **67:** 2953-2962

Bräutigam A, Schliesky S, Külahoglu C, Osborne CP, Weber APM (2014) Towards an integrative model of C_4 photosynthetic subtypes: insights from comparative transcriptome analysis

of NAD-ME, NADP-ME, and PEP-CK C_4 species. Journal of Experimental Botany **65:** 3579-3593

Brown NJ, Newell CA, Stanley S, Chen JE, Perrin AJ, Kajala K, Hibberd JM (2011) Independent and Parallel Recruitment of Preexisting Mechanisms Underlying C_4 Photosynthesis. Science **331:** 1436-1439

Brown NJ, Palmer BG, Stanley S, Hajaji H, Janacek SH, Astley HM, Parsley K, Kajala K, Quick WP, Trenkamp S, Fernie AR, Maurino VG, Hibberd JM (2010) C_4 acid decarboxylases required for C_4 photosynthesis are active in the mid-vein of the C_3 species Arabidopsis thaliana, and are important in sugar and amino acid metabolism. The Plant Journal **61:** 122-133

Brown WV (1977) The Kranz syndrome and its subtypes in grass systematics. Memoirs of the Torrey Botanical Club **23:** 1-97

Burgess SJ, Reyna-Llorens I, Stevenson SR, Singh P, Jaeger K, Hibberd JM (2019) Genome-Wide Transcription Factor Binding in Leaves from C_3 and C_4 Grasses. PNAS **31:** 2297-2314

Cacefo V, Ribas AF, Zilliani RR, Neris DM, Domingues DS, Moro AL, Vieira LGE (2019) Decarboxylation mechanisms of C_4 photosynthesis in Saccharum spp.: increased PEPCK activity under water-limiting conditions. BMC Plant Biology **19:** 144

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden T (2009) BLAST+: architecture and applications. BMC Bioinformatics **10:** 421

Chang YM, Liu WY, Shih AC, Shen MN, Lu CH, Lu MY, Yang HW, Wang TY, Chen SC, Chen SM, Li WH, Ku MS (2012) Characterizing regulatory and functional differentiation between

maize mesophyll and bundle sheath cells by transcriptomic analysis. Plant Physiol **160:** 165-177

Christin P-A, Samaritani E, Petitpierre B, Salamin N, Besnard G (2009) Evolutionary insights on C_4 photosynthetic subtypes in grasses from genomics and phylogenetics. Genome Biology and Evolution **1:** 221-230

Christin PA, Arakaki M, Osborne CP, Edwards EJ (2015) Genetic enablers underlying the clustered evolutionary origins of C_4 photosynthesis in angiosperms. Molecular Biology and Evolution **32:** 846-858

Covshoff S, Furbank RT, Leegood RC, Hibberd JM (2013) Leaf rolling allows quantification of mRNA abundance in mesophyll cells of sorghum. Journal of Experimental Botany **64:** 807-813

DOE-JGI (2017) *Sorghum bicolor* v3.1. *In*,

Edwards GE, Kanai R, Black CC (1971) Phosphoenolpyruvate carboxykinase in leaves of certain plants which fix CO2 by the C_4-dicarboxylic acid cycle of photosynthesis. Biochemical and Biophysical Research Communications **45:** 278-285

Furbank RT (2011) Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? Journal of Experimental Botany **62:** 3103-3108

Furbank RT (2016) Walking the C_4 pathway: past, present, and future. Journal of Experimental Botany **67:** 4057-4066

Furbank RT, Stitt M, Foyer CH (1985) Intercellular compartmentation of sucrose synthesis in leaves of Zea mays L. Planta **164:** 172-178

592 **Gould SJ** (1989) Wonderful life: the Burgess Shale and the nature of history, Ed 1st. W.W.
593 Norton, New York
594 **Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,**
595 **Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma**
596 **F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A** (2011) Full-length
597 transcriptome assembly from RNA-Seq data without a reference genome. Nat
598 Biotechnol **29:** 644-652
599 **Grass Phylogeny Working Group II** (2012) New grass phylogeny resolves deep evolutionary
600 relationships and discovers C_4 origins. New Phytologist **193:** 304-312
601 **Gutierrez M, Gracen VE, Edwards GE** (1974) Biochemical and cytological relationships in C_4
602 plants. Planta **119:** 279-300
603 **Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D,**
604 **Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N,**
605 **Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A**
606 (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity
607 platform for reference generation and analysis. Nat. Protocols **8:** 1494-1512
608 **Hatch M, Kagawa T, Craig S** (1975) Subdivision of C_4-Pathway Species Based on Differing C_4 Acid
609 Decarboxylating Systems and Ultrastructural Features. Functional Plant Biology **2:** 111-
610 128
611 **Hatch M, Slack C** (1966) Photosynthesis by sugar-cane leaves. A new carboxylation reaction and
612 the pathway of sugar formation. Biochemical Journal **101:** 103-111

Hatch MD (1987) C4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics **895:** 81-106

Hatch MD (1992) I can't believe my luck. Photosynthesis Research **33:** 1-14

Hatch MD, Kagawa T (1974) Activity, location and role of NAD malic enzyme in leaves with C_4-pathway photosynthesis. Austral. J. Pl. Physiol **1:** 357-369

Hatch MD, Kagawa T (1976) Photosynthetic activities of isolated bundle sheath cells in relation to differing mechanisms of C_4 pathway photosynthesis. Archives of Biochemistry and Biophysics **175:** 39-53

Häusler RE, Rademacher T, Li J, Lipka V, Fischer KL, Schubert S, Kreuzaler F, Hirsch HJ (2001) Single and double overexpression of C_4-cycle genes had differential effects on the pattern of endogenous enzymes, attenuation of photorespiration and on contents of UV protectants in transgenic potato and tobacco plants. Journal of Experimental Botany **52:** 1785-1803

Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber Andreas PM, Lercher Martin J (2013) Predicting C_4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape. Cell **153:** 1579-1588

Hibberd JM, Covshoff S (2010) The regulation of gene expression required for C_4 photosynthesis. Annu Rev Plant Biol **61:** 181-207

Hibberd JM, Quick WP (2002) Characteristics of C_4 photosynthesis in stems and petioles of C_3 flowering plants. Nature **415:** 451-454

Hua L, Stevenson SR, Reyna-Llorens I, Xiong H, Kopriva S, Hibberd JM (2021) The bundle sheath of rice is conditioned to play an active role in water transport as well as sulfur assimilation and jasmonic acid synthesis. 2021.2004.2016.440137

John CR, Smith-Unna RD, Woodfield H, Covshoff S, Hibberd JM (2014) Evolutionary Convergence of Cell-Specific Gene Expression in Independent Lineages of C_4 Grasses. Plant Physiology **165:** 62-75

Kopriva S, Weber APM (2021) Genetic encoding of complex traits. J Exp Bot **72:** 1-3

Koteyeva NK, Voznesenskaya EV, Edwards GE (2015) An assessment of the capacity for phosphoenolpyruvate carboxykinase to contribute to C_4 photosynthesis. Plant Science **235:** 70-80

Langdale JA (2011) C_4 cycles: past, present, and future research on C_4 photosynthesis. Plant Cell **23:** 3879-3892

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods **9**

Leegood RC (1985) The intercellular compartmentation of metabolites in leaves of Zea mays L. Planta **164:** 163-171

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics **12**

Lin C, Tai Y, Liu D, Ku M (1993) Photosynthetic mechanisms of weeds in taiwan. Functional Plant Biology **20:** 757-769

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15**

655 **Majeran W, Friso G, Ponnala L, Connolly B, Huang M, Reidel E, Zhang C, Asakura Y, Bhuiyan**
656 **NH, Sun Q, Turgeon R, van Wijk KJ** (2010) Structural and Metabolic Transitions of C(4)
657 Leaf Development and Differentiation Defined by Microscopy and Quantitative
658 Proteomics in Maize. The Plant Cell **22:** 3509-3542
659 **Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U** (2014)
660 The role of photorespiration during the evolution of C_4 photosynthesis in the genus
661 Flaveria. eLife **3:** e02478
662 **Marshall DM, Muhaidat R, Brown NJ, Liu Z, Stanley S, Griffiths H, Sage RF, Hibberd JM** (2007)
663 Cleome, a genus closely related to Arabidopsis, contains species spanning a
664 developmental progression from C_3 to C_4 photosynthesis. **51:** 886-896
665 **McCarthy DJ, Chen Y, Smyth GK** (2012) Differential expression analysis of multifactor RNA-Seq
666 experiments with respect to biological variation. Nucleic Acids Research **40:** 4288-4297
667 **Meyer E, Aspinwall MJ, Lowry DB, Palacio-Mejía JD, Logan TL, Fay PA, Juenger TE** (2014)
668 Integrating transcriptional, metabolomic, and physiological responses to drought stress
669 and recovery in switchgrass (Panicum virgatum L.). BMC Genomics **15:** 527
670 **Niklaus M, Kelly S** (2019) The molecular evolution of C_4 photosynthesis: opportunities for
671 understanding and improving the world's most productive plants. J Exp Bot **70:** 795-804
672 **Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G,**
673 **Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T,**
674 **Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M,**
675 **Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC,**
676 **Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R,**

677 **Peterson DG, Mehboob ur R, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS**
678 (2009) The Sorghum bicolor genome and the diversification of grasses. Nature **457:** 551-
679 556

680 **Pick TR, Brautigam A, Schluter U, Denton AK, Colmsee C, Scholz U, Fahnenstich H, Pieruschka**
681 **R, Rascher U, Sonnewald U, Weber AP** (2011) Systems analysis of a maize leaf
682 developmental gradient redefines the current C_4 model and provides candidates for
683 regulation. Plant Cell **23:** 4208-4220

684 **Pinto H, Powell JR, Sharwood RE, Tissue DT, Ghannoum O** (2016) Variations in nitrogen use
685 efficiency reflect the biochemical subtype while variations in water use efficiency reflect
686 the evolutionary lineage of C_4 grasses at inter-glacial CO_2. **39:** 514-526

687 **Porra RJ, Thompson WA, Kriedemann PE** (1989) Determination of accurate extinction
688 coefficients and simultaneous equations for assaying chlorophylls a and b extracted with
689 four different solvents: verification of the concentration of chlorophyll standards by
690 atomic absorption spectroscopy. Biochimica et Biophysica Acta (BBA) - Bioenergetics
691 **975:** 384-394

692 **Prendergast H, Hattersley P, Stone N** (1987) New structural/biochemical associations in leaf
693 blades of C_4 grasses (Poaceae). Functional Plant Biology **14:** 403-420

694 **Raghavendra AS** (1980) Characteristics of plant species intermediate between C_3 and C_4
695 pathways of photosynthesis: their focus of mechanism and evolution of C_4 syndrome.
696 Photosynthetica **14:** 271-273

697 **Rao X, Dixon RA** (2016) The Differences between NAD-ME and NADP-ME Subtypes of C_4
698 Photosynthesis: More than Decarboxylating Enzymes. Frontiers in Plant Science **7**

699 **Rao X, Lu N, Li G, Nakashima J, Tang Y, Dixon RA** (2016) Comparative cell-specific
700 transcriptomics reveals differentiation of C_4 photosynthesis pathways in switchgrass and
701 other C_4 lineages. Journal of Experimental Botany **67:** 1649-1662
702 **Rawsthorne S** (1992) C_3–C_4 intermediate photosynthesis: linking physiology to gene expression.
703 **2:** 267-274
704 **Reyna-Llorens I, Burgess SJ, Reeves G, Singh P, Stevenson SR, Williams BP, Stanley S, Hibberd**
705 **JM** (2018) Ancient duons may underpin spatial patterning of gene expression in C_4
706 leaves. PNAS **115:** 1931-1936
707 **Robinson MD, McCarthy DJ, Smyth GK** (2010) edgeR: a Bioconductor package for differential
708 expression analysis of digital gene expression data. Bioinformatics **26**
709 **Robinson MD, Oshlack A** (2010) A scaling normalization method for differential expression
710 analysis of RNA-seq data. Genome Biology **11:** R25
711 **Sage RF** (2001) Environmental and evolutionary preconditions for the origin and diversification
712 of the C_4 photosynthetic syndrome. Plant Biology **3:** 202-213
713 **Sage RF** (2004) The evolution of C_4 photosynthesis. New Phytologist **161:** 341-370
714 **Sage RF, Christin PA, Edwards EJ** (2011) The C_4 plant lineages of planet earth. Journal of
715 Experimental Botany **62:** 3155-3169
716 **Sage RF, Sage TL, Kocacinar F** (2012) Photorespiration and the evolution of C_4 photosynthesis.
717 Annual Review of Plant Biology **63:** 19-47
718 **Schluter U, Weber APM** (2020) Regulation and Evolution of C_4 Photosynthesis. Annu Rev Plant
719 Biol **71:** 183-215

Sharwood R, Sonawane BV, Ghannoum O, Whitney SJJoEB (2016) Improved analysis of C_4 and C_3 photosynthesis via refined in vitro assays of their carbon fixation biochemistry. **67:** 3137 - 3148

Sharwood RE, Sonawane BV, Ghannoum O (2014) Photosynthetic flexibility in maize exposed to salinity and shade. Journal of Experimental Botany **65:** 3715-3724

Sheen J-Y, Bogorad L (1985) Differential Expression of the Ribulose Bisphosphate Carboxylase Large Subunit Gene in Bundle Sheath and Mesophyll Cells of Developing Maize Leaves Is Influenced by Light. Plant Physiology **79:** 1072-1076

Sheen J (1999) C_4 Gene Expression. Annual Review of Plant Physiology & Plant Molecular Biology **50:** 187

Shen W, Ye L, Ma J, Yuan Z, Zheng B, LV C, Zhu Z, Chen X, Gao Z, Chen G (2016) The existence of C_4-bundle-sheath-like photosynthesis in the mid-vein of C_4 rice. Rice **9:** 20

Sommer M, Brautigam A, Weber AP (2012) The dicotyledonous NAD malic enzyme C_4 plant Cleome gynandra displays age-dependent plasticity of C_4 decarboxylation biochemistry. Plant Biol (Stuttg) **14:** 621-629

Sonawane BV, Sharwood RE, Whitney S, Ghannoum O (2018) Shade compromises the photosynthetic efficiency of NADP-ME less than that of PEP-CK and NAD-ME C_4 grasses. Journal of Experimental Botany **69:** 3053-3068

Strable J, Satterlee JW (2021) Detecting Spaciotemporal Transcript Accumulation in Maize by RNA In Situ Hybridization. Bio-protocol **11:** e3931

Tausta SL, Li P, Si Y, Gandotra N, Liu P, Sun Q, Brutnell TP, Nelson T (2014) Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C_4-related processes. Journal of Experimental Botany

Vicentini A, Barber JC, Aliscioni SS, Giussani LM, Kellogg EA (2008) The age of the grasses and clusters of origins of C_4 photosynthesis. Global Change Biology **14:** 2963-2977

von Caemmerer S, Ghannoum O, Furbank RT (2017) C_4 photosynthesis: 50 years of discovery and innovation. Journal of Experimental Botany **68:** 97-102

Walker RP, Acheson RM, Técsi LI, Leegood RC (1997) Phosphoenolpyruvate Carboxykinase in C_4 Plants: Its Role and Regulation. Functional Plant Biology **24:** 459-468

Walker RP, Trevanion SJ, Leegood RC (1995) Phosphoenolpyruvate carboxykinase from higher plants: Purification from cucumber and evidence of rapid proteolytic cleavage in extracts from a range of plant tissues. Planta **196:** 58-63

Wang C, Guo L, Li Y, Wang Z (2012) Systematic Comparison of C_3 and C_4 Plants Based on Metabolic Network Analysis. BMC Systems Biology **6:** S9

Wang Y, Bräutigam A, Weber APM, Zhu X-G (2014) Three distinct biochemical subtypes of C_4 photosynthesis? A modelling analysis. Journal of Experimental Botany **65:** 3567-3578

Washburn JD, Bird KA, Conant GC, Pires JC (2016) Convergent Evolution and the Origin of Complex Phenotypes in the Age of Systems Biology. International Journal of Plant Sciences **177:** 305-318

Washburn JD, Schnable JC, Conant GC, Brutnell TP, Shao Y, Zhang Y, Ludwig M, Davidse G, Pires JC (2017) Genome-Guided Phylo-Transcriptomic Methods and the Nuclear Phylogentic Tree of the Paniceae Grasses. Sci Rep **7:** 13528

762 **Washburn JD, Schnable JC, Davidse G, Pires JC** (2015) Phylogeny and photosynthesis of the
763 grass tribe Paniceae. American Journal of Botany **102:** 1493-1505
764 **Weissmann S, Ma F, Furuyama K, Gierse J, Berg H, Shao Y, Taniguchi M, Allen DK, Brutnell TP**
765 (2016) Interactions of C_4 Subtype Metabolic Activities and Transport in Maize Are
766 Revealed through the Characterization of DCT2 Mutants. The Plant Cell **28:** 466-484
767 **Williams BP, Burgess SJ, Reyna-Llorens I, Knerova J, Aubry S, Stanley S, Hibberd JM** (2016) An
768 Untranslated cis-Element Regulates the Accumulation of Multiple C_4 Enzymes in
769 Gynandropsis gynandra Mesophyll Cells. The Plant Cell **28:** 454-465
770 **Wingler A, Robert PW, Zhi-Hui C, Leegood RC** (1999) Phosphoenolpyruvate Carboxykinase Is
771 Involved in the Decarboxylation of Aspartate in the Bundle Sheath of Maize. Plant
772 Physiology **120:** 539-545
773 **Zhang J-Y, Lee Y-C, Torres-Jerez I, Wang M, Yin Y, Chou W-C, He J, Shen H, Srivastava AC,**
774 **Pennacchio C, Lindquist E, Grimwood J, Schmutz J, Xu Y, Sharma M, Sharma R, Bartley**
775 **LE, Ronald PC, Saha MC, Dixon RA, Tang Y, Udvardi MK** (2013) Development of an
776 integrated transcript sequence database and a gene expression atlas for gene discovery
777 and analysis in switchgrass (Panicum virgatum L.). The Plant Journal **74:** 160-173

778
779

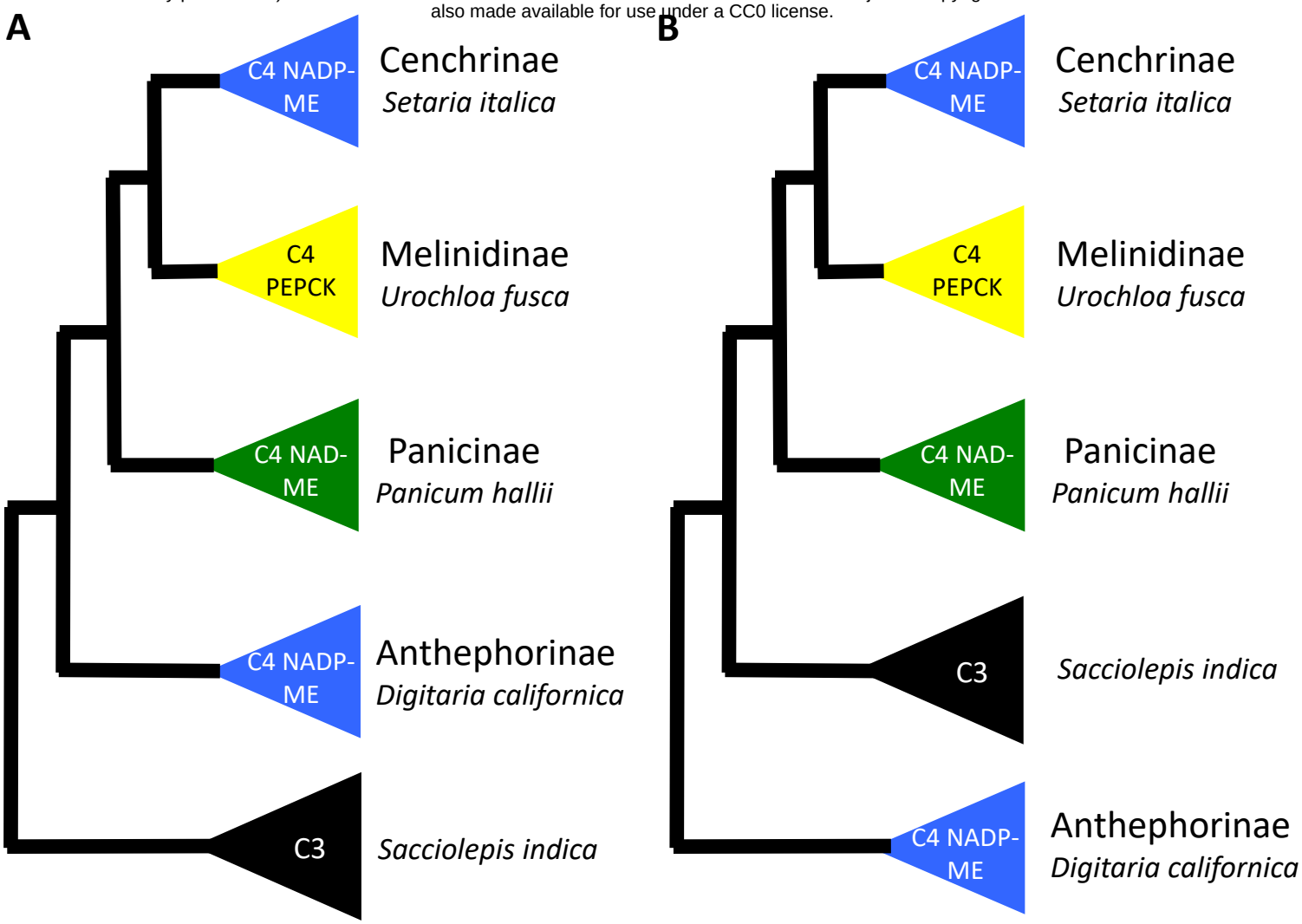

Figure 1. Phylogenetic relationships between a subset of species in the grass tribe Paniceae (Poaceae).

The photosynthetic type (C_3 or C_4) and C_4 sub-type of each species is labeled in the colored triangle next to it. *NADP-ME* = *NADP-DEPENDENT MALIC ENZYME*, *PCK* = *PHOSPHONENOLPYRUVATE CARBOXYKINASE*, *NAD-ME* = *NAD-DEPENDENT MALIC ENZYME*. A) Phylogeny based on nuclear genes (Vicentini et al. 2008 and Washburn et al. 2017). B) Phylogeny based on chloroplast genes (Washburn et al. 2015).

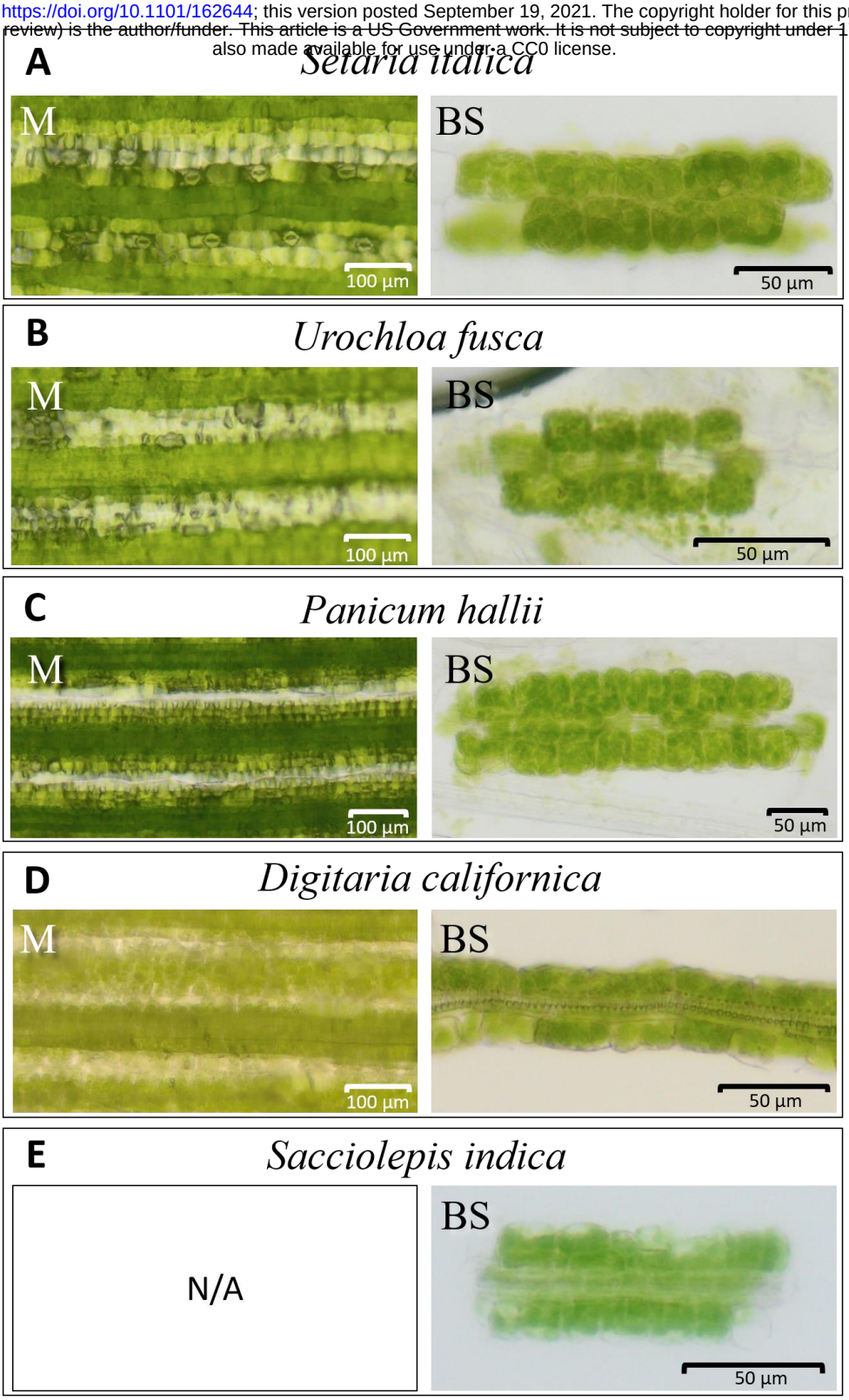

Figure 2. Representative whole leaf and bundle strands.
Images from leaves that have been rolled to remove mesophyll (M) contents or bundle sheath (BS) strands after isolation. A) *Setaria italic*a, B) *Urochloa fusca*, C) *Panicum hallii*, D) *Digitaria californica*, E) *Sacciolepis indica*. All species use the C_4 pathway except E, *Sacciolepis indica* which is a C_3 plant. The bands of cells with low chlorophyll content in M images represent the position of mesophyll cells that have collapsed and had their contents expelled during the rolling procedure. Scale bars are depicted.

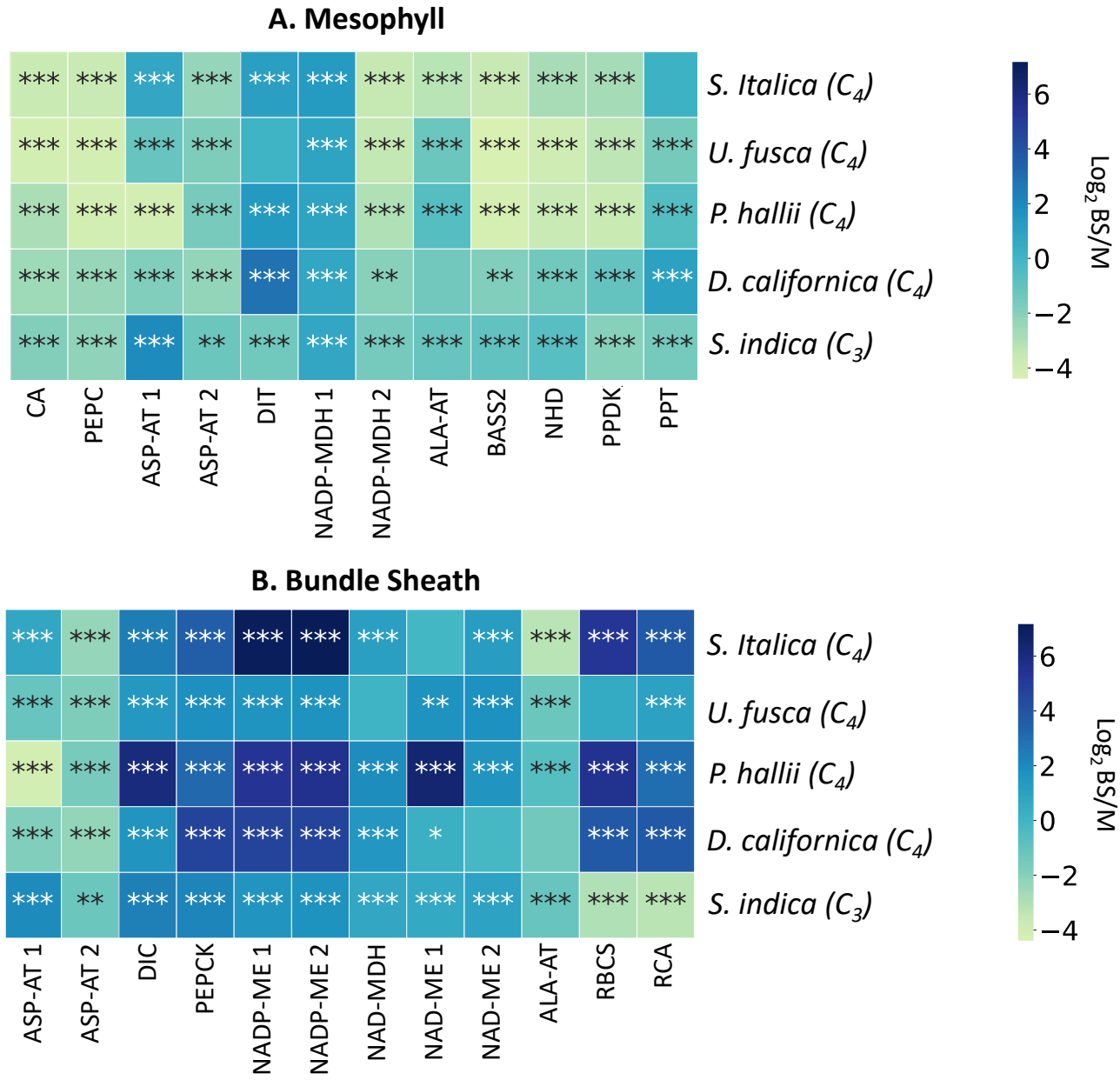

Figure 3. Log_2 fold change between mesophyll (M) and bundle sheath (BS) enriched mRNA transcripts. Species used are *Setaria italica, Urochloa fusca, Panicum hallii* and *Digitaria californica*, and *Sacciolepis indica*. Note that for *S. indica*, a C_3 species, whole leaf data is used in place of M. Genes depicted encode proteins of the core C_4 cycle that are known to be preferentially expressed in either: A) M, or B) BS cells. The number of asterisks in each box represents the p-value. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. *CA = CARBONIC ANHYDRASE, PEPC = PHOSPHOENOLPYRUVATE CARBOXYLASE, NADP-MDH = NADP-DEPENDENT MALATE DEHYDROGENASE, PPDK = PYRUVATE,ORTHOPHOSPHATE DIKINASE, NADP-ME = NADP-DEPENDENT MALIC ENZYME, PEPCK = PHOSPHONENOLPYRUVATE CARBOXYKINASE, NAD-ME = NAD-DEPENDENT MALIC ENZYME, NAD-MDH = NADP-DEPENDENT MALATE DEHYDROGENASE, RCA = RUBISCO ACTIVASE, RBCS = RUBISCO SMALL SUBUNIT, ASP-AT = ASPARAGINE-AMINOTRANSFERASE,* and *ALA-AT = ALANINE-AMINO TRANSFERASE*. DIT = *DICARBOXYLATE TRANSPORTER 1* , BASS2 = *SODIUM BILE ACID SYMPORTER 2*, NHD = *SODIUM:HYDROGEN ANTIPORTER*, DIC = *MITOCHONDRIAL DICARBOXYLATE CARRIER, PPT = PHOSPHATE/PHOSPHOENOLPYRUVATE TRANSLOCATOR*. The addition of a space and a number after the enzyme name indicates that multiple genes where mapped that may perform this function.

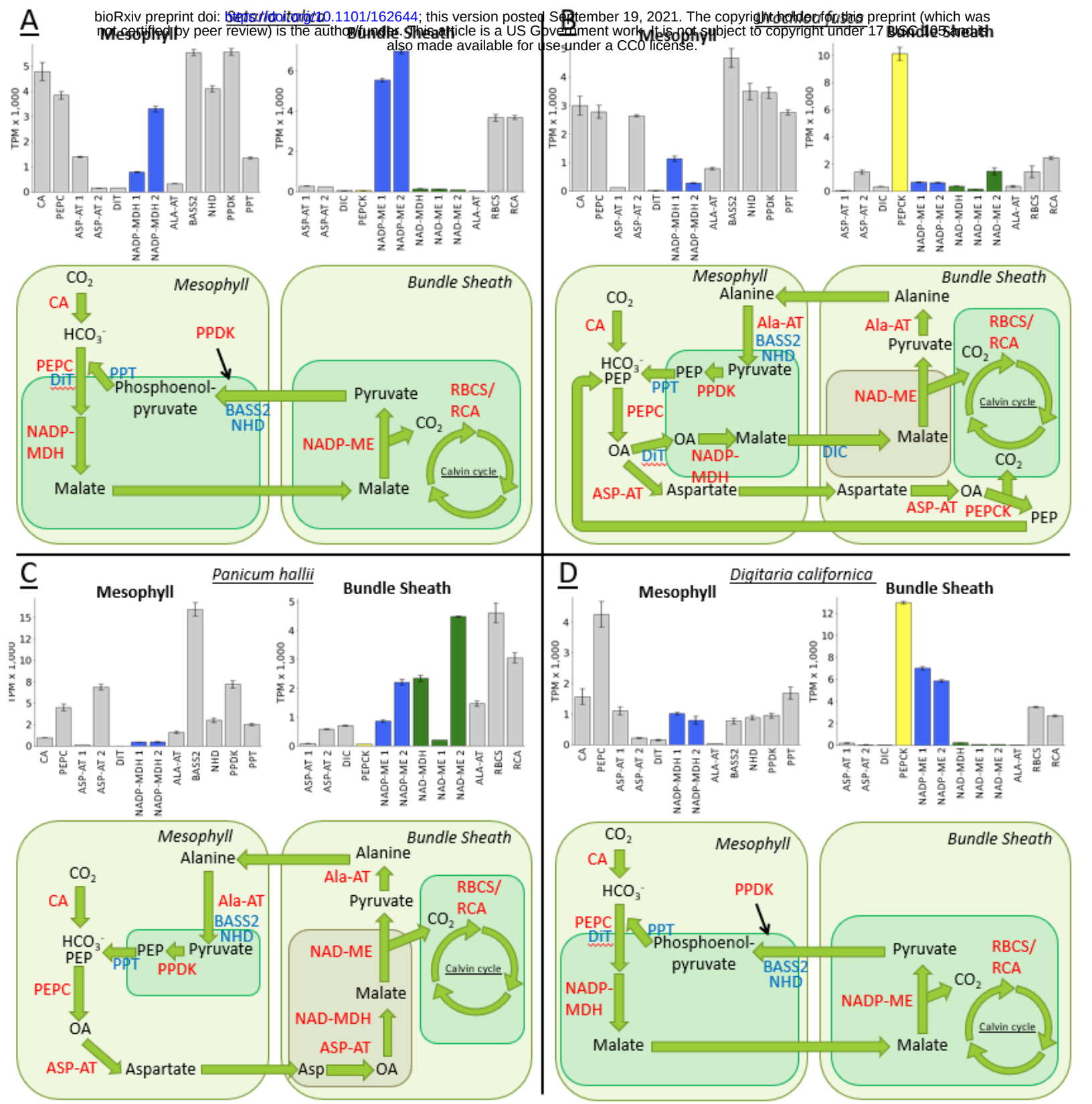

Figure 4. Relative transcript abundance between core C_4 enzymes within mesophyll (M) and bundle sheath (BS) extracts. Data is displayed for: A) *Setaria italica,* B) *Urochloa fusca,* C) *Panicum hallii,* and D) *Digitaria californica*. The schematics below each histogram indicate the enzyme complement associated with each of the three biochemical sub-types. *CA = CARBONIC ANHYDRASE*, *PEPC = PHOSPHOENOLPYRUVATE CARBOXYLASE*, *NADP-MDH = NADP-DEPENDENT MALATE DEHYDROGENASE, PPDK = PYRUVATE,ORTHOPHOSPHATE DIKINASE*, *NADP-ME = NADP-DEPENDENT MALIC ENZYME*, *PEPCK = PHOSPHONENOLPYRUVATE CARBOXYKINASE*, *NAD-ME = NAD-DEPENDENT MALIC ENZYME*, *NAD-MDH = NADP-DEPENDENT MALATE DEHYDROGENASE*, *RCA = RUBISCO ACTIVASE*, *RBCS = RUBISCO SMALL SUBUNIT*, *ASP-AT = ASPARAGINE-AMINOTRANSFERASE,* and *ALA-AT = ALANINE-AMINO TRANSFERASE*. DIT = *DICARBOXYLATE TRANSPORTER 1* , BASS2 = *SODIUM BILE ACID SYMPORTER 2*, NHD = *SODIUM:HYDROGEN ANTIPORTER*, DIC = *MITOCHONDRIAL DICARBOXYLATE CARRIER, PPT = PHOSPHATE/PHOSPHOENOLPYRUVATE TRANSLOCATOR.* The addition of a space and a number after the enzyme name indicates that multiple genes where mapped that may perform this function. Error bars are plus or minus the standard error across replicates.

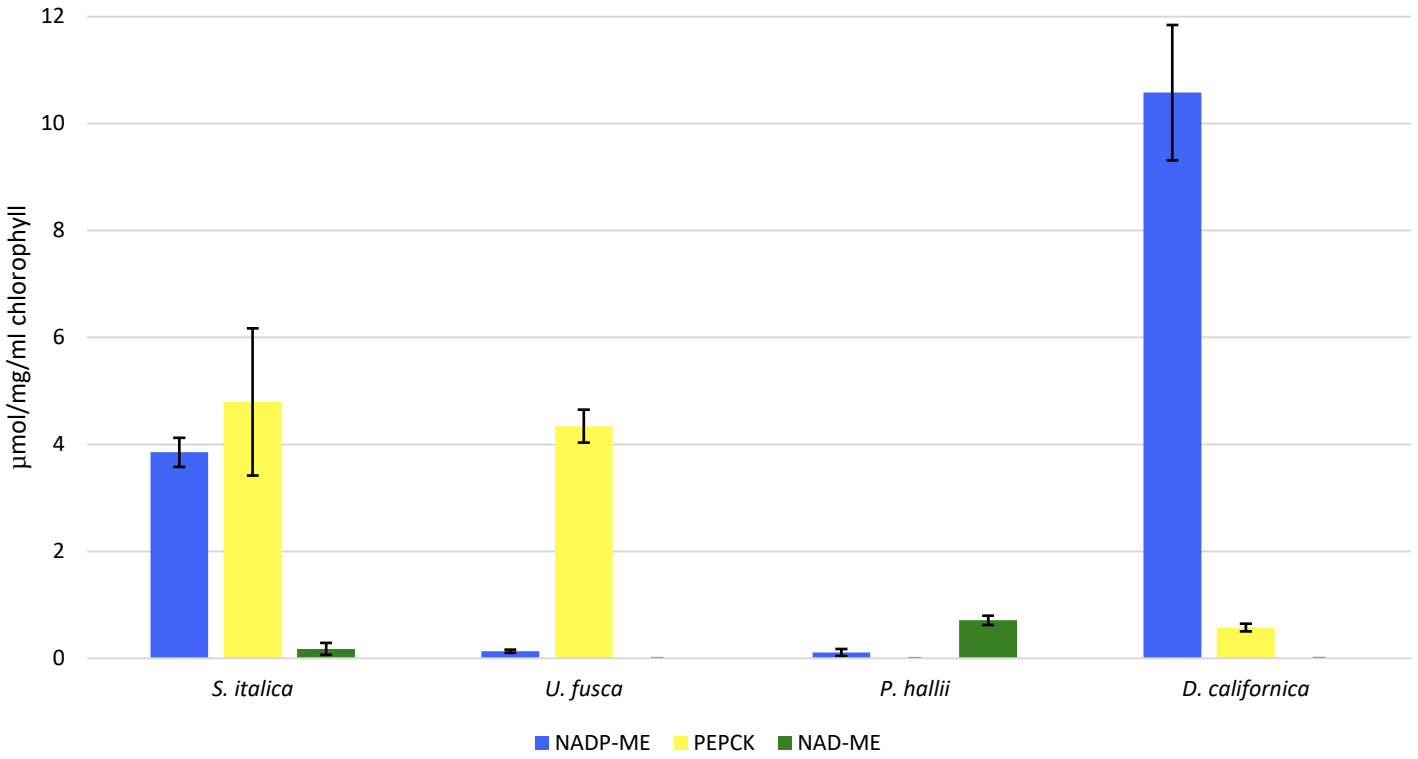

Figure 5. Enzyme activities of C_4 decarboxylases. *NADP-DEPENDENT MALIC ENZYME* (*NADP-ME*), *PHOSPHOENOLPYRUVATE CARBOXYKINASE* (*PEPCK*), and *NAD-DEPENDENT MALIC ENZYME* (*NAD-ME*) for *Setaria italica, Urochloa fusca, Panicum hallii,* and *Digitaria californica*. Error bars represent plus or minus the standard error across replicates.

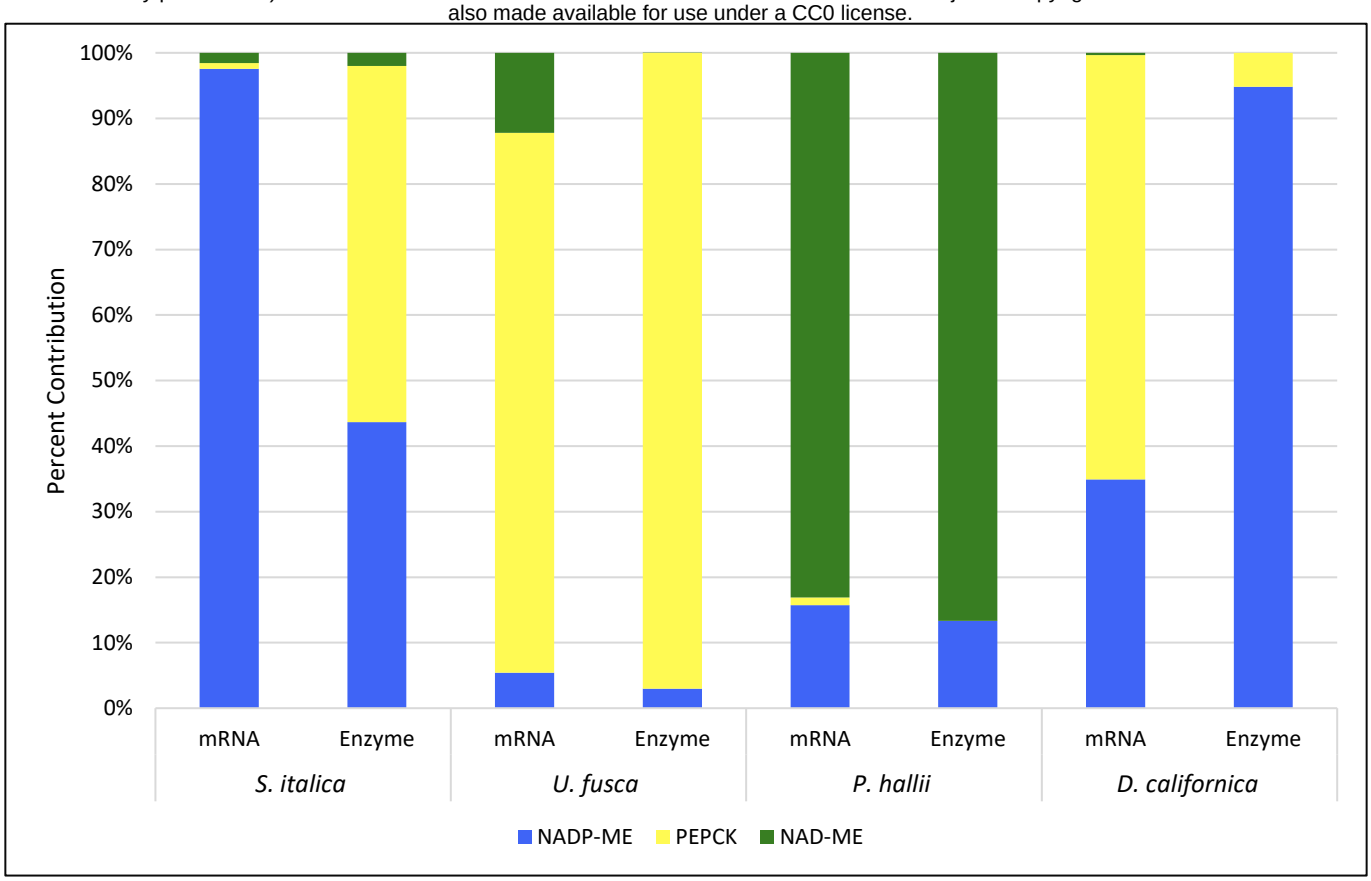

Figure 6. Relative transcript accumulation and enzyme activities of C_4 decarboxylases. *NADP-DEPENDENT MALIC ENZYME* (*NADP-ME*), *PHOSPHOENOLPYRUVATE CARBOXYKINASE* (*PEPCK*), and *NAD-DEPENDENT MALIC ENZYME* (*NAD-ME*) transcript accumulation and enzyme activities for *Setaria italica, Urochloa fusca, Panicum hallii,* and *Digitaria californica*. Values are represented as a percentage of the total of all three decarboxylase values. Enzyme and mRNA data were not collected at the same time, but under closely matching environmental conditions.

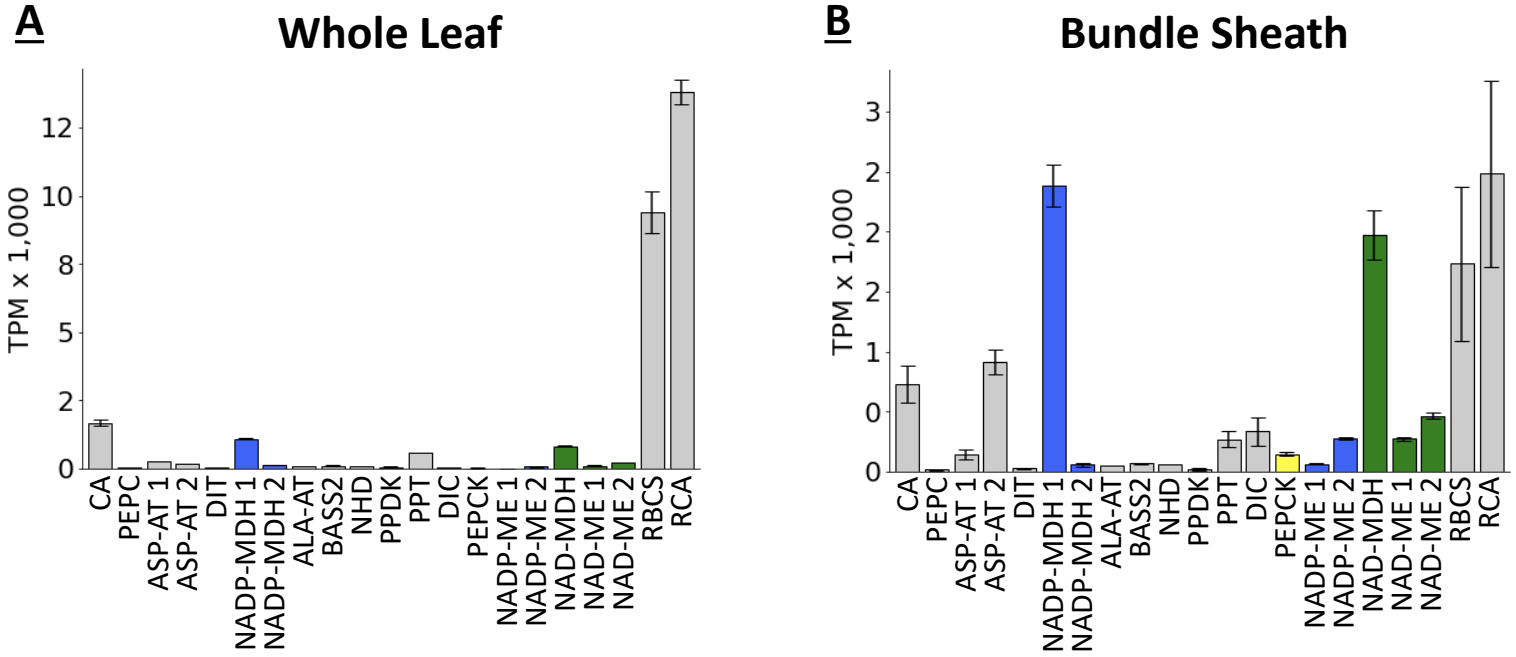

Figure 7. Relative transcript abundance between core C_4 enzymes within whole leaf and bundle sheath (BS) extracts from *Sacciolipis indica*. A) Whole leaf, B) Bundle sheath. *CA = CARBONIC ANHYDRASE*, *PEPC = PHOSPHOENOLPYRUVATE CARBOXYLASE*, *NADP-MDH = NADP-DEPENDENT MALATE DEHYDROGENASE, PPDK = PYRUVATE,ORTHOPHOSPHATE DIKINASE*, *NADP-ME = NADP-DEPENDENT MALIC ENZYME*, *PEPCK = PHOSPHONENOLPYRUVATE CARBOXYKINASE*, *NAD-ME = NAD-DEPENDENT MALIC ENZYME*, *NAD-MDH = NADP-DEPENDENT MALATE DEHYDROGENASE*, *RCA = RUBISCO ACTIVASE*, *RBCS = RUBISCO SMALL SUBUNIT*, *ASP-AT = ASPARAGINE-AMINOTRANSFERASE,* and *ALA-AT = ALANINE-AMINO TRANSFERASE*. DIT = *DICARBOXYLATE TRANSPORTER 1* , BASS2 = *SODIUM BILE ACID SYMPORTER 2*, NHD = *SODIUM:HYDROGEN ANTIPORTER*, DIC = *MITOCHONDRIAL DICARBOXYLATE CARRIER, PPT = PHOSPHATE/PHOSPHOENOLPYRUVATE TRANSLOCATOR.* The addition of a space and a number after the enzyme name indicates that multiple genes where mapped that may perform this function. Error bars are plus or minus the standard error across replicates.